NBER WORKING PAPER SERIES

BEN-PORATH MEETS LAZEAR:
LIFETIME SKILL INVESTMENT AND OCCUPATION CHOICE WITH MULTIPLE SKILLS

Costas Cavounidis
Kevin Lang

Working Paper 23367
http://www.nber.org/papers/w23367

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
April 2017

Ben-Porath meets Lazear: Lifetime Skill Investment and Occupation Choice with Multiple
Skills
Costas Cavounidis and Kevin Lang
NBER Working Paper No. 23367
April 2017
JEL No. J01,J24,J3

ABSTRACT

We develop a fairly general and tractable model of investment when workers can invest in
multiple skills and different jobs put different weights on those skills. In addition to expected
findings such as that younger workers are more likely than older workers to respond to a demand
shock by investing in skills whose value unexpectedly increases, we derive some less obvious
results. Credit constraints may affect investment even when they do not bind it equilibrium. If
there are mobility costs, firms will generally have an incentive to invest in some of their workers'
skills even when there are a large number of similar competitors, and, in equilibrium, there can be
overinvestment in all skills. Worker skill accumulation resembles learning by doing even in its
absence. We demonstrate how the model can be simulated to show the effect of a shock to the
price of individual skills.

Costas Cavounidis
Department of Economics
Boston University
270 Bay State Road
Boston, MA 02215
ccavouni@bu.edu

Kevin Lang
Department of Economics
Boston University
270 Bay State Road
Boston, MA 02215
and NBER
lang@bu.edu

Ben-Porath meets Lazear: Lifetime Skill Investment and Occupation Choice with Multiple Skills[*]

Costas Cavounidis[†]and Kevin Lang[‡]

March 1, 2017

Abstract

We develop a fairly general and tractable model of investment when workers can invest in multiple skills and different jobs put different weights on those skills. In addition to expected findings such as that younger workers are more likely than older workers to respond to a demand shock by investing in skills whose value unexpectedly increases, we derive some less obvious results. Credit constraints may affect investment even when they do not bind it equilibrium. If there are mobility costs, firms will generally have an incentive to invest in some of their workers' skills even when there are a large number of similar competitors, and, in equilibrium, there can be overinvestment in all skills. Worker skill accumulation resembles learning by doing even in its absence. We demonstrate how the model can be simulated to show the effect of a shock to the price of individual skills.

1 Introduction

For roughly five decades, the Ben-Porath (1967) model has justifiably served as the workhorse model of investment in homogeneous general human capital over the lifetime. But labor economists increasingly rely on models of multiple skills and tasks to understand earnings variation across individuals and over the lifetime. The Autor, Levy and Murnane (2003) classification of skills as manual and cognitive and tasks as routine and non-routine plays a key role in much of the literature on recent changes in the distribution of earnings. Similarly,

[†]Department of Economics, Boston University. email: ccavouni@bu.edu

[‡]Department of Economics, Boston University and NBER/IZA. email: lang@bu.edu.

the work of Heckman and coauthors (e.g. Cunha and Heckman 2008) focuses on the dynamics of investment in cognitive and noncognitive skills, particularly prior to labor market entry. Urzua and Prada (forthcoming) find that also accounting for mechanical ability greatly affects how we should think about investment in education. Bowlus, Mori and Robinson (forthcoming) explore how skill use evolves over the lifecycle. Altonji (2010), in particular, emphasizes the need for a research agenda that recognizes that skill is multidimensional and that jobs differ in their requirements.

This paper is intended to contribute to that agenda. It draws heavily on the insights of Lazear (2009) in viewing jobs as putting linear weights on skills but drops the assumption that the sum of the weights must be one. We make three main contributions:

1. We develop a relatively tractable model of post-labor market entry job choice and investment in skills. In its most abstract form, it is very flexible; workers may generalize or specialize over their lifetimes. But we parallel limited general results with an example that permits stronger conclusions and could be used for calibration.

2. We show that the comparative statics of the model are consistent with what we expect from such a model. Workers who have relatively more of one skill choose jobs that put more weight on that skill. Workers with lower discount rates and longer remaining lifetimes will tend to invest more heavily in skills that unexpectedly become more valuable.

3. We also provide some results that are, at least to us, to varying degrees surprising:

   (a) The process of skill investment and job choice generates persistence even if there is no learning by doing; workers always continue to invest in any skill they already have although not necessarily sufficiently to offset depreciation.

   (b) Credit constraints can influence investment even when in equilibrium they do not appear to be binding.

   (c) Even when all skills are completely general in the sense that there is an arbitrarily large number of jobs requiring similar skills, in the presence of mobility costs and firm bargaining power, firms generally have an incentive to invest in some, but not all, skills used in the job.

   (d) In equilibrium, between them, the worker and firm may overinvest in every skill. This happens quite generally with only two skills and for a class of feasible job sets regardless of the number of skills.

We begin with a simple two-skill example with two periods. We then generalize the two-period model and derive a number of analytic results. We conclude by examining continuous time where we generally rely on a specific functional form and numeric solutions. The continuous-time example considers the case of individuals who enter the market with equal

stocks of three skills. Initially it is optimal to invest most heavily in skill 2 and least in skill 3. Workers are then hit with a shock which would be mildly positive for those workers if it arrived just when they enter the market but which increases the value of skill 3 and reduces the value of skill 2. Not surprisingly, older workers continue in jobs that weight skill two heavily despite the adverse shock. Middle-aged workers partially adjust by gradually increasing their stock of skill 1 and shift to skill 1-intensive jobs while relatively younger workers shift to skill 3. Younger workers have a longer lifetime to exploit their skills over, and so most investment is undertaken when young. Thus, it is (perhaps surprisingly) only the very youngest workers who benefit from the shock.

2   A Simple Example

The worker begins period 0 endowed with skill levels $S_1 \geq 0$ and $S_2 \geq 0$. We treat premarket investment as exogenous. We do, however, assume that the worker can arrive in the labor market with something other than the skills that are optimal for her. This may be due to uncertainty; the value of skills in the future may be unknown, and the worker or those investing in her may wish to diversify against this uncertainty. Schooling may be insufficiently individualized or premarket skill investment may reflect goals other than maximizing market earnings.

A job J is a vector of non-negative weights on the worker's skills. At job J, she produces $J_1 S_1 + J_2 S_2$. The worker chooses from a set of jobs J given by

$$(a_1 J_1)^\sigma + (a_2 J_2)^\sigma \leq 1 \tag{1}$$

with $a_1 > 0$, $a_2 > 0$ and $\sigma > 1$. Optimal choice implies a job on the boundary of J where (1) holds with equality. Such a job that only uses one skill puts a weight of $1/a_i$ on it. When $\sigma = 2$ and $a_1 = a_2 = 1$, the boundary is the unit quarter circle in the positive quadrant; there, a job using both skills equally puts a weight of $\sqrt{.5}$ on each.

As $\sigma \to 1$, the trade-off between the skills - given by the northeast boundary - tends to a straight line. The limit is thus the (excluded) case where workers always choose to use only one skill. As $\sigma \to \infty$, the job set becomes a square, and it is thus disadvantageous to move away from using both skills equally; in effect there is only one skill.

2.1   Job Choice

In order for the model to be sensible and helpful, we need to allow for individual heterogeneity. We would not want the optimal skill vector to be the same for everyone. At the same time, we

do not wish to make the set of available jobs individual-specific. Consequently, we introduce a weighting matrix which allows the value of skills within a job to vary both over time and across individuals. Various basketball skills are less valuable for an unusually short worker than for one who is unusually tall. Similarly, knowing how to shoe horses is a skill that has declined in value even though the job persists.

A worker who chooses job J earns $\alpha_1 J_1 + \alpha_2 J_2$. She therefore maximizes

$$\alpha_1 J_1 S_1 + \alpha_2 J_2 S_2 - \lambda \left((a_1 J_1)^\sigma + (a_2 J_2)^\sigma - 1\right). \tag{2}$$

This problem has the same solution as

$$A_1 J_1 S_1 + A_2 J_2 S_2 - \lambda \left(J_1^\sigma + J_2^\sigma - 1\right) \tag{3}$$

where $A_i = \alpha_i / a_i$ and has the useful interpretation that A_i is the maximum weight on skill i in any job. We will show a similar result holds in general and will focus on this formulation throughout the paper. Changes in A can be thought of capturing changes over time in the value of a skill or differences in quality of that skill across individuals.

Maximizing gives the first-order conditions

$$A_i S_i = \lambda \sigma J_i^{\sigma - 1} \tag{4}$$

along with the constraint. Solving the first-order conditions, we have

$$J_i = \frac{(A_i S_i)^{\frac{1}{\sigma - 1}}}{\left((A_1 S_1)^{\frac{\sigma}{\sigma - 1}} + (A_2 S_2)^{\frac{\sigma}{\sigma - 1}}\right)^{\frac{1}{\sigma}}}. \tag{5}$$

Note that $dJ_i / dS_i > 0$; the higher a worker's skill, the more weight the job she chooses puts on it.

Finally, using (5), we get the value of the skill endowment

$$V(S) = \left((A_1 S_1)^{\frac{\sigma}{\sigma - 1}} + (A_2 S_2)^{\frac{\sigma}{\sigma - 1}}\right)^{\frac{\sigma - 1}{\sigma}}. \tag{6}$$

Note that this resembles a CES production function except that the exterior exponent is less than 1 rather than greater than or equal to 1. This is significant because it means that the function is convex rather than concave - as a skill increases, production would increase linearly if the job remained constant; however, the worker re-optimizes and increases the weight on that skill.

2.2 Investment

We now extend the example to two periods and allow the worker to invest in skills following production in period 0, increasing them by I at cost

$$C(I) = I_1^\rho + I_2^\rho. \tag{7}$$

Note that since skill typically has no natural scale, we can normalize the coefficients on I_i rather than writing $\alpha_i I_i^\rho$. Of course, this normalization will affect S_i and A_i, but this simplifies the problem. Following the investment choice the worker again chooses a job, so that the worker's lifetime problem is to maximize the Lagrangian

$$A_1 J_{01} S_1 + A_2 J_{02} S_2 + \beta \left(A_1 J_{11} \left(\Delta_1 S_1 + I_1 \right) + A_2 J_{12} \left(\Delta_2 S_2 + I_2 \right) \right)$$
$$- \left(I_1^\rho + I_2^\rho \right) - \lambda \left(J_{01}^\sigma + J_{02}^\sigma - 1 \right) - \mu \left(J_{11}^\sigma + J_{12}^\sigma - 1 \right) \tag{8}$$

where J_0 and J_1 are the jobs in periods 0 and 1, β is the discount factor and Δ_i is the rate at which skill i does not depreciate (1 minus the depreciation rate) between periods 0 and 1. It should be apparent that the problem is separable. First period job choice and investment do not depend on each other. Separating investment and second-period job choice and using the formula for V from (6) to simplify it, we get the maximand

$$\beta \left(\left(A_1 \left(\Delta_1 S_1 + I_1 \right) \right)^{\frac{\sigma}{\sigma-1}} + \left(A_2 \left(\Delta_2 S_2 + I_2 \right) \right)^{\frac{\sigma}{\sigma-1}} \right)^{\frac{\sigma-1}{\sigma}} - \left(I_1^\rho + I_2^\rho \right), \tag{9}$$

which yields the first-order conditions for I:

$$\beta \frac{A_i^{\frac{\sigma}{\sigma-1}} \left(\Delta_i S_i + I_i \right)^{\frac{1}{\sigma-1}}}{\left(\left(A_1 \left(\Delta_1 S_1 + I_1 \right) \right)^{\frac{\sigma}{\sigma-1}} + \left(A_2 \left(\Delta_2 S_2 + I_2 \right) \right)^{\frac{\sigma}{\sigma-1}} \right)^{\frac{1}{\sigma}}} - \rho I_i^{\rho-1}. = 0 \tag{10}$$

Claim 1 *Investment in skill i is increasing in the quantity of the skill carried forward from the previous period. That is $dI_i/d\left(\Delta_i S_i \right) > 0$.*

Proof. Apply the implicit function theorem. Since V is convex, its second derivative with respect to S_i is positive while the second derivative of (10) with respect to I_i is negative by the second-order conditions. ∎

Thus skill builds on itself. A worker who has a high level of skill chooses a job that makes greater use of that skill. Knowing that she will be in a similar job next period, the worker chooses to invest more in the type of skill that the worker uses. In a manner somewhat analogous to Lazear (2009), workers invest in skills that make them particularly good at the

type of job they currently occupy even though there is no learning by doing.

Claim 1 also implies that workers invest less in skills that depreciate more rapidly. Note that this occurs even though the investment itself does not depreciate. Instead, because the skill depreciates, workers know they will choose a job that makes less use of it.

Somewhat informally, we can view workers with higher values of β as being 'younger' in the sense that they put more weight on the future. It is evident from (10) that investment is increasing in β.

All of the results in this subsection carry over to the general two-period model.

2.3 Credit Constraints

For the remainder of the example, it is convenient to set $\Delta_i = 1 = \beta$, i.e. there is no depreciation or discounting.

Heretofore we have assumed that the worker is free to invest as much as she wants. In contrast, in Ben-Porath workers can only invest up to their current production. Restrictions on the worker's ability to invest can be used to make our model more closely resemble Ben-Porath's.

More significantly, in our model, such constraints can have important effects even when they appear not to bind. Consider the following example. As $\sigma \rightarrow 1$, the worker chooses a job that specializes in whichever skill she has in greater supply. So she either maximizes $A_1 (S_1 + I_1) - I_1^\rho$ or its equivalent for I_2. After maximizing and substituting, she compares $A_1 S_1 + A_1^{\frac{\rho}{\rho-1}} \left(\rho^{-\frac{1}{\rho-1}} - \rho^{-\frac{\rho}{\rho-1}} \right)$ with its equivalent for S_2. Normalize A_1 to 1 and let $\rho = 2$, $S_1 = 2$ and $S_2 = 0$. She will be indifferent between investing in the two skills if $A_2 = 3$. But if she invests in skill 1, she will set I_1 to .5 at a cost of .25. If she invests in skill 2, she will choose $I_2 = 1.5$ at a cost of 2.25, making a net wage of 2.25 in both cases. However, if she is constrained to spend less than 2.25, she will strictly prefer to invest in skill 1. Provided that the constraint exceeds .25, the constraint is not binding even though it eliminates her indifference. By continuity, there are values of $\sigma > 1$, $A_2 > 3$ and the constraint such that if unconstrained, she would strictly prefer to invest in skill 2 but if constrained invests in skill 1 while spending less than the constraint.

That is, although production at individual jobs is linear, job choice produces a convex skill value function, and thereby we can have a non-binding but choice-altering resource constraint. In the general model, we provide sufficient conditions for this to occur.

2.4 Mobility Costs and Firm Investment in General Human Capital

Since Becker (1971) economists have understood that workers should pay for their general training. Yet firms appear to pay for their workers to acquire general skills. Acemoglu and Pischke (1999) argue convincingly that a fixed mobility cost is insufficient to overturn this result. They argue that the firm must be able to capture some of the worker's increased productivity. In Lazear (2009), firms pay for general skills because they have some monopsony power.

In the example here, we show that in our model a fixed mobility cost combined with some firm bargaining power is sufficient for the firm to invest in one of the two skills provided that the mobility cost is sufficient to prevent mobility in equilibrium. Furthermore, the worker overinvests in one skill in order to improve her bargaining position while the firm overinvests in the other in order to deter the worker from overinvesting quite so much.

To fix ideas, consider the case where $S = [1, 0]^T$, $A = [1, 2]^T$, $\sigma = \rho = 2$, $\Delta = \mathbb{I}_{2 \times 2}$ and $\beta = 1$. That is, the worker is initially endowed with only a unit of skill 1, but skill 2 is potentially a more valuable investment since A_2 is twice A_1.

If there were no mobility cost, the worker would choose $J_0 = [1, 0]^T$, $I = [\frac{1}{3}, \frac{\sqrt{5}}{3}]^T$ and $J_1 = [\frac{2}{3}, \frac{\sqrt{5}}{3}]^T$, for a total payoff of $\frac{7}{3}$. In other words, the worker starts in a job that uses only skill 1, invests primarily in skill 2, and then chooses a period 1 job that puts slightly more weight on skill 2 than skill 1.

With a mobility cost of m levied if the worker chooses to change jobs between periods, the worker will either act as in the no-cost case and pay m, or take a single job solving

$$\max_{J,I} A_1 J_1(S_1 + (S_1 + I_1)) + A_2 J_2(S_2 + (S_2 + I_2)) - I_1^2 - I_2^2 \tag{11}$$

which, given the parameter assumptions above, reduces to

$$\max_{J,I} J_1(2 + I_1) + 2J_2 I_2 - I_1^2 - I_2^2 \tag{12}$$

subject to $J_1^2 + J_2^2 = 1$. This gives the corner solution $I = [.5, 0]$, $J = [1, 0]$, for a payoff of 2.25. Comparing the maxima under each scheme, we deduce the worker moves if $m \leq 1/12$.

If the incumbent firm can extract some of the avoided mobility cost when the worker does not move, things are quite different. Consider firms which, following production and investment by the worker in the first period, can make a take it or leave it offer for period 1 employment. Outside firms behave competitively and each offer the worker her productivity, but the incumbent firm effectively has monopsony power. If such a deal leaves the firm with positive surplus, the worker will stay at the incumbent firm and receive a wage

equal to her outside option, which is simply the highest production her skills can achieve elsewhere, minus m. But this perversely incentivizes the worker so that even when she remains at the incumbent firm, she invests *as though she was at a firm with different skill weights*. That is, the worker will invest $I = [\frac{1}{3}, \frac{\sqrt{5}}{3}]^T$, as in the case without mobility costs, despite the fact that she will not actually move to the job $[\frac{2}{3}, \frac{\sqrt{5}}{3}]^T$. Since the market in period 0 is competitive, firms make offers that foresee their monopsony power. Consequently, the worker will get a total payoff equal to her lifetime production net of investment costs, but subject to the incentive distortion.

In other words, the worker's job choice will solve

$$\max_J J_1 \left(2 + \frac{1}{3}\right) + J_2 \left(\frac{\sqrt{5}}{3}\right)$$

which yields $J = [\frac{7}{\sqrt{69}}, \frac{2\sqrt{5}}{\sqrt{69}}]$. That is, the worker's second-period distortionary incentive to invest as though she were moving to the job $[\frac{2}{3}, \frac{\sqrt{5}}{3}]^T$ will also distort her original job choice. In the end, the worker attains a reduced payoff of 2.102.

If the firm can commit to invest in the worker's skills, this inefficiency can be partially blunted. Suppose that firms can make period-0 offers that include some provision of skill investment $\widetilde{I}$. The firm incurs a cost for this investment $C(\widetilde{I})$, and the worker can invest further so that her total investment is I, by expending $C(I) - C(\widetilde{I})$. As the worker has an inefficiently high desired level of investment in skill 2, the firm will want to affect that level by committing to a level of investment in skill 1 beyond what the worker would choose on her own, $\frac{1}{3}$. By using the investment FOCs, we can show that the worker will respond to a firm investment $\widetilde{I}_1 > \frac{1}{3}, \widetilde{I}_2 = 0$ by choosing

$$I_2(\widetilde{I}_1) = \frac{\sqrt{(1 - I_1)(3 + I_1)}}{2}.$$

The offer from each firm in period 0 will therefore maximize total worker production subject to this constraint, so that the worker chooses the offer maximizing her ex-ante payoff. The job choice ends up being $J \approx [.9, .435]^T$ and the total investment $I = [.567, .621]^T$ for a total payoff of 2.144. Thus, firms' ability to commit to investing in the worker reduces the inefficiency.

Notice that the investment in skill 1 offered by the firm, .567, is greater than the worker would ever choose in the absence of the incentive problem. This is because investment in skill 1 reduces the worker's incentive to invest in skill 2. Suppose instead that the firm merely committed to investment in skill 1 at the optimal level for production at the current

8

job. Then, further skill investment in skill 1 would result in only a second-order loss of net production at the incumbent firm, but the gain in net production by reducing incentives to invest in skill 2 is first-order as the worker's investment level in skill 2 is above that which would satisfy the FOC of net production at the current firm. In other words, the firm commits to an inefficiently high level of investment in skill 1 in order to dampen the worker's inefficiently high level of investment in skill 2, as the two are seen by the worker as substitutes. The end result is overinvestment in both skills given the worker's job, but for different reasons.

3   The Two-Period Model

In this section we generalize the basic two-period model with no uncertainty. We reproduce the principal results from the example. In section 4, we derive sufficient conditions for the credit constraint and overinvestment results. That section introduces the functional form we rely on for the continuous time section. Readers who are not interested in the technical details of the general two-period model can skip to section 4.

There exist N different skills. A worker is endowed with a *skill vector* $S \geq 0^1$ so that her level of ability in skill n is S_n. We will discuss investment shortly, but first consider a worker with a fixed vector of skills.

The worker will choose a job from $\mathcal{J} \subseteq \mathbb{R}_+^N$, the *job set*. The job set represents the collection of production technologies at different jobs, in the form of the set of available skill weight vectors from which the worker can choose. This set is nonempty, strictly convex, compact and can be described in terms of a strictly convex, smooth function $F : \mathbb{R}^N \to \mathbb{R}$ and the positive orthant so that $\mathcal{J} \equiv \{J \in \mathbb{R}_+^N | F(J) \leq 0\}$.

A worker with skill vector S at job J receives a wage

$$W(J, S) = (AJ)^T S \tag{13}$$

where A is a diagonal matrix. It may be useful to think of A as allowing us to capture shifts over time or across individuals in the value of a skill in all jobs. We can scale A and $\mathcal{J}$ so that A_i can be interpreted as the maximum weight that any job puts on skill i.

We further assume that $\nabla^2 F(J)$ is positive definite when $\nabla F(J) >> 0$. Finally, if some $J_i = 0$ and $F(J) = 0$ then $\partial F(J)/\partial J_i \leq 0$. In other words, if there is a job that does not use

---

[1]Throughout this paper we use vector inequality notation as follows: $x >> y \Leftrightarrow \forall i \ x_i > y_i$; $x > y \Leftrightarrow \forall i \ x_i \geq y_i$ and $x \neq y$; $x \geq y \Leftrightarrow \forall i \ x_i \geq y_i$. Vectors are columns unless otherwise noted and A^T denotes the transpose of A. ∇ denotes the gradient operator and ∇^2 the Hessian operator. The identity matrix is $\mathbb{I}$. Finally, for a diagonal matrix A, A_i denotes the ith term on the main diagonal.

a skill, then there are jobs that put some weight on that skill, without reducing the weight on the other skills much. This will ensure the worker will choose a job that uses, at least a little, every skill she possesses. The job set therefore tells us how different skills can be combined to produce.

3.1 Single job selection

A worker who must choose a job from $\mathcal{J}$ to maximize her wage will solve the constrained program

$$\max_{J \in \mathcal{J}} W(J, S) = \max_{J \in \mathcal{J}} (AJ)^T S. \tag{P_0}$$

We can form the associated Lagrangian

$$\mathcal{L}_0 = (AJ)^T S - \lambda F(J). \tag{14}$$

When $S \neq 0$, this has a unique (from strict convexity) solution satisfying

$$AS - \lambda \nabla F(J^*) = 0 \tag{15}$$
$$F(J^*) = 0. \tag{16}$$

That such a point is a maximizer also follows from the strict convexity of F. Furthermore, since each job has linear skills weights, $J^*(\cdot)$ is homogeneous of degree 0; doubling all of a worker's skills does not change her choice of job.

3.1.1 Relation to skill endowment

As noted, when $S >> 0$, the first order condition and our assumptions guarantee that $J^* >> 0$: the worker puts at least some weight on all skills she possesses. We show now, as is quite intuitive, that this is a general result; the worker puts more weight on a skill when she has a higher endowment of that skill.

Proposition 1 *The weight that the optimal job places on a skill is increasing in that skill:*

$$\frac{\partial J_i^*}{\partial S_i} > 0. \tag{17}$$

Proof. see appendix for all remaining proofs. ■

Therefore, as a worker's particular skill grows, she will choose a job that puts more weight on it.

10

The value the worker attains with skills S is $V(S) = (AJ^*(S))^T S$. The Envelope Theorem tells us that $\nabla V(S) = AJ^*(S)$. From this and (17) we can see that the value is strictly convex in any particular skill. As her ith skill improves, the worker not only gains by becoming linearly better at her old job, but also gains by selecting jobs that increasingly involve this improved skill.

3.2 Investment

We now assume the worker lives two periods, chooses a job from $\mathcal{J}$ for each (denoted J_0 and J_1) and also has a chance to invest in the first period, with the investment increasing her skills in the second period. She will discount payoffs in the second period to the first by a factor of β.

Her skills will evolve between the two periods by way of an $N \times N$ diagonal, positive definite *non-depreciation matrix* $\Delta \leq \mathbb{I}_{N \times N}$ and an *investment vector* I.[2] Starting with skills S, the worker will have a skill vector $S' = \Delta S + I$ in the second period. That is, Δ_i is the part of the endowment in skill i that does not depreciate by period 1 due to aging. The rate of depreciation may differ among skills.

On the other hand, the worker chooses the investment I and pays $C(I)$ at the time of the investment. We impose restrictions to ensure that workers always want to invest at least a little in any skill used on the job they plan to choose, that investment is finite and that we can use standard calculus. Formally, we assume that $C : \mathbb{R}^N_+ \to \mathbb{R}_+$ is twice-differentiable, strictly increasing in each dimension of I, $C(0) = 0, \nabla C(0) = 0$, and $\nabla^2 C(I)$ is diagonal and positive definite on $\mathbb{R}^N_{++}$. We also assume that $\partial C(I)/\partial I_i$ is unbounded above.

The worker solves the problem

$$\max_{I \in \mathbb{R}^N_+, \ J_0, J_1 \in \mathcal{J}} J_0^T A S + \beta J_1^T A(\Delta S + I) - C(I) \qquad (P_{01})$$

For this problem, we form the Lagrangian

$$\mathcal{L}_{01} = J_0^T A S + \beta J_1^T A(\Delta S + I) - C(I) - \lambda F(J_1) - \mu F(J_0). \qquad (L_{01})$$

Note that J_t corresponds to the vector $(J_{t1}, J_{t2})^T$ in the example in Section 2.

Clearly, the part related to J_0 is entirely separable from the rest and follows the discussion

---

[2] Diagonality implies that depreciation of a skill depends only on the amount of that skill the worker possesses and not on the level of other skills.

of job choice with exogenous skills above. The rest of the problem becomes

$$\max_{I \in \mathbb{R}_+^N, \; J_1 \in \mathcal{J}} \beta J_1^T A (\Delta S + I) - C(I) \qquad (P_1)$$

$$\mathcal{L}_1 = \beta J_1^T A (\Delta S + I) - C(I) - \lambda F(J_1). \qquad (L_1)$$

The Lagrangian (L_1) is solved with the first order conditions[3]

$$\beta A (\Delta S + I^*) - \lambda \nabla F(J_1^*) = 0 \qquad (18)$$
$$\beta J_1^* A - \nabla C(I^*) = 0 \qquad (19)$$
$$F(J_1^*) = 0 \qquad (20)$$

and the negative semi-definite (non-bordered) Hessian, in blocks,

$$H(S, I^*, J_1^*) = \begin{bmatrix} -\lambda \nabla^2 F(J_1^*) & \beta A \\ \beta A & -\nabla^2 C(I^*) \end{bmatrix} \qquad (21)$$

3.2.1 Investment and Skill Persistence

One of the striking implications of the model is that prior investment in skills tends to make skills persistent. We show first that the worker will never completely abandon investment in a skill that she has already. Then we show that investment in a skill is weakly increasing in the level of that skill. In effect, we have a dynamic that looks very much like learning-by-doing. A worker who has a high level of some skill knows that, despite some depreciation, she will have a lot of it next period as well. And since she will have a lot next period, she will choose a job that will put a lot of weight on that skill as well. But this makes it valuable to invest even more in the skill.

Proposition 2 *A worker always continues to invest in any skill she already possesses:*

$$S_i > 0 \Rightarrow I_i > 0. \qquad (22)$$

As all skills the worker has any initial ability with will be given weight in period 1, the worker is incentivized to invest a positive amount in them as the marginal cost of doing so is 0 at an investment of 0. This does not imply that all skills improve; depreciation can dominate investment.

---

[3]Equation (19) has a solution as the additive separability of $C(I)$ across its components as well as the unboundedness above of each $\frac{\partial}{\partial I_i} C(I)$ give us that ∇C is a bijection.

We are not guaranteed a unique solution to the maximization problem, although we will have it in most generic cases. We assume a unique solution for the remainder of the two-period model and also that $S >> 0$ so that we have an interior solution, i.e. $J_1^* >> 0, I^* >> 0$. Then we have

Proposition 3 *Investment is weakly increasing in the existing endowment of a skill:*

$$\frac{\partial I_i^*}{\partial S_i} \geq 0. \tag{23}$$

That is to say, workers invest in skills at which they are already good. This is produced by the fact that costs to improve a skill do not depend on that skill's previous level. This is an expression of *specialization persistence*. Highly skilled, specialized workers will take a second-period job that is largely determined by their endowment, and are therefore incentivized to invest in a way that reflects their initial specialization.

Using (17) and the fact the J_0-part of (P_{01}) is identical to (P_0), we also have

$$\frac{\partial J_{0i}^*}{\partial S_i} \geq 0 \tag{24}$$

In other words, both skill investment and the first job's skill use are correlated with the initial endowment. This occurs despite the fact that investment costs do not depend on the first job. Therefore, what appears as learning-by-doing may instead simply be the product of aligned incentives.[4]

3.3 Comparative Statics

3.3.1 Depreciation

Skill depreciation and initial skills S enter the problem multiplicatively and identically outside of first-period job choice which is in turn entirely separable from the rest of the problem. As a consequence, the effects of Δ_i on investment and second period job choice are symmetric to those of S_i; we have $\partial I_i^*/\partial \Delta_i = (\partial I_i^*/\partial S_i) S_i/\Delta_i \geq 0$ and $\partial J_{1i}^*/\partial \Delta_i = (\partial J_{1i}^*/\partial S_i) S_i/\Delta_i > 0$.

This is significant. Despite the fact that the new investment will not depreciate by the time it is used in production, the fact that initial ability in that skill will have, means it is less worthwhile to invest in it - no one wants to 'run to stay in place'. This can simply be

---

[4]A learning-by-doing approach where it is cheaper to invest in skills used at the first job can explain internships, apprenticeships and other cases where short-term productivity (not just income) is foregone in order to facilitate learning. However, these make up a small part of total employment, so we will instead proceed with the more parsimonious model.

understood by the fact that $\partial^2 V(S)/(\partial S_i)^2 = \partial J_i^*(S)/\partial S_i > 0$; the second-period value of skills is convex in each argument, so increasing depreciation reduces a skill's marginal value.

3.3.2 Age

Although the model does not explicitly account for age, we can alter β to change the 'length' of the second period. This proportionately raises the value of both S and I. The total expenditure on investment will increase with the remaining time.

Proposition 4 *Let $\beta > \beta'$, I^* be a solution to the problem with discount factor β and $I^{*\prime}$ be a solution to the problem with discount factor β'. Then $C(I^*) > C(I^{*\prime})$.*

Therefore, younger workers spend more on investment, as is intuitive. However, this result only addresses total investment costs. It does not necessarily mean that investment in any particular individual skill will increase in age. This is illustrated by the following example.

Convex example. Take $N = 2$, $\Delta = \mathbb{I}$ with $C(I) = \frac{2}{3}I_1^{3/2} + \frac{3}{4}I_2^{4/3}$ and $F(J) = J_1^2 + J_2^2 - 1$. Suppose a worker has $S = (0,0)^T$. Then $V(\Delta S + I) = V(I) = \sqrt{I_1^2 + I_2^2}$. Fixing $C(I)$, solving for I_2 and taking a second derivative of the objective w.r.t. I_1, the result is positive for all $I_1 > 0$. Therefore, for any fixed amount of spending, corner solutions are optimal; thus, we need only examine corner solutions. For $\beta = .5\sqrt{3}$, the solution is $I = (.75, 0)$ while for $\beta = \sqrt{2}$, the solution is $I = (0, 2\sqrt{2})$; total investment costs are $.25\sqrt{3}$ and 3 respectively. Therefore, in this example, the younger worker will spend more on skill investment overall, but less on skill 1 in particular. The intuition here is as follows: investment costs are initially lower for skill 1, but rise more sharply. A worker with insufficient time to exploit her skills after investment will therefore want to make a small total investment, and skill 1 is better suited for that; by contrast, a worker with a higher β will wish to invest more in total, and skill 2 is better suited for large investments.

4 The Diagonal Form Case

We now work with a specific functional form both to address questions of specialization and generalization, and to have a flexible but workable framework for continuous time. To do this, we will assume $F - 1$ and C are diagonal forms. We will also derive conditions under which there is overinvestment in all skills in this case and conditions under which *ex ante* a credit constraint is binding even though it appears not to be *ex post*.

14

4.1 Setup

We will be working with job sets which share the form

$$F(J) = \sum_i J_i^\sigma - 1 \tag{25}$$

with $\sigma > 1$. The job-choice problem with skills S is therefore $\max_{J \in \mathbb{R}_+^N} J^T S$ subject to (25). Recall that we can always renormalize the job and that therefore the assumption that the coefficient on J_i is 1 is without loss of generality.

A diagonal form for $F + 1$ implies a $\sigma-$spherical shape; this means that skills are inter-related in the same ways, so that we can ignore questions of complementarity versus substitutability and focus on questions of specialization versus generalization. The cost function we will be using is simply a diagonal form with exponent $\rho > 1$ and unit coefficients.[5]

$$C(I) = \sum_i I_i^\rho. \tag{26}$$

4.2 Solution

Solving for the first period job choice, we obtain

$$V(S) = \left(\sum_i (A_i S_i)^{\frac{\sigma}{\sigma-1}} \right)^{\frac{\sigma-1}{\sigma}} \tag{27}$$

$$J_i^*(S) = \frac{A_i^{\frac{1}{\sigma-1}} S_i^{\frac{1}{\sigma-1}}}{\left(\sum_j (A_j S_j)^{\frac{\sigma}{\sigma-1}} \right)^{\frac{1}{\sigma}}} = A_i^{\frac{1}{\sigma-1}} S_i^{\frac{1}{\sigma-1}} V(S)^{-\frac{1}{\sigma-1}} \tag{28}$$

The investment problem is

$$\max_{J \in \mathbb{R}_+^N, I \in \mathbb{R}_+^N} \beta J^T A (\Delta S + I) \; - \; \sum_i I_i^\rho \tag{29}$$

$$\text{s.t.} \sum_i J_i^\sigma - 1 \; \leq \; 0. \tag{30}$$

---

[5]Unit coefficients are used without loss of generality; using diagonal K as cost coefficients, endowment S with productivity A would generate the same job choices, production and investment costs as unit cost coefficients, endowment $K^{\frac{1}{\rho}} S$ and productivity $K^{-\frac{1}{\rho}} A$. As we are not in the business of defining an absolute notion of a unit of skill, we allow ourselves this normalization.

15

Swapping S for $\Delta S + I$ we can use (27) to write (29) as

$$\max_{I \in \mathbb{R}^N_+} \left[\beta \left(\sum_i (A_i(\Delta_i S_i + I_i))^{\frac{\sigma}{\sigma-1}} \right)^{\frac{\sigma-1}{\sigma}} - \sum_i I_i^\rho \right]. \tag{31}$$

From there, we can compute the first order condition with respect to I_i as

$$\beta \frac{A_i^{\frac{\sigma}{\sigma-1}} (\Delta_i S_i + I_i)^{\frac{1}{\sigma-1}}}{\left(\sum_j (A_j(\Delta_j S_j + I_j))^{\frac{\sigma}{\sigma-1}} \right)^{\frac{1}{\sigma}}} = \rho I_i^{\rho-1} \tag{32}$$

and arrive at

$$\left(\frac{A_i}{A_j} \right)^\sigma \frac{\Delta_i S_i + I_i}{\Delta_j S_j + I_j} = \left(\frac{I_i}{I_j} \right)^{(\rho-1)(\sigma-1)}. \tag{33}$$

Corner solutions with $I_i = 0$ require $S_i = 0$ as usual, but here it is also necessary that $(\rho-1)(\sigma-1) \leq 1$ for them to exist.[6]

4.3 Specialization and Generalization

So far we have analyzed period 1 job selection but, as that problem is separable from period 0 job selection, we have not examined the relation between the two. In this subsection, we examine the conditions under which workers generalize or specialize in their skill set. We restrict ourselves to the case where $\Delta = \delta \mathbb{I}$ for some $\delta < 1$ so skills depreciate at the same relative rate $1 - \delta$.

In this case, a single measure of the problem's convexity can deliver many results simultaneously.

$$\frac{\partial ln \left(\frac{\frac{\partial C(I)}{\partial I_i}}{\frac{\partial C(I)}{\partial I_j}} \right)}{\partial \left(\frac{I_i}{I_j} \right)} \bigg/ \frac{\partial ln \left(\frac{\frac{\partial V(S)}{\partial S_i}}{\frac{\partial V(S)}{\partial S_j}} \right)}{\partial \left(\frac{S_i}{S_j} \right)} = (\rho-1)(\sigma-1) \tag{34}$$

The ratio of the elasticity of the relative marginal costs to the elasticity of the of the relative marginal products, $(\rho-1)(\sigma-1)$, tells us how fast costs grow as we move away from balanced investment, relative to how production grows. If $(\rho-1)(\sigma-1) > 1$, costs grow too fast for specialization to occur, and the worker generalizes. If, on the other hand, $(\rho-1)(\sigma-1) < 1$,

---

[6]To see this, notice that when $S_i = 0$, the left hand side of (32) goes to 0 slower than the right hand side as $I_i \downarrow 0$ when $(\rho-1)(\sigma-1) > 1$.

the costs grow too slow relative to the gains from job selection, so the worker will specialize.

We show the problem in the generalization case is unique.

Proposition 5 *When $(\rho - 1)(\sigma - 1) \geq 1$ and $S \neq 0$ the global maximizer of (31) is unique.*

Now, we can examine the evolution of job choice in the generalization world. When $(\rho - 1)(\sigma - 1) > 1$, the optimal second-period job when $S = 0$ defines an attractor for job weight ratios. When the ratio of the weights assigned to two skills in the first period is not the same as in the S=0 problem, the following period the ratios of skill weights will be closer to those when $S = 0$. We regard this as generalization - regardless of their initial skill endowments, all workers in such a model move towards a common, non-extreme skill set.

Proposition 6 *If $(\rho - 1)(\sigma - 1) > 1$, $\exists \delta : \Delta = \delta \mathbb{I}$ then for each i, j, if $J_{0i}^*(S)/J_{0j}^*(S) > J_{1i}^*(0)/J_{1j}^*(0)$ then $J_{0i}^*(S)/J_{0j}^*(S) > J_{1i}^*(S)/J_{1j}^*(S)$.*

4.4   Credit Constraints and Relation to the Ben-Porath Model

Other than addressing the multiplicity of skills, our model so far differs from the standard Ben-Porath model in two ways. First, it is set in discrete time. This is unimportant. Later in the paper, we explore the diagonal version of the model in continuous time. The second is that the standard presentation of the Ben-Porath model assumes that investment takes the form of foregone production. Workers devote a fraction of their time to investing. In contrast, we treat investment as a cost. Workers can purchase skills.

For the most part this distinction is unimportant and largely a matter of convenience. We can think of someone who is capable of earning \$$x$ and chooses to invest \$$y$ as foregoing a proportion y/x of her income. Where it becomes important is when the optimal $y > x$. In some ways this can be resolved by simply imposing the additional constraint that $C(I^*) \leq J_0^T S$ which would only slightly complicate the model.

However, we show in this section that a constraint can work differently than it does in the Ben-Porath model. In that model, when the constraint binds, the individual spends all of her resources on investment, which is typically interpreted as being engaged in full-time schooling. We show here that when there are multiple skills, this conclusion is no longer valid. The skill value function V is weakly convex, with the convexity inequality strict for all but parallel skill vectors. As a consequence, on R_{++}^N, $V(\cdot)$ is strictly convex in each component of S, despite the fact that any individual job's production function is linear. This can make the investment problem non-convex, and therefore produce solutions affected by the constraint, but without the constraint binding with equality. We derive sufficient conditions for this to be true.

In a world of specialization, when skill productivity varies sufficiently, there are always endowments and budget constraints such that the budget constraint does not bind with equality but affects the optimum. Let $I_c^*(\cdot)$ denote the function mapping the endowment S into optimal investment when the constraint is c.

Proposition 7 *If* $(\rho - 1)(\sigma - 1) < 1$ *and* $A_1 < max_i A_i =: A_2$ *such that* $\left(\frac{A_2}{A_1}\right)^{\frac{\rho}{\rho-1}} (1 -$
$\frac{\rho}{\delta(\rho-1)} \left(\frac{A_2}{A_1}\right)^{\frac{(\sigma-1)(\rho-1)-1}{\rho-1}}) > 1,$[7] *then* $\exists S, c : C(I_c^*(S)) < c$ *and* $C(I^*(S)) > c$.

The basic intuition can be seen in Figure 1 which shows, in an example, how the budget constraint affects both total investment and the particular skills invested in. As the worker is endowed with much skill 1, for low values of the constraint he simply continues investing in that skill. It's not worth investing in skill 1 for long, as it's productivity is mediocre, so investment is constant when $c \in [.25, 1]$. However, once the constraint is greater than 1 the worker specializes heavily in skill 2, and the constraint once again binds until $c = 4$. Beyond that, further investment is inefficient, and relaxing the constraint further has no effect on net output.
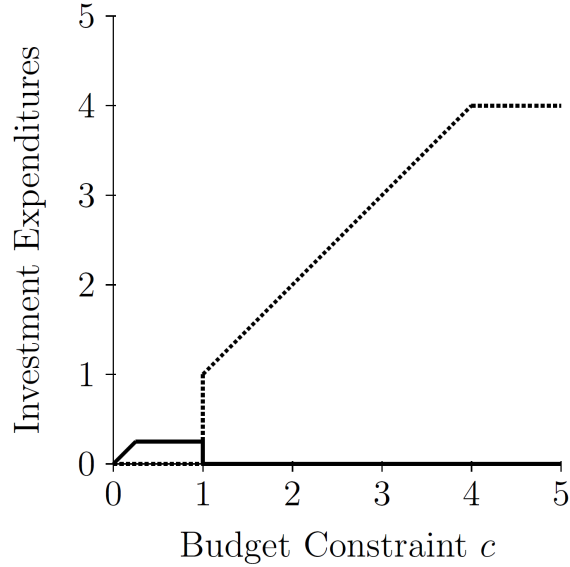

Figure 1: Investment expenditure in skill 1 (solid line) and 2 (dotted line) as a function of the budget constraint c when $A_1 = 1, A_2 = 4, S_1 = 2.75, S_2 = 0, \rho = 2, \sigma \approx 1$.

---

[7]Notice that since $(\sigma - 1)(\rho - 1) < 1$, for large enough A_2/A_1 this holds.

4.5 Mobility Costs and Overinvestment

Consider again the problem that arises when firms can offer wages to the worker and mobility costs give the incumbent firm local monopsony power. Once again, suppose that to counteract the worker's inefficient investment incentives, the firm may commit to provide some skill investment as part of the period 0 job offer. We now derive a sufficient condition for overinvestment in all skills if mobility costs are sufficient to prevent mobility in equilibrium. It turns out that this is always true provided that the worker would move in the absence of the mobility cost and that the condition for workers to generalize over their lifetime is satisfied.[8]

Let the mobility cost be m. Denote the worker's investment best response function mapping the firm's investment commitment to total investment by $I^W(\cdot)$. Note that I^W refers to the combined investment regardless of whether carried out by the firm or worker.

Proposition 8 *Let C and $F + 1$ be diagonal forms with $(\sigma - 1)(\rho - 1) > 1$. Suppose the worker with skill endowment $S >> 0$ does not move in period 1, but would absent the mobility cost. Then the optimal contract (J, I^F, w_0) satisfies $I^W(I^F) >> I^*(J)$.*

The intuition for this result is simple: to the worker, investments in different skills are substitutes. The worker's incentives are to overinvest in certain skills relative to the current job's weights in order to improve the outside option for bargaining purposes. Then, by increasing investment in other skills - those not overinvested in - the firm can dampen the worker's incentives. The firm wants to commit to overinvest in these counterweight-skills, as at the appropriate level of investment for the current job, the direct effect on net production has a zero FOC (by definition), but the efficiency gain from reducing excessive investment elsewhere is a first-order effect.

5 Continuous Time

We now wish to move the model to continuous time. We will keep the Diagonal Form structure in doing so, and retain much of the relevant intuition.

5.1 Setup

The problem is now defined over an interval in continuous time $[0, T]$, which is discounted at a rate r. The worker possesses skills $S(t)$ at time t; the productivity matrix is A and

---

[8]It additionally holds in the two-skill case in a more general setting, provided that the worker's investment is a continuous function of the firm's commitment.

the worker chooses jobs $J(t)$ from the job set $\mathcal{J} = \{J \in \mathbb{R}_+^N | \sum_i J_i^\sigma \le 1\}$ so that her time-t instantaneous production is $J(t)^T AS(t)$. Skills depreciate at relative rates given by the diagonal matrix $D := \mathbb{I} - \Delta$, counterbalanced by investment $I(t)$, so that

$$\frac{d}{dt}S(t) = -DS(t) + I(t). \tag{35}$$

However, investment is costly, with time-t instantaneous cost $C(I(t)) = \sum_i I_i(t)^\rho$ - the diagonal form. Endowed with initial skills S_0, the worker therefore seeks to maximize her lifetime utility by solving

$$\max_{J:[0,T]\to\mathcal{J},\ I:[0,T]\to\mathbb{R}_{++}^N} \int_0^\infty e^{-rt}\left[J(t)^T AS(t) - \sum_i I_i(t)^\rho\right] dt \tag{36}$$

$$s.t.\ \frac{d}{dt}S(t) = -DS(t) + I(t) \tag{37}$$

$$S(0) = S_0 \tag{38}$$

The worker chooses $I(t)$ and $J(t)$ optimally. However, as J does not influence the state variable, it is chosen according to (27). Thus, we can bypass job selection for the moment and reduce the problem to

$$\max_{I:[0,T]\to\mathbb{R}_{++}^N} \int_0^\infty e^{-rt}\left[\left(\sum_i (A_i S_i(t))^{\frac{\sigma}{\sigma-1}}\right)^{\frac{\sigma-1}{\sigma}} - \sum_i I_i(t)^\rho\right] dt \tag{39}$$

$$s.t.\ \frac{d}{dt}S(t) = -DS(t) + I(t) \tag{40}$$

$$S(0) = S_0. \tag{41}$$

We therefore construct the Hamiltonian

$$H = e^{-rt}\left[\left(\sum_i (A_i S_i(t))^{\frac{\sigma}{\sigma-1}}\right)^{\frac{\sigma-1}{\sigma}} - \sum_i I_i(t)^\rho\right] + \sum_i \mu_i(-D_i S_i(t) + I_i(t)). \tag{42}$$

The solution is given by the i equations

$$\frac{A_i^{\frac{\sigma}{\sigma-1}}(S_i(t))^{\frac{1}{\sigma-1}}}{\left(\sum_i (A_i S_i(t))^{\frac{\sigma}{\sigma-1}}\right)^{\frac{1}{\sigma}}} + \rho(\rho-1)I_i(t)^{\rho-2}\frac{dI_i(t)}{dt} - (r + D_i)\rho I_i(t)^{\rho-1} = 0 \tag{43}$$

along with the motion equations for skills

$$\frac{d}{dt}S(t) = -DS(t) + I(t), \tag{44}$$

the initial condition $S(0) = S_0$, and the transversality condition $I(T) = 0$.

5.2   Ben-Porath Case

In the special case where all skills grow at the same rate, the first term in (43) can be replaced with K_i, a skill-specific term that is constant over time. In this case the solution to (43) becomes

$$I_i(t) = \left(K_i \frac{1 - e^{(r+D_i)(t-T)}}{\rho(r + D_i)}\right)^{\frac{1}{\rho-1}}. \tag{45}$$

Note that, as required for constant K_i, the ratio of investment in any two skills is constant.

Figure 2 graphs an example of this 'Ben-Porath case'. The worker enters the market with twice as many units of skill 1 as of skill 2. Although the values of σ and ρ are chosen so that there is a tendency to generalize skills, this is exactly offset by the greater initial endowment of skill 1 and the higher productivity of skill 1 for the worker so that the worker's investment maintains the 2 : 1 ratio of skill 1 to skill 2. Net output (the wage) shows the classic hump-shaped pattern of the Ben-Porath model and peaks later than gross output. Since investment reaches 0 at exactly time T, this is the point at which the two are equal.

Unlike in the Ben-Porath model, we allow for investment in excess of production. In this and the other examples given, production net of investment starts out negative, meaning the worker is borrowing to finance the early stages of her skill investment.

5.3   Jobs and Skills Over the Lifecycle

It is generally not possible to obtain a closed form solution for (43). We can, however, solve the system numerically for given values of A, D, r and S_0. To demonstrate the potential usefulness of this approach, we present a scenario that we find particularly interesting.

The scenario has three skills which we refer to as skills 1, 2 and 3, but the reader may think of them as nonroutine manual, routine cognitive and nonroutine cognitive. Our chief example considers a worker subject to an unanticipated shock that increases the value of nonroutine cognitive skills while also decreasing the value of routine cognitive tasks. In reality, combinations of skills may have differing degrees of synergy. Nevertheless, we think even the simple example using diagonal form, which treats skills complementarity symmetrically, provides some useful insights.
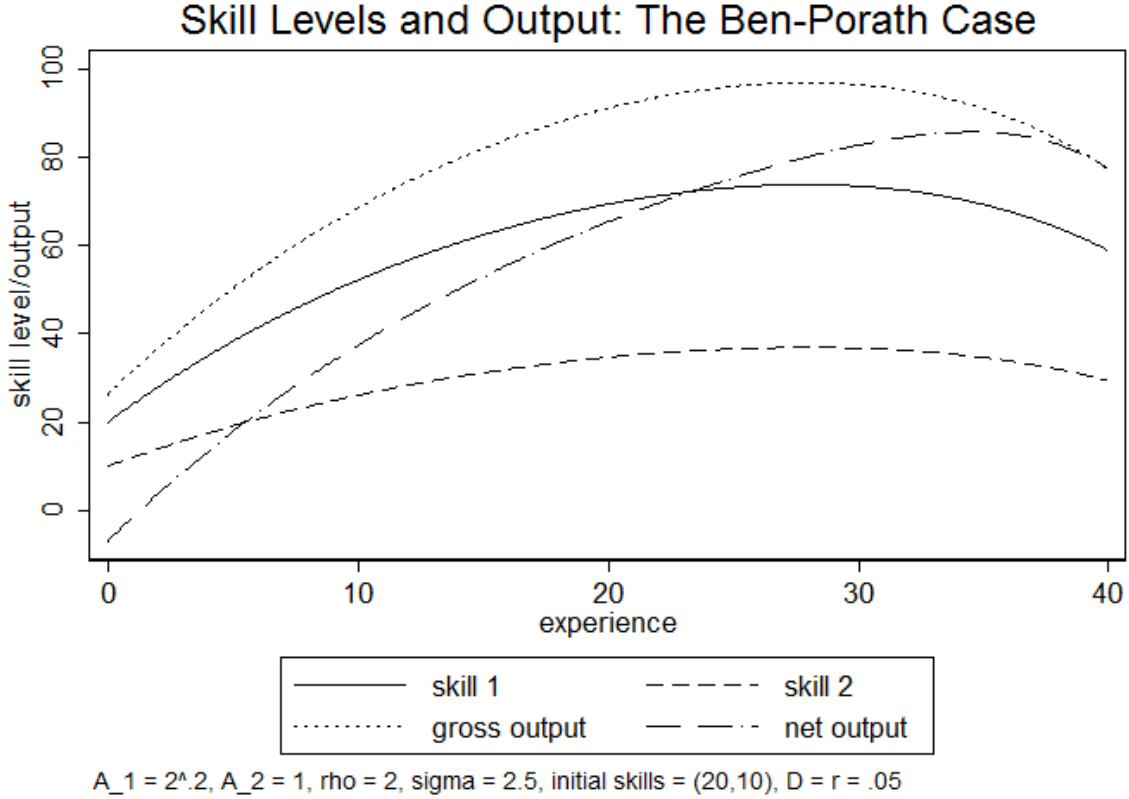
Figure 2: The case of proportional skill evolution.

We consider an individual who arrives in the labor market with 10 units of each skill. Initially the second skill is the most valuable ($A_2 = 1.2$); the first lies in the middle ($A_1 = 1.13$) and the third skill is the least valuable ($A_3 = .8$) to the worker. The worker is assumed to be in the labor force for forty years. We consider an unanticipated shock that occurs in either the worker's 10th, 20th or 30th year (and for some purposes the 35th year) in the market. The shock reduces A_2 to .8 and increases A_3 to 1.25 while leaving A_1 at 1.13. If workers typically arrive in the market with similar amounts of skills 2 and 3, the shock represents a mild form of positive shock for the youngest workers.

Figure 3 shows the path of the three skill weights for the individual if she experiences no shock and at 10, 20 or 30 years of experience. The top left corner shows the baseline with no shock. Absent the shock, the worker specializes in skill 2.

Continuing clockwise, we see that if the shock arrives when she has thirty years experience, she immediately mechanically (since A_2 falls) finds herself a job that puts less weight on skill 2. Comparing the two panels, it is also evident that she increases her net investment in skills 1 and 3 relative to her net investment in skill 2 once the shock hits since the weights

22

Job Weights by Experience by Timing of Shock

Before shock: A_1.13 = 1, A_2 = 1.2, A_3=.8, rho = 2, sigma = 2, initial skills = (10,10,10), D = r = .05
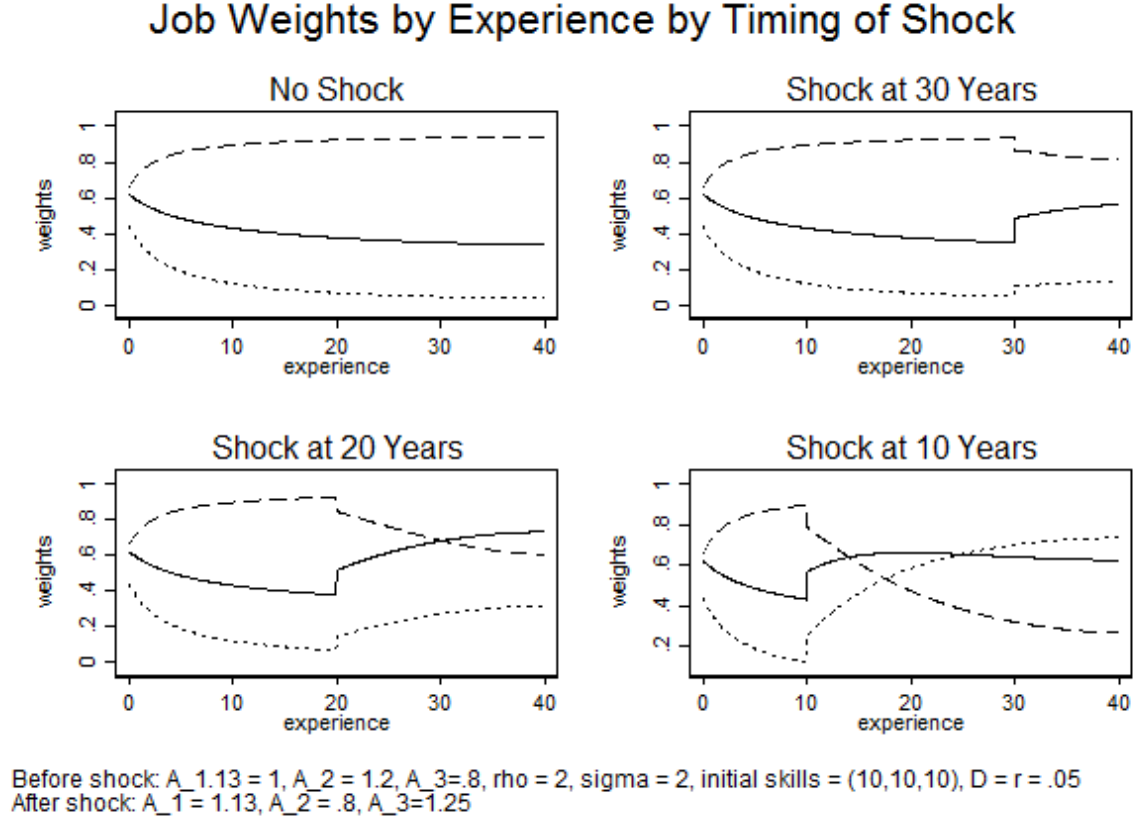After shock: A_1 = 1.13, A_2 = .8, A_3=1.25

Figure 3: Job Weights by Experience by Timing of Shock

on the former now grow rather than shrink. In fact, most of the adjustment can be shown to come from reduced investment in skill 2 rather than increased investment in the other skills. This can be seen most clearly in Figure 4 which shows the stocks of skills at each point in time. In the end, she adjusts very little. She continues to work in jobs that focus on the skill in which she has accumulated a large stock even though the value of that stock has fallen by about one third.

A shock at twenty years of experience has a more noticeable effect on shifting the job weights. But because her stock of skill 3 has depreciated so much over twenty years, by the end of her career, she shifts towards a job that places the most weight on skill 1. Much, but not all, of the increased weight on skill 3 reflects the greatly increased value of that skill in all jobs rather than a very large shift towards investment in skill 3.

Only when the shock arrives sufficiently early in her career does she adjust by investing much more heavily in skill 3 and somewhat more in skill 1 so that ten years after the shock, she works in a job that places the greatest weight on skill 3.

Figure 5 shows net output over time. As the worker has invested most heavily in a skill

Skill Levels by Experience by Timing of Shock

Before shock: A_1 = 1.13, A_2 = 1.2, A_3=.8, rho = 2, sigma = 2, initial skills = (10,10,10), D = r = .05
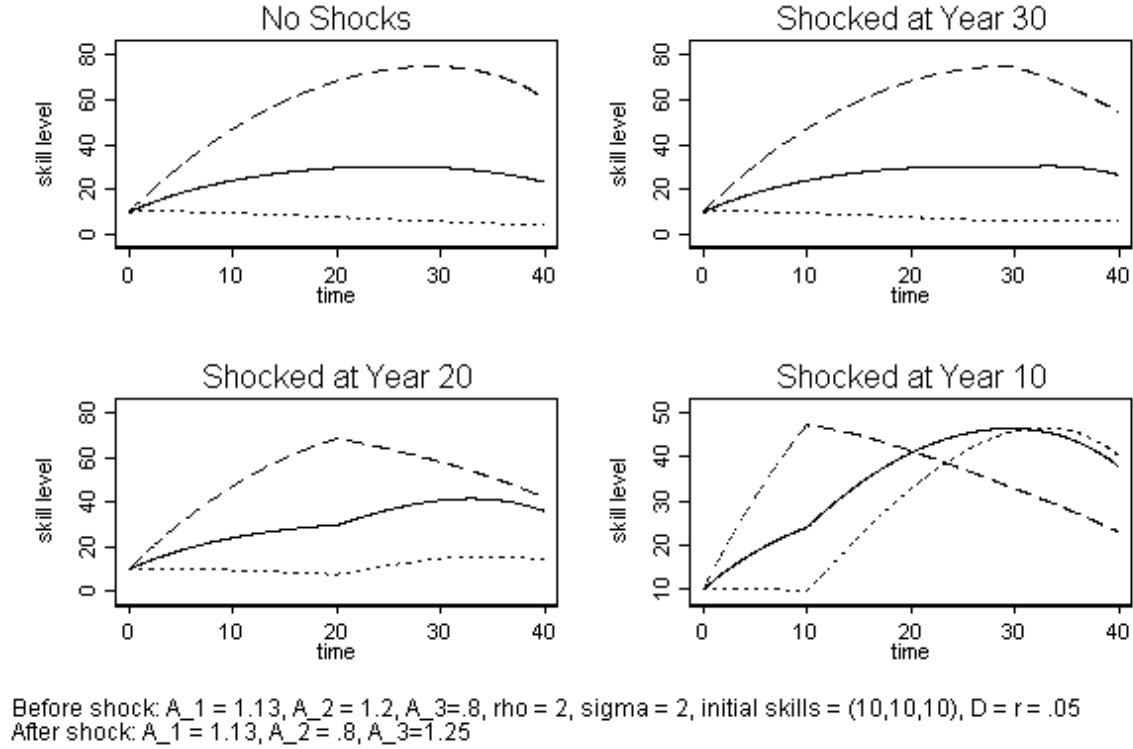After shock: A_1 = 1.13, A_2 = .8, A_3=1.25

Figure 4: Skill Levels by Experience by Timing of Shock

whose value is reduced, the worker suffers an immediate adverse shock to net output. The magnitude of the shock will depend largely on how much of skill 2 she has accumulated relative to skill 3. As a consequence, the individual shocked at 10 years of experience suffers an earlier but smaller output shock. Compared to a similar person suffering a shock at 20 years of experience, she has higher output at nearly every later experience level. When the shock hits the similar person at 30 years, the one shocked at 10 years has already recovered sufficiently to have higher net output. The person shocked at 20 years of experience fares almost as badly in the last 30 years of work as the person shocked at 30 years.

More generally, in this example in which the shock is in a sense positive, a worker who begins her career just as the shock hits will benefit. One who ends her career just as the shock hits will be unaffected. By continuity there will be a range of low experience levels at which the effect of the shock will be positive. We expect, but have not shown, that the effect of the shock is U-shaped.

We can broaden this discussion to ask who is most adversely affected by the shock. The easiest way to answer this question is by comparing the present value of net output. In
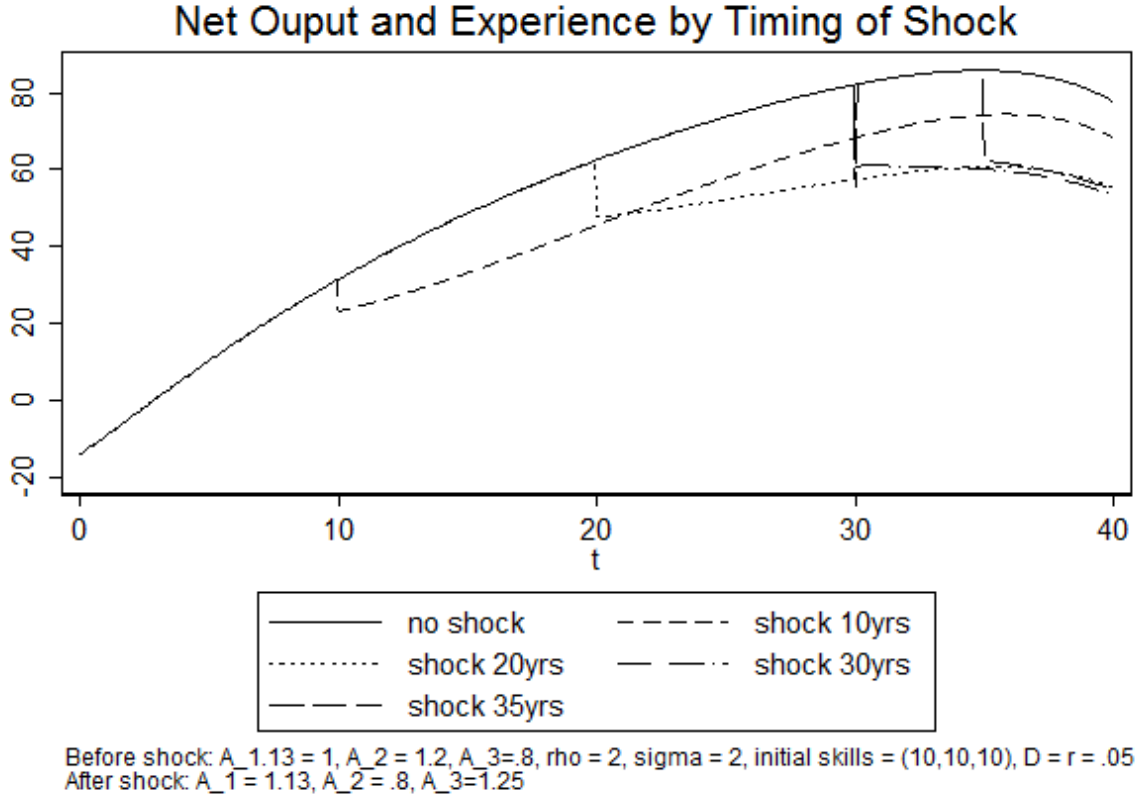
Figure 5: Net Output and Experience by Timing of Shock

our example, this value is highest for the person who is never shocked and decreases as the timing of the shock moves to 35 to 30 to 20 to 10 years of experience. Since we know that a shock at 0 years of experience would have a positive effect, the significant point is that a positive shock can have a negative effect for a very long time. In 5, the individual shocked at 10 years of experience never returns to the net output level that she would have reached in the absence of a shock.

At the same time, the PDV of net output may be a misleading measure of how adversely the worker is affected. It requires that utility be linear in income. If workers are risk averse and smooth consumption over their lifetimes, then a worker subjected to an unanticipated negative shock will have lower lifetime utility than one with the same PDV of net output but who does not experience an unanticipated shock, since the latter, but not the former, can fully smooth consumption.

If the workers in our sample smooth consumption over their lifetimes, very young workers will have accumulated less debt than somewhat older workers while workers nearing retirement will have accumulated more retirement savings than those somewhat further from

retirement. Therefore, very young workers and those nearing retirement do not need to re-duce the flow of consumption by as much as someone in between. We continue our example by assuming that people live for another 20 years following retirement and smooth their consumption perfectly except for the effect of the unanticipated shock. Here we find that a worker shocked at 20 years of experience must reduce her consumption by almost half (47 percent) relative to what she had anticipated. In contrast, workers shocked at 10 and 30 years of experience must reduce their consumption by 35 percent and 36 percent.

Perhaps the most striking aspect of the example is the length of time for which an ultimately positive shock can be negative. The same worker who spends her entire career after the shock will earn 6.3 percent more over her lifetime than if she finished her career before the shock hit. Yet even a worker who was only five years into her career never recovers from the shock and suffers a decline of about one-third in future consumption. This is because skill investment is extremely front-loaded to allow for longer exploitation time, so the loss is great even when the shock hits early. While ours is an example, not a calibration exercise, we find this duration and magnitude of the effect striking.

6 Discussion and Conclusion

We believe that our model provides both qualitative and quantitative insights. It provides as explanation for firm investments in general skills that draws on the insight of Lazear (2009) and is complementary to the story told there. It also suggests that nonconvexities arise naturally in a model of multiple skills and that these nonconvexities can create settings in which credit constraints affect behavior even though they are not binding in equilibrium.

When extended to a continuous time setting, our model is tractable and suggests that large shocks, even if positive on net, can have long-lasting adverse effects on even relatively young workers. While a full treatment of the effect of trade deals which raise the price of some skills and lower others would require us to model complementarities among skills more fully and to consider building mobility costs into the simulation, our example should make us think very carefully about the winners and losers and perhaps even the political economy issues.

Since Ricardo, arguments for trade and technological innovation have relied on compen-sating transfers. But our model suggests two additional considerations. First, losers may be difficult to detect. Losers include a) workers who continue using similar skills but the value of those skills has declined, b) workers who are able to shift to jobs whose value has not been greatly altered but whose value is less than that of the job the worker would otherwise have held, and c) even some workers who eventually shift to jobs that place a relatively

high weight on the skills that have increased in value. Moreover, our analysis suggests that the importance of credit constraints for limiting transitions to better jobs may be hidden because workers may be unable to afford to acquire the optimal set of new skills even though they do not appear to be credit constrained. While it may be optimal for some workers to continue to work in jobs that have declined in value, other workers may cling to such jobs because they cannot afford retraining. For workers to adjust more adequately to skill-biased technical and trade shocks, transfer and retraining policies may have to start as soon as foresight permits.

7 References

Acemoglu, Daron and Steve Pischke. "Beyond Becker: Training in Imperfect Labor Markets," *Economic Journal*, 109 (February 1999): F112-142.

Altonji, Joseph G., "Multiple Skills, Multiple Types of Education, and the Labor Market: A Research Agenda" (September 2010). American Economic Association, *Ten Years and Beyond: Economists Answer NSF's Call for Long-Term Research Agendas*. Available at SSRN: http://ssrn.com/abstract=1888514 or http://dx.doi.org/10.2139/ssrn.1888514

Autor, David H., Frank Levy, and Richard J. Murnane, "The Skill Content of Recent Technological Change: An Empirical Exploration," *Quarterly Journal of Economics,* 118(4) (November 2003):1279-1333.

Becker, Gary S. *Human Capital,* 2nd ed., Chicago: Chicago University Press, 1975.

Ben-Porath, Yoram, "The Production of Human Capital and the Life Cycle of Earnings," *Journal of Political Economy,* 75(4, Part 1) (August 1967): 352-365.

Bowlus, Audra J., Hiroaki Mori and Chris Robinson, "Ageing and the Skill Portfolio: Evidence from Job Based Skill Measures," *Journal of the Economics of Ageing*, 7 (April 2016):89-103.

Cunha, Flavio and James J. Heckman, "Formulating, Identifying, and Estimating the Technology of Cognitive and Noncognitive Skill Formation," *Journal of Human Resources* 43 (2008), 738-782.

Fiacco, Anthony V. "Sensitivity analysis for nonlinear programming using penalty methods." Mathematical programming 10.1 (1976): 287-311.

Lazear, Edward, "Firm-Specific Human Capital: A Skill-Weights Approach," *Journal of Political Economy*, 117 (October 2009): 914 - 940.

Prada, Maria F., and Sergio S. Urzua^{o}a. "One Size Does Not Fit All: Multiple Dimensions of Ability, College Attendance and Wages". *Journal of Labor Economics*, forthcoming..

A Proofs

A.1 Proof of Proposition 1

From Fiacco (1976), we have that J^* is differentiable with respect to S. Taking a derivative of (15) with respect to S_i, we have

$$[0, ...0, A_i, 0, ..., 0]^T - \frac{\partial \lambda}{\partial S_i} \nabla F(J^*) - \lambda \nabla^2 F(J^*) \frac{\partial J^*}{\partial S_i} = 0. \tag{46}$$

Differentiating (16) with respect to S_i, we obtain $\frac{\partial J^*}{\partial S_i}^T \nabla F(J^*) = 0$. Premultiplying (46) with $\frac{\partial J^*}{\partial S_i}^T$ and using that fact, we have

$$A_i \frac{\partial J_i^*}{\partial S_i} - \frac{\partial \lambda}{\partial S_i} \frac{\partial J^*}{\partial S_i}^T \nabla F(J^*) = \lambda \frac{\partial J^*}{\partial S_i}^T \nabla^2 F(J^*) \frac{\partial J^*}{\partial S_i} \tag{47}$$

$$\lambda^{-1} A_i \frac{\partial J_i^*}{\partial S_i} = \frac{\partial J^*}{\partial S_i}^T \nabla^2 F(J^*) \frac{\partial J^*}{\partial S_i} \tag{48}$$

As $S >> 0$, $\nabla F(J^*) > 0$ and therefore by assumption $\nabla^2 F(J^*)$ is positive definite; thus, the right hand side term is positive as a quadratic form on a positive definite matrix and we have $\frac{\partial J_i^*}{\partial S_i} > 0$ as required.

A.2 Proof of Proposition 2

Substituting for J_1^* using (19) in (18) and (20) we have

$$\beta A(\Delta S + I^*) - \lambda \nabla F(\beta^{-1} A^{-1} \nabla C(I^*)) = 0 \tag{49}$$

$$F(\beta^{-1} A^{-1} \nabla C(I^*)) = 0 \tag{50}$$

and also, substituting in the block Hessian:

$$H(S, I^*, \beta^{-1} A^{-1} \nabla C(I^*)) = \begin{bmatrix} -\lambda \nabla^2 F(\beta^{-1} A^{-1} \nabla C(I^*)) & \beta A \\ \beta A & -\nabla^2 C(I^*) \end{bmatrix} \tag{51}$$

From (49), $S_i > 0$ implies that $\frac{\partial}{\partial J_i} F(J) > 0$. From this and the assumption that $J_i = 0 \Rightarrow \frac{\partial}{\partial J_i} F(J) \leq 0$, we have that $\frac{\partial}{\partial I_i} C(I^*) > 0$; but as we've assumed that $\nabla C(0) = 0$ and that C is additively separable , it must be that $I_i^* > 0$ to satisfy $\frac{\partial}{\partial I_i} C(I^*) > 0$.

A.3 Proof of Proposition 3

Differentiating (49) and (50) with respect to S_i, recalling that Δ and A are diagonal and suppressing functional arguments, we have

$$[0,..,0,\beta A_i \Delta_i, 0,..,0]^T + \beta A \frac{\partial I^*}{\partial S_i} - \frac{\partial \lambda}{\partial S_i} \nabla F - \lambda \beta^{-1} \nabla^2 F A^{-1} \nabla^2 C \frac{\partial I^*}{\partial S_i} = 0 \qquad (52)$$

$$(\nabla F)^T A^{-1} \nabla^2 C \frac{\partial I^*}{\partial S_i} = 0. \qquad (53)$$

Premultiplying (52) with $(\frac{\partial I^*}{\partial S_i})^T \nabla^2 C(I^*) A^{-1}$ and recalling $\nabla^2 C$ is diagonal, we get

$$\beta \Delta_i \frac{\partial I_i^*}{\partial S_i} \frac{\partial^2 C}{\partial I_i^{*2}} + \beta \left(\frac{\partial I^*}{\partial S_i}\right)^T \nabla^2 C \frac{\partial I^*}{\partial S_i} = \lambda \beta^{-1} \left(\frac{\partial I^*}{\partial S_i}\right)^T \nabla^2 C A^{-1} \nabla^2 F A^{-1} \nabla^2 C \frac{\partial I^*}{\partial S_i} \qquad (54)$$

$$\Delta_{ii} \frac{\partial I_i^*}{\partial S_i} \frac{\partial^2 C}{\partial I_i^{*2}} = -\beta^{-2} \left(\frac{\partial I^*}{\partial S_i}\right)^T \nabla^2 C A^{-1} \left[-\lambda \nabla^2 F + \beta^2 A (\nabla^2 C)^{-1} A\right] A^{-1} \nabla^2 C \frac{\partial I^*}{\partial S_i} \qquad (55)$$

As $-\lambda \nabla^2 F + (\beta A)(\nabla^2 C)^{-1}(\beta A)$ is the Schur complement of the (negative semi-definite) Hessian with respect to the (negative definite) block $-\nabla^2 C$, it is negative semi-definite. Therefore, as the negative of a quadratic form on a negative semi-definite matrix, the right hand side of (55) a whole is nonnegative; thus proving the proposition.

A.4 Proof of Proposition 4

Suppose I^* is a solution to the problem with discount β and $I^{*\prime}$ is one with β'. Then, optimality implies

$$\beta V(\Delta S + I^*) - C(I^*) \geq \beta V(\Delta S + I^{*\prime}) - C(I^{*\prime}) \qquad (56)$$

$$\beta' V(\Delta S + I^{*\prime}) - C(I^{*\prime}) \geq \beta' V(\Delta S + I^*) - C(I^*) \qquad (57)$$

so that, after some manipulation

$$(C(I^{*\prime}) - C(I^*))(\beta' - \beta) \geq 0 \qquad (58)$$

$$C(I^*) \geq C(I^{*\prime}) \qquad (59)$$

Now, supposing $C(I^*) = C(I^{*\prime})$ for contradiction, we have $V(\Delta S + I^*) = V(\Delta S + I^{*\prime})$ or else one of the objective functions is improvable. Then, from the first order condition for $I^{*\prime}$ we have $\beta' \nabla V(\Delta S + I^{*\prime}) = \nabla C(I^{*\prime})$ and thus $\beta \nabla V(\Delta S + I^{*\prime}) >> \nabla C(I^{*\prime})$. This means that

$I^{*\prime}$ also achieves the supposed maximum in the β problem, but with positive first derivatives in the maximizers and therefore I^* is not a maximizer. Hence, it must be the case that $C(I^*) > C(I^{*\prime})$.

A.5 Proof of Proposition 5

Proof. From (32) and (27) we have that global maximizers must satisfy

$$\beta A_{ii}^{\frac{\sigma}{\sigma-1}}(\Delta_{ii}S_i + I_i^*)^{\frac{1}{\sigma-1}}(V(\Delta S + I^*))^{-\frac{1}{\sigma-1}} = \rho I_i^{*\rho-1} \tag{60}$$

$$\Leftrightarrow \beta^\sigma \rho^{1-\sigma} A_{ii}^\sigma \frac{(\Delta_{ii}S_i + I_i^*)}{I_i^{*(\rho-1)(\sigma-1)}} = \beta V(\Delta S + I^*)) \tag{61}$$

Suppose the global maximum is Π; then for any global maximizer I^*,

$$\beta V(\Delta S + I^{*\prime})) - \sum_j I_j^{*\rho} = \Pi, \tag{62}$$

and thus from (61)

$$\beta^\sigma \rho^{1-\sigma} A_{ii}^\sigma \frac{(\Delta_{ii}S_i + I_i^*)}{I_i^{*(\rho-1)(\sigma-1)}} - \sum_j I_j^{*\rho} = \Pi. \tag{63}$$

As $(\rho-1)(\sigma-1) \geq 1$, the leftmost component is strictly decreasing in I_i^* for each i and therefore injective. Thus two global maximizers with the same total investment costs $\sum_j I_j^{*\rho}$ are identical. Suppose now that two global maximizers I^* and $I^{*\prime}$ exist, and $\sum_j I_j^{*\rho} < \sum_j I_j^{*\prime\rho}$; then from the strict decreasing nature of the left-most term, for each i we have $I_i^* > I_i^{*\prime}$. But this contradicts $\sum_j I_j^{*\rho} < \sum_j I_j^{*\prime\rho}$. Therefore the global maximizer I^* is unique. ∎

A.6 Proof of Proposition 6

Proof. Suppose $(\rho-1)(\sigma-1) > 1$, $\Delta = \delta\mathbb{I}$ and $J_{0i}^*(S)/J_{0j}^*(S) > J_{1i}^*(0)/J_{1j}^*(0)$. First, we find $J_1^*(0)$ from the FOC:

$$I_i^*(0) = \left(\frac{\beta}{\rho}\right)^{\frac{1}{\rho-1}} \frac{A_i^{\frac{\sigma}{(\sigma-1)(\rho-1)-1}}}{\left(\sum_j A_j^{\frac{\rho\sigma}{(\sigma-1)(\rho-1)-1}}\right)^{\frac{1}{(\rho-1)\sigma}}} \tag{64}$$

and then

$$J_{1i}^*(0) = \frac{A_i^{\frac{\rho}{(\rho-1)(\sigma-1)-1}}}{\left(\sum_j A_j^{\frac{\sigma\rho}{(\rho-1)(\sigma-1)-1}}\right)^{\frac{1}{\sigma}}}. \tag{65}$$

From this, $J_{0i}^*(S)/J_{0j}^*(S) > J_{1i}^*(0)/J_{1j}^*(0)$ and the job FOC we have

$$\frac{A_{ii}^{\frac{1}{\sigma-1}} S_i^{\frac{1}{\sigma-1}}}{A_{jj}^{\frac{1}{\sigma-1}} S_j^{\frac{1}{\sigma-1}}} > \frac{A_{ii}^{\frac{\rho}{(\rho-1)(\sigma-1)-1}}}{A_{jj}^{\frac{\rho}{(\rho-1)(\sigma-1)-1}}} \Leftrightarrow \left(\frac{A_i}{A_j}\right)^{\sigma} < \left(\frac{S_i}{S_j}\right)^{(\rho-1)(\sigma-1)-1}. \tag{66}$$

Now, as the solution for $S >> 0$ is interior, recall from (32)

$$\left(\frac{A_i}{A_j}\right)^{\sigma} \left(\frac{\delta S_i + I_i}{\delta S_j + I_j}\right) = \left(\frac{I_i}{I_j}\right)^{(\rho-1)(\sigma-1)}.$$

so that using (66) we obtain

$$\left(\frac{S_i}{S_j}\right)^{(\rho-1)(\sigma-1)-1} \frac{\delta S_i + I_i}{\delta S_j + I_j} > \left(\frac{I_i}{I_j}\right)^{(\rho-1)(\sigma-1)} \tag{67}$$

$$\frac{\delta + \frac{I_i}{S_i}}{\delta + \frac{I_j}{S_j}} > \left(\frac{\frac{I_i}{S_i}}{\frac{I_j}{S_j}}\right)^{(\rho-1)(\sigma-1)} \tag{68}$$

$$\delta \left(\frac{I_i}{S_i}\right)^{-(\rho-1)(\sigma-1)} + \left(\frac{I_i}{S_i}\right)^{1-(\rho-1)(\sigma-1)} > \delta \left(\frac{I_j}{S_j}\right)^{-(\rho-1)(\sigma-1)} + \left(\frac{I_j}{S_j}\right)^{1-(\rho-1)(\sigma-1)} \tag{69}$$

But given that $(\rho - 1)(\sigma - 1) > 1$, the expression $\delta x^{-(\rho-1)(\sigma-1)} + x^{1-(\rho-1)(\sigma-1)}$ is necessarily strictly decreasing in x; therefore (69) implies $\frac{I_i}{S_i} < \frac{I_j}{S_j}$. This in turn implies

$$\frac{S_i}{S_j} > \frac{\delta S_i + I_i}{\delta S_j + I_j} \tag{70}$$

and from this and (28) we have

$$\frac{J_{0i}^*(S)}{J_{0j}^*(S)} = \frac{A_{ii}^{\frac{1}{\sigma-1}} S_i^{\frac{1}{\sigma-1}}}{A_{jj}^{\frac{1}{\sigma-1}} S_j^{\frac{1}{\sigma-1}}} > \frac{A_{ii}^{\frac{1}{\sigma-1}} (\delta S_i + I_i)^{\frac{1}{\sigma-1}}}{A_{jj}^{\frac{1}{\sigma-1}} (\delta S_j + I_j)^{\frac{1}{\sigma-1}}} = \frac{J_{1i}^*(S)}{J_{1j}^*(S)} \tag{71}$$

as required. ∎

A.7 Proof of Proposition 7

Proof. Fix $k > 1$ such that $\left(\frac{A_2}{A_1}\right)^{\frac{\rho}{\rho-1}} \left(1 - k \frac{\rho}{\delta(\rho-1)} \left(\frac{A_2}{A_1}\right)^{\frac{(\sigma-1)(\rho-1)-1}{\rho-1}}\right) > 1$, possible ex hypothesi. Set $c = \left(\frac{\delta A_1}{\rho}\right)^{\frac{\rho}{\rho-1}} k^\rho$ and $S = \left(\left(\frac{A_2}{A_1}\right)^{\sigma} c^{\frac{1}{\rho}}, 0, 0...\right)^T$, First, suppose that the constraint c binds. Consider the problem of allocating c across the different skills; each $i > 1$ is allocated I_i and

31

the remainder goes to skill 1.

$$\max \Pi_{bind} = \max \left([A_1(S_1 + (c - \sum_{j>1} I_j^\rho)^{\frac{1}{\rho}})]^{\frac{\sigma}{\sigma-1}} + \sum_{j>1} (A_j I_j)^{\frac{\sigma}{\sigma-1}} \right)^{\frac{\sigma-1}{\sigma}}$$

then (ignoring the outer power) the first order condition with respect to I_i is

$$\frac{\partial \Pi_{bind}}{\partial I_i} = \frac{\sigma}{\sigma-1} \left(A_i^{\frac{\sigma}{\sigma-1}} I_i^{\frac{1}{\sigma-1}} - A_1^{\frac{\sigma}{\sigma-1}} [S_1 + (c - \sum_{j>1} I_j^\rho)^{\frac{1}{\rho}}]^{\frac{1}{\sigma-1}} (c - \sum_{j>1} I_j^\rho)^{\frac{1-\rho}{\rho}} I_i^{\rho-1} \right)$$

which is 0 for $I_i = 0$. For $I_i > 0$, recalling $(\rho - 1)(\sigma - 1) < 1$,

$$\frac{\partial \Pi_{bind}}{\partial I_i} = \frac{\sigma}{\sigma-1} I_i^{\frac{1}{\sigma-1}} \left(A_i^{\frac{\sigma}{\sigma-1}} - A_1^{\frac{\sigma}{\sigma-1}} [S_1 + (c - \sum_{j>1} I_j^\rho)^{\frac{1}{\rho}}]^{\frac{1}{\sigma-1}} (c - \sum_{j>1} I_j^\rho)^{\frac{1-\rho}{\rho}} I_i^{\frac{(\rho-1)(\sigma-1)-1}{\sigma-1}} \right)$$

$$< \frac{\sigma}{\sigma-1} I_i^{\frac{1}{\sigma-1}} \left(A_i^{\frac{\sigma}{\sigma-1}} - A_1^{\frac{\sigma}{\sigma-1}} S_1^{\frac{1}{\sigma-1}} c^{\frac{1-\rho}{\rho}} c^{\frac{(\rho-1)(\sigma-1)-1}{\sigma-1}} \right) = \frac{\sigma}{\sigma-1} I_i^{\frac{1}{\sigma-1}} \left(A_i^{\frac{\sigma}{\sigma-1}} - A_1^{\frac{\sigma}{\sigma-1}} S_1^{\frac{1}{\sigma-1}} c^{\frac{-1}{\rho(\sigma-1)}} \right) =$$

$$\frac{\sigma}{\sigma-1} I_i^{\frac{1}{\sigma-1}} \left(A_i^{\frac{\sigma}{\sigma-1}} - A_2^{\frac{\sigma}{\sigma-1}} \right) \le 0.$$

Therefore, if the budget constraint is to bind, it must be that it is spent only on skill 1. If the worker invests only in skill 1, she solves

$$\max_{I_1} \left[\delta \left((A_1(S_1 + I_1))^{\frac{\sigma}{\sigma-1}} \right)^{\frac{\sigma-1}{\sigma}} - I_1^\rho \right]$$

so that $I_1^* = \left(\frac{\delta A_1}{\rho} \right)^{\frac{1}{\rho-1}}$ for an expenditure of $\left(\frac{\delta A_1}{\rho} \right)^{\frac{\rho}{\rho-1}} < c$. As a consequence of these two points, the budget constraint c does not bind. To construct a lower bound for the unconstrained worker's payoff, we suppose the worker only invests in skill 2 and suppose the job that only puts weight on skill 2 is chosen. This results in a payoff of $A_2 I_2^* - I_2^{*\rho} = \left(\frac{\delta A_2}{\rho} \right)^{\frac{\rho}{\rho-1}} (\rho - 1)$. We want to show this is greater than the constrained payoff, that is,

$$\left(\frac{\delta A_2}{\rho} \right)^{\frac{\rho}{\rho-1}} (\rho - 1) > A_1 S_1 + \left(\frac{\delta A_1}{\rho} \right)^{\frac{\rho}{\rho-1}} (\rho - 1)$$

which, after rearrangement and division by $(\frac{\delta A_1}{\rho})^{\frac{\rho}{\rho-1}}(\rho - 1)$ becomes

$$
\left(\frac{A_2}{A_1}\right)^{\frac{\rho}{\rho-1}}\left(1 - k\frac{\rho}{\delta(\rho-1)}\left(\frac{A_2}{A_1}\right)^{\frac{(\sigma-1)(\rho-1)-1}{\rho-1}}\right) > 1
$$

which we know to be true; therefore, the unconstrained problem yields strictly higher utility.
∎

A.8 Proof of Proposition 8

Step 1: J is not optimal in period 1 in the absence of a mobility cost.
The contract maximizes

$$
\max\{\max_{\hat{J}}[\hat{J}^T A(S + \delta\Delta S + \delta I^W(I^F))], \max_{\hat{J}}[\hat{J}^T AS] + \delta\max_{\hat{J}}[\hat{J}^T A(\Delta S + I^W(I^F)] - m\}. \tag{72}
$$

As the worker does not move, we have

$$
J = \arg\max_{\hat{J}} \hat{J}^T A(S + \delta\Delta S + \delta I^W(I^F)) \tag{73}
$$

From the fact the worker would move absent the mobility cost, we have

$$
\max_{\hat{J}} \hat{J}^T A(S + \delta\Delta S + \delta I^W(I^F)) < \max_{\hat{J}}[\hat{J}^T AS] + \delta\max_{\hat{J}}[\hat{J}^T A(\Delta S + I^W(I^F))] \tag{74}
$$

Thus S and $\Delta S + I^W(I^F)$ are not parallel. Then from J solving $\nabla F(J) \parallel A(S + \delta\Delta S + \delta I^W(I^F))$ uniquely on $F(J) = 0$, we have that $\nabla F(J) \nparallel A(\Delta S + I^W(I^F))$. And therefore $J \neq \arg\max_{\hat{J}}[\hat{J}^T A(\Delta S + I^W(I^F))]$.

Step 2: No skill is underinvested in, and at least one is overinvested in. Given I_F, the worker chooses I_W to maximize

$$
\max_{I^W} \delta V(\Delta S + I^W) - c(I^W) + c(I^F) \qquad \text{s.t. } I_W \geq I_F \tag{75}
$$

which given the diagonal forms assumption and the exogeneity of I^F is the same problem as

$$
\max_{I^W}\left[\delta\left(\sum_i(A_i(\Delta_i S_i + I_i^W))^{\frac{\sigma-1}{\sigma}}\right)^{\frac{\sigma-1}{\sigma}}\right] - \sum_i I_i^{W\rho} \qquad \text{s.t. } I_W \geq I_F \tag{76}
$$

The first order condition for I_i^W (when the i'th constraint does not bind and thus $I_i^W >$

$I_i^F)$ is

$$\delta A_i^{\frac{\sigma}{\sigma-1}} (\Delta_i S_i + I_i^W)^{\frac{1}{\sigma-1}} \left(\sum_j (A_j (\Delta_j S_i + I_j^W))^{\frac{\sigma}{\sigma-1}} \right)^{\frac{-1}{\sigma}} - \rho I_i^{W\ \rho-1} = 0 \qquad (77)$$

which can be rewritten as

$$\delta A_i^{\sigma} \frac{\Delta_i S_i + I_i^W}{I_i^{W(\rho-1)(\sigma-1)}} = \rho V(\Delta S + I^W) \qquad (78)$$

As $(\rho-1)(\sigma-1) > 1$ by hypothesis, the left hand side is decreasing in I_i^W. Therefore I_i^W is decreasing in V.

If $\frac{dV(\Delta S + I^W)}{dI_j^F} \leq 0$ then for any i such that $I_i^W(I^F) > I_i^F$, we would have $\frac{dI_i^W(I^F)}{dI_i^F} > 0$ and thus as $V(\Delta S + I^W) = (\sum_k (A_k (\Delta S + I^W))^{\frac{\sigma}{\sigma-1}})^{\frac{\sigma-1}{\sigma}}$, V would have to increase, a contradiction. Thus, we have shown that if $I_i^W(I^F) > I_i^F$ and $I_j^W(I^F) = I_j^F$ then $dI_i^W(I^F)/dI_j^F < 0$ $(\star)$.

Ex ante, the contract maximizes

$$\max_{I^F} \delta J^T A(\Delta S + I^W(I^F)) - C(I^W(I^F)) \qquad (79)$$

If for all i we have $I_i^W(I^F) \leq I_i^*(J)$ then $J^*(\Delta S + I^*(J)) \leq J$ and therefore either J is in the interior of $\mathcal{J}$ and therefore a job where the worker is more productive both periods exists (a contradiction) or J is optimal in the second period in the absence of a moving cost, not the case by Step 1. Therefore there is an i for which $I_i^W(I^F) > I_i^*(J)$.

Suppose there is an i for which $I_i^W(I^F) < I_i^*(J)$. Then setting $\forall j, I_j^{F\prime} := \max\{I_j^W, I_j^*(J)\}$, from $\star$ we have that $I^W(I^{F\prime}) = I^{F\prime}$. Therefore, $I^{F\prime}$ improves the objective (79), a contradiction. So $I^W(I^F) \gneq I^*(J)$.

Step 3: Every skill is overinvested in.

Now suppose $\exists i : I_i^W(I^F) = I_i^*(J)$. As $I^W \gneq I^*(J)$, there must be a j such that $I_j^W(I^F) > I_j^*(J)$. Two cases are of interest.

Case 1. Suppose (78) holds for I_j^W. Then, define $\bar{I}^F = I^W(I^F)$; we have $I^W(\bar{I}^F) = I^W(I^F)$ and, as I^F is part of an optimal contract, so is $\bar{I}^F$ (albeit with a compensating period-0 wage). We will consider increasing $\hat{I}_i^F$ to effect an increase in V and through it will implement a decrease in $\hat{I}_j^F$ without affecting $(\bar{I}_k)_{k \notin \{i,j\}}$.

We define the auxiliary function $\hat{I}_j^F(\hat{I}_i^F)$ implicitly by

$$\delta A_j^\sigma \frac{\Delta_j S_j + \hat{I}_j^F}{\hat{I}_j^{F(\rho-1)(\sigma-1)}} = \rho V\left((\Delta_i S_i + \hat{I}_i^F), (\Delta_j S_j + \hat{I}_j^F), (\Delta_k S_k + \bar{I}_k^F)_{k\notin\{i,j\}}\right) \tag{80}$$

As $S \gg 0$, we have $\partial V/\partial \hat{I}_i^F > 0$ and $\partial V/\partial \hat{I}_j^F > 0$; furthermore, the left hand side is decreasing in $\hat{I}_j^F$ as $(\rho-1)(\sigma-1) > 1$. Therefore, we have $d\hat{I}_j^F(\hat{I}_i^F)/d\hat{I}_i^F < 0$.

Consider now perturbing the optimal contract $\bar{I}^F$ by increasing $\bar{I}_i^F$ and lowering $\bar{I}_j^F$ along $\hat{I}_j^F(\cdot)$. As $V\left((\Delta_i S_i + \hat{I}_i^F), (\Delta_j S_j + \hat{I}_j^F), (\Delta_k S_k + \bar{I}_k^F)_{k\notin\{i,j\}}\right) \geq V(\bar{I}^F)$ when $\hat{I}_i^F \geq \bar{I}_i^F$, we have that $I^W(\hat{I}_i^F, \hat{I}_j^F(\hat{I}_i^F), (\bar{I}_k)_{k\notin\{i,j\}}) = (\hat{I}_i^F, \hat{I}_j^F(\hat{I}_i^F), (\bar{I}_k)_{k\notin\{i,j\}})$.

Written solely in terms of $\hat{I}_i^F$ (and keeping constant in skills other than i and j), the objective function is

$$\delta J^T A((\Delta_i S_i + \hat{I}_i^F), (\Delta_j S_j + \bar{I}_j^F(\hat{I}_i^F)), (\Delta_k S_k + \bar{I}_k^F)_{k\notin\{i,j\}}) - C((\hat{I}_i^F, \hat{I}_j^F(\hat{I}_i^F), (\bar{I}_k)_{k\notin\{i,j\}})) \tag{81}$$

and is ex hypothesi maximized at $\hat{I}_i^F = \bar{I}_i^F$. Taking a right derivative of the objective with respect to $\hat{I}_i^F$ we get

$$\delta A_i J_i + \delta A_j J_j \frac{d\hat{I}_j^F(\hat{I}_i^F)}{d\hat{I}_i^F}) - \rho\hat{I}_i^{F\rho-1} - \frac{d\hat{I}_j^F(\hat{I}_i^F)}{d\hat{I}_i^F}\rho(\hat{I}_j^F(\hat{I}_i^F))^{\rho-1} \tag{82}$$

$$= (\delta A_i J_i - \rho\bar{I}_i^{F\rho-1}) + \frac{d\hat{I}_j^F(\hat{I}_i^F)}{d\hat{I}_i^F}(\delta A_j J_j - \rho\hat{I}_j^F(\bar{I}_i^F)^{\rho-1}) \tag{83}$$

But by assumption $\bar{I}_i^F = I_i^*(J)$, so that $\delta A_i J_i - \rho\bar{I}_i^{F\rho-1} = 0$ and $\hat{I}_j^F(\bar{I}_i^F) = \bar{I}_j^F > I_j^*(J)$ so that $\delta A_j J_j - \rho\hat{I}_j^F(\bar{I}_i^F)^{\rho-1} < 0$. Furthermore, we have that $\frac{d\hat{I}_j^F(\hat{I}_i^F)}{d\hat{I}_i^F} < 0$. Therefore evaluated at $\bar{I}_i^F$, the restricted objective function's right derivative is positive. As a result, there exists a $\hat{I}_i^F > \bar{I}_i^F$ so that $(\hat{I}_i^F, \hat{I}_j^F(\hat{I}_i^F), (\bar{I}_k)_{k\notin\{i,j\}})$ improves the objective function over the assumed maximizer I^F, a contradiction.

Case 2. Now suppose instead that (78) does not hold for any I_j^W. Consider again $\bar{I}^F = I^W(I^F)$, which is again optimal under the hypothesis that I^F is. Define $\hat{I}_j^F(\hat{I}_i^F)$ by $V\left((\Delta_i S_i + \hat{I}_i^F), (\Delta_j S_j + \hat{I}_j^F), (\Delta_k S_k + \bar{I}_k^F)_{k\notin\{i,j\}}\right) = V(\Delta S + \bar{I}^F)$ when $\hat{I}_i^F \geq \bar{I}_i^F$ is small enough for a solution to exist. In other words, $\hat{I}_j^F$ adjusts to $\hat{I}_i^F$ so as to keep production at the outside option job constant.

Then as V is constant along $(\hat{I}_i^F, \hat{I}_j^F(\hat{I}_i^F))$ for $\hat{I}_i^F > \bar{I}_i^F$, skills $k \notin \{i,j\}$ stay constant. As

V has strictly positive (as $S >> 0$) and continuous partials, $d\hat{I}_j^F(\hat{I}_i^F)/d\hat{I}_i^F > 0$; the rest of the argument follows as in Case 1.