Ben-Porath meets Lazear: Microfoundations for Dynamic Skill Formation*

Costas Cavounidis† and Kevin Lang‡

January 23, 2019

Abstract

We provide microfoundations for dynamic skill formation with a model of investment in multiple skills, when jobs place different weights on skills. We show that credit constraints may affect investment even when workers do not exhaust their credit. Firms may invest in their workers' skills even when there are many similar competitors. Firm and worker incentives can lead to overinvestment. Optimal skill accumulation resembles, but is not, learning by doing. An example shows that shocks to skill productivity benefiting new workers but lowering one skill's value may adversely affect even relatively young workers, and adjustment may be discontinuous in age.

1 Introduction

We develop a tractable two-period model of dynamic skill formation with investment in multiple skills and heterogeneous jobs. Our model draws heavily on the insights of Lazear (2009) in viewing jobs as putting linear weights on skills. We place this in the context of lifetime investment in the spirit of Ben-Porath (1967).

Our approach makes several contributions. First, it offers a more realistic treatment of the varieties of human capital and the jobs that use them. We depart from treating manual/routine/abstract work as a hierarchy, and thus can account for occupations such as veterinarian and kindergarten teacher that combine manual with abstract tasks. Second,

*This research was funded in part by NSF grant SES-1260917. We are grateful to Arun Advani, Dan Bernhardt, Mirko Draca, Ed Lazear, Paul Oyer, Pascual Restrepo and participants in CRETE and the economic theory workshop at Boston University for helpful comments. The usual caveat applies.

†Department of Economics, University of Warwick. email: c.cavounidis@warwick.ac.uk

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

the model explains the differential career paths of workers by skills and age. In particular, we address the patterns of displacement caused by technological shocks. Third, the model provides new insights into the financing of human capital acquisition and the distortions its imperfections may create, to which we now turn.

The model provides a novel explanation for why firms help workers invest in general skills and why workers frequently invest in skills that they do not intend to use much,[1] with markedly different implications from those of Acemoglu and Pischke (1999a&b). We do this by introducing wage bargaining and mobility frictions, so that workers have an incentive to invest for their best outside option in order to raise their wage at their current job. If firms can offer contracts that guarantee financing of some human capital investment, even if workers' own investments cannot be contracted, this inefficiency can be blunted. This happens in a somewhat peculiar way. Optimal contracts commit the firm to overinvest in skills useful at the current job, relative to the first best, in order to temper workers' incentive to overinvest in other skills.

Our model can also explain why, even when credit constraints appear to be unimportant, some workers do not undertake what appear to be highly profitable investments. In their excellent review of credit constraints in education, Lochner and Monge-Naranjo (2012) discuss how when investment is constrained by borrowing limits, individuals receive a marginal return on investment that exceeds the gross interest rate and borrow up to their credit limit. This result is correct when earnings are a concave function of (homogeneous) human capital or, more precisely, when there is a decreasing marginal return to spending, broadly defined, on schooling. However, in our model the complementarity between occupation choice and skill choice can create a natural local convexity in the returns to spending on training in some skill. Then, when a borrowing limit prohibits undertaking the unconstrained investment, the worker may prefer a much smaller investment such that she borrows strictly less than her borrowing limit. The worker will appear unconstrained - the borrowing constraint will not bind with equality - but the constraint will nevertheless affect the outcome.

But are such constraints common? As an example, we note that apprenticeships in painting are typically two or three years. Plumbing apprenticeships are generally four or five years and typically require coursework. For instance, in Montana, there is no license for

[1] We provide two anecdotes. While working in a very demanding job as Director of Player Development for the San Diego Padres, Theo Epstein, currently President of Basbeall Operations for the Chicago Cubs, earned a law degree. While contract law and labor law were arguably relevant to his job, the full degree included courses in criminal and constitutional law, skills unlikely to be useful in player development or baseball management, more generally. Marie (a real person, known to one of us), while working in a demanding advertising job, pursued a BA through the adult education section of a local university even though she was unlikely ever to work in a job for which it was important for her to have a degree.

painters.[2] In contrast, to become a journeyman plumber in Montana, a worker can complete the Associate of Applied Science degree in Plumbing Technology at Montana State University - Northern, for which tuition and fees over the two years are approximately $23,000. After degree completion, apprenticeships last three years.[3] Under plausible assumptions, the internal rate of return on this investment relative to becoming a painter is in excess of 11 percent.[4] Without question, there are many reasons that a worker might choose to become a painter rather than a plumber, but at least one plausible explanation is that potential plumbers face credit constraints.

How is this consistent with the consensus that credit constraints were relatively unimportant for determining schooling levels in the United States in the late 1970s and early 1980s (Lochner and Monge-Naranjo 2012), and possibly even for the 1960s (Lang and Ruud 1986)? Perhaps credit constraints were unimportant then. But the earnings gap between plumbers and painters was, if anything, higher in that period (Mellor 1985). Our model suggests that the discrepancy between the high returns to investing in plumbing skills and the apparent absence of credit constraints is due to the fact that job choice can lead to non-concave returns to investment, and that standard methods treat those for whom constraints do not bind with equality as unconstrained.[5]

We show conditions under which investment is discontinuous in the age of the worker, but the full import of this finding is better understood in terms of workers' responses to unanticipated changes in skill prices. Therefore, we investigate a continuous-time version of the model, albeit in a world with no mobility costs or credit constraints. We provide a simple example with two skills in which workers with less than roughly four years of experience move swiftly to acquire the skill that has become more valuable while more experienced workers, who are already heavily invested in the skill that has become less valuable, do not.

We then use the continuous-time model to explore the effects of asymmetric shocks to the production technology on the careers and earnings of workers of different ages. Two effects

[2]https://paintingleads.com/painting-contractor-license-requirements/ accessed 11/18/2017.

[3]Information accessed on the Montana State University - Northern website 11/17/2017.

[4]According to payscale.com (accessed 11/18/2017), the average painter makes $36,000 per year, compared with the average apprentice plumber who makes $30,000. But the apprentice plumber is on her way to being a journeyman plumber for whom average annual earnings are $52,000, and potentially even more as a master plumber. These numbers are consistent with Bureau of Labor Statistics averages. If we assume that the painter apprentices earn $30,000 per year for two years and then earn the average of $36,000 every year for 43 years and if we further assume that the plumber pays tuition over two years and earns nothing, then earns the average apprentice salary of $30,000 for three years before earning the average of $52,000 for journeymen plumbers for 40 years, the internal rate of return is just over 11 percent. We view these assumptions as optimistic for painters and pessimistic for plumbers.

[5]Bernhardt and Backus's (1990) model also predicts that credit-constrained workers will choose different jobs - those with a flatter wage profile. Their result is based on consumption-smoothing rather than human capital investment.

lead older workers to adjust less to shocks than young ones:

1. Horizon: The time over which to exploit new skills is declining in age. This makes new skills less valuable to older workers.

2. Inertia: Mature workers are more skilled and, possibly, more specialized than fresh ones. Therefore, acquiring new skills leads to a smaller shift in jobs. This makes skill adjustment less effective.

These two effects combine so that workers respond to shocks to the value of skills in predictable ways. Autor and Dorn (2009) show that as employment in routine cognitive jobs declined, the proportion of older workers in these jobs also grew. This reflects some combination of reduced inflows and increased outflows of younger workers. The data used in that paper do not allow them to estimate the relative importance of these two effects. In Appendix B, we provide some small supplementary evidence using panel data that the desktop computer revolution caused younger workers, but not older workers, to exit routine-cognitive intensive jobs.

Of course, it is unsurprising that older workers are less likely to leave adversely shocked jobs or to move to positively shocked jobs. This will happen in any model in which mobility is a costly investment and thus has a horizon effect. But tautologically a single skill model cannot explain movement to a job that is intensive in a different skill. We expect, and our model implies, that how the worker invests in response to a negative shock depends on her previously accumulated skills (inertia).[6] Single-skill models might predict that when scanner technology reduces the value of skill in cashiering, workers may move away from such jobs. Our model helps explain who remains a cashier, who trains to be an electrician and who trains to be a store manager.

We show by way of an example that since investment is front-loaded, the inertia effect can ‘kick in’ early. In fact, due to this front-loading, shocks beneficial to the youngest workers can be strongly adverse to workers only slightly older. This is potentially important for both positive and normative reasons. On the one hand, it helps explain the disaffection and opposition to free trade among relatively young workers whom we might expect to be able to adapt to and benefit from technology shocks that make otherwise similar new entrants better off. On the other, this finding suggests that if it wishes to address the adverse effects of trade

[6]We also observe this inertia effect outside of the mainstream labor market. Sviatschi (2018) shows early experience in the illegal cocoa industry increases later activity in the illegal drug trade by more than predicted by reduced school attendance. A single-skill model does not predict this. Rather children build on their parents’ early investment in their knowledge of the industry as our model predicts. Similarly Arora (2018) finds, consistent with a model of heterogeneous skills, that decreasing punishment for crimes committed at a relatively young age increases criminality at older ages at which punishment is not reduced.

or technology shocks, government must act swiftly to subsidize retraining, even before there is large-scale displacement.

We are certainly not the first to address the dynamics of human capital investment. 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. Prada and Urzua (2017) find that also accounting for mechanical ability greatly affects how we should think about investment in education. Bowlus, Mori and Robinson (2016) explore how skill use evolves over the life-cycle. Sanders and Taber's (2012) two-period model is closest to ours but follows Lazear in assuming that the worker will randomly meet only one firm. Altonji (2010) emphasizes the growing need for a research agenda that recognizes that skill is multidimensional rather than hierarchical, and that jobs differ in their requirements.

2 The Two-Period Model[7]

We will now turn to the tractable, two-period model for which we prove the results outlined in the introduction. For a more general model, and a discussion of which results hold in that broader setting, we refer the interested reader to this paper's online appendix.

There exist N different skills. The worker begins period 1 endowed with a vector of skill levels $S \in \mathbb{R}_+^N$. 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 is a vector $J \in R_+^N$ of weights applied to the worker's skills. At job J, a worker with skills S produces

$$\Sigma_n A_n J_n S_n \tag{1}$$

where $A >> 0$. The worker chooses a job $J \in \mathbb{R}_+^N$ which must satisfy

$$\Sigma_n J_n^{\sigma} \leq 1 \tag{2}$$

where $\sigma > 1$. A is the productive efficiency of different skills, representing how the current technology uses each skill. A_n has the useful interpretation that it is the maximum weight on skill n in any job: a job that only uses skill n puts weight of A_n on it. This allows us to

[7] Throughout this paper we use vector inequality notation as follows: $x >> y \Leftrightarrow \forall n\ x_n > y_n$; $x > y \Leftrightarrow \forall n\ x_n \geq y_n$ and $x \neq y$; $x \geq y \Leftrightarrow \forall n\ x_n \geq y_n$.

discuss technology in terms of changes in A while holding the set of jobs constant.

Differences in A can capture differences in the value of a skill over time. Knowing how to shoe horses is a skill that has declined in value even though the job persists. We can also capture differences across workers in their capacity to use (or equivalently, later in the paper, learn) that skill. An individual who is 'good at math' has a higher A_{math} and will turn the same amount of math education into a higher level of production in math-intensive jobs.

Workers will choose their job to maximize production (or wages) given S. Optimal job choice implies a job J such that $\Sigma_n J_n^\sigma = 1$. With only two skills, when $\sigma = 2$, the northeast boundary of the set of available jobs is the unit quarter circle in the positive quadrant; there, a boundary job using both skills equally puts a weight of $A_n\sqrt{.5}$ 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.

The worker chooses J to maximize the Lagrangian

$$\Sigma_n A_n J_n S_n - \lambda \left(\Sigma_n J_n^\sigma - 1 \right). \tag{3}$$

Maximization gives the first-order conditions

$$A_n S_n = \lambda \sigma J_n^{\sigma-1} \tag{4}$$

along with the constraint. Solving, we have

$$J_n = \frac{(A_n S_n)^{\frac{1}{\sigma-1}}}{\left(\sum_n [A_n S_n]^{\frac{\sigma}{\sigma-1}}\right)^{\frac{1}{\sigma}}}. \tag{5}$$

Note that $dJ_n/dS_n \geq 0$; the higher a worker's skill, the more weight the job she chooses puts on it. Workers naturally choose jobs they're good at.

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

$$V(S) = \max_{J \in \mathbb{R}_+^N : \Sigma J_n^\sigma \leq 1} \sum_n A_n J_n S_n = \left(\sum_n [A_n S_n]^{\frac{\sigma}{\sigma-1}} \right)^{\frac{\sigma-1}{\sigma}}. \tag{6}$$

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.

We now augment the model with a second productive period, and allow the worker to invest in skills following production in period 1, increasing them by I at cost

$$C(I) = \Sigma_n I_n^\rho \tag{7}$$

with $\rho > 1$. Note that since skill typically has no natural scale, we can normalize the coefficients on I_n rather than writing $\psi_n I_n^\rho$. Of course, this normalization will affect S_n and A_n, 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

$$\underbrace{\Sigma_n A_n J_{1,n} S_n}_{\text{Period 1 Production}} + \beta \underbrace{\Sigma_n A_n J_{2,n} (\delta S_n + I_n)}_{\text{Period 2 Production}} - \underbrace{\Sigma_n I_n^\rho}_{\text{Investment Cost}} \underbrace{-\lambda \left(\Sigma_n J_{1,n}^\sigma - 1\right) - \mu \left(\Sigma_n J_{2,n}^\sigma - 1\right)}_{\text{Job Choice Constraints}}, \tag{8}$$

where J_1 and J_2 are the jobs in periods 1 and 2, β is the discount factor and δ is the rate at which skills do not depreciate (1 minus the depreciation rate) between periods 1 and 2.

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

$$\max_{I \in \mathbb{R}_+^N} \left[\beta \left(\Sigma [A_n(\delta S_n + I_n)]^{\frac{\sigma}{\sigma-1}} \right)^{\frac{\sigma-1}{\sigma}} - \Sigma I_n^\rho \right], \tag{9}$$

which yields the first-order conditions for I:

$$\beta \frac{A_n^{\frac{\sigma}{\sigma-1}} (\delta S_n + I_n)^{\frac{1}{\sigma-1}}}{\left(\Sigma_m [A_m(\delta S_m + I_m)]^{\frac{\sigma}{\sigma-1}} \right)^{\frac{1}{\sigma}}} - \rho I_n^{\rho-1} = 0 \tag{10}$$

and arrive at

$$\left(\frac{A_n}{A_m}\right)^\sigma \frac{\delta S_n + I_n}{\delta S_m + I_m} = \left(\frac{I_n}{I_m}\right)^{(\rho-1)(\sigma-1)}. \tag{11}$$

The solution is guaranteed to be unique when $(\sigma-1)(\rho-1) > 1$ and $S \neq 0$. The existence of a corner solution with $I_n = 0$ requires $S_n = 0$ and that $(\rho-1)(\sigma-1) \leq 1$.[8]

[8] To see this, notice that when $S_n = 0$, the left hand side of (11) goes to 0 more slowly than the right hand side as $I_n \downarrow 0$ when $(\rho-1)(\sigma-1) > 1$. That the marginal cost of investment is 0 at 0 drives the fact that workers invest in all skills they are endowed with - without such an assumption, some of our results would have to be qualified in ways that are not easily stated in terms of primitives.

Note that the standard presentation of the Ben-Porath model assumes that investment takes the form of foregone production while we treat investment as a cost. This distinction is largely a matter of convenience. Someone who is capable of earning $\$x$ and chooses to invest $\$y$ foregoes a proportion y/x of her income. Since tautologically, in the Ben-Porath model, post-schooling workers never devote their entire potential income to human capital investment, with a single skill, our model resembles a Ben-Porath model of post-schooling investment.

2.1 Implications

Although it makes intuitive sense, no existing model is designed to explain why older workers who have invested more heavily in a skill respond less to a shock to the value of skills. We will address this point more directly in the continuous-time model. Here we lay the groundwork by showing both inertia in investment and the effect of a shorter horizon on investment.

Proposition 1 *Investment in skill n is increasing in the endowment of skill n. That is, if $S'_n > S_n$ and for $m \neq n$ we have $S'_m = S_m$, and there is an $m \neq n$ s.t. $S_m \neq 0$, then $I^{*\prime}_n > I^*_n$. Furthermore, the period-2 weight on skill n is increasing in the endowment of skill n; $J^{*\prime}_{2,n} > J^*_{2,n}$.*

Proof. See appendix for all proofs. ■

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 she currently uses. An alignment of incentives causes workers with a higher stock of a skill to both (i) choose initial jobs where they use more of that skill and (ii) invest more in that skill. 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.*[9] A similar alignment of incentives operates in Jovanovic (1979), where a worker with a higher level of firm-specific human capital invests more in firm-specific human capital as finding a better match is unlikely, making the worker more likely to continue to use it.

With heterogeneous skill depreciation rates, Proposition 1 is easily extendable to also imply that workers invest less in skills that depreciate more rapidly. Note that this occurs

[9] Note that the proposition does not say that workers always specialize (see the working paper for conditions under which workers will tend to specialize and when they instead converge towards the same job). What the proposition says is that if two otherwise identical workers enter the market but one has more of skill n than the other, that worker will invest more in skill n even if both will invest more heavily in some other skill. More generally, in an infinite period model, the first worker will invest more in n in every period even if in infinite time they would end up at the same job.

even though the investment itself does not depreciate. Instead, because their initial skill depreciates, workers will optimally choose a job that makes less use of it; as a consequence, they invest less in the skill. In addition, although the model does not explicitly account for age, we can alter β to change the 'length' of the second period. Increasing β (i.e. raising the remaining lifetime), proportionately raises the value of a given endowment and investment. Total expenditure on investment increases with the remaining time.

Proposition 2 *Let $\beta > \beta'$, and let 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 and as in the Ben-Porath model. However, this result only addresses total investment costs. It does not necessarily mean that investment in any particular individual skill will decrease with age.

Combining propositions 1 and 2 suggests an intuitively appealing explanation for why older workers are less likely to shift occupations in response to a skill shock. They both invest less overall in additional skills due to a (shortened) horizon effect and are more heavily invested in an existing stock of skills, the inertia effect. Therefore, their stock of skills after investing more closely resembles their initial stock of skills, and they will tend to remain in similar jobs. We address this point more fully in continuous time.

So far, all results carry over to the general two-period model in the online appendix. For the model used here, we can do more: we find mild sufficient conditions for the optimal investment and its cost to be discontinuous in β.

Proposition 3 *If $(\rho-1)(\sigma-1) < 1$ and the productivity of skills is not identical, there is an endowment such that investment cannot be continuous in β. That is, if $(\rho-1)(\sigma-1) < 1$ and $\min_n A_n < \max_n A_n$ there is an endowment S such that, if $\mathcal{I}^*(\cdot)$ maps each β into the set of optimal investments for that β, then $\mathcal{I}^*$ does not admit a continuous singlevalued selection.*[10]

Informally, β captures the length of the second period relative to the first or, even more informally, the worker's age. When the worker is old, there is little investment and the worker's second-period job choice and first-period job choice are relatively similar. When the worker is young, it makes sense to shift more towards jobs that are intensive in a highly valued skill even if the worker was initially endowed with more of a low-value skill. The shift between the two is discontinuous in β, and thus informally the worker's age, when $(\rho-1)(\sigma-1) < 1$, as this condition ensures that a varied portfolio is both expensive and not that productive. We expand on this point in the discussion of the continuous-time model.

[10] A singlevalued selection from a correspondence $F : X \rightrightarrows Y$ is a function $f : X \to Y$ such that $f(x) = y \implies y \in F(x)$.

3 Extensions of the Two-Period Model

3.1 Credit Constraints

With a single skill, we could fully replicate a Ben-Porath model by imposing the additional constraint that investment be (weakly) less than production, which would only slightly complicate the model. As in Ben-Porath, we could then have workers (for certain endowments) fully specialize in investment in the first period. In this section we show that, with multiple skills, a credit constraint can work very differently than it does in the Ben-Porath model. In particular, workers' behavior may be influenced by credit constraints even when the constraint does not bind with equality in equilibrium.

Recall that because a worker with more skill n chooses a job that is more intensive in n, the skill value function, V, is weakly convex[11] even though each individual job's production function is linear. Put differently, there can be (locally) increasing returns to investment in a skill such that a large investment may be worthwhile even if a small investment is not. This can make the investment problem non-convex, and therefore produce solutions affected by the constraint, but without the constraint binding with equality. This corresponds to our example in the introduction, where unconstrained workers choose to be plumbers, but constrained workers choose to be painters, in which case they do not borrow up to their borrowing limit. We derive sufficient conditions for this to be true.

If $(\rho-1)(\sigma-1) < 1$, workers are driven to specialize, but their specialization may depend on their ability to invest. As a result, if $(\rho-1)(\sigma-1) < 1$ and skill productivity is sufficiently varied, there is an endowment and a constraint such that the constraint affects the outcome, but does not bind with equality.

Proposition 4 *Assume $(\rho-1)(\sigma-1) < 1$. There exists an $x > 0$ such that if for some n, m we have $A_n \cdot x < A_m$, then there exist S, c such that if I^* solves the unconstrained investment problem with endowment S and I_c^* solves the investment problem with endowment S and constraint c, then $C(I^*) > c$ and $C(I_c^*) < c$.*

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 its productivity is mediocre, so investment is constant for an intermediate constraint. However, once the constraint is greater than 1 the worker specializes heavily in skill 2, and the constraint once again binds

[11] With the convexity inequality strict for all but parallel skill vectors.

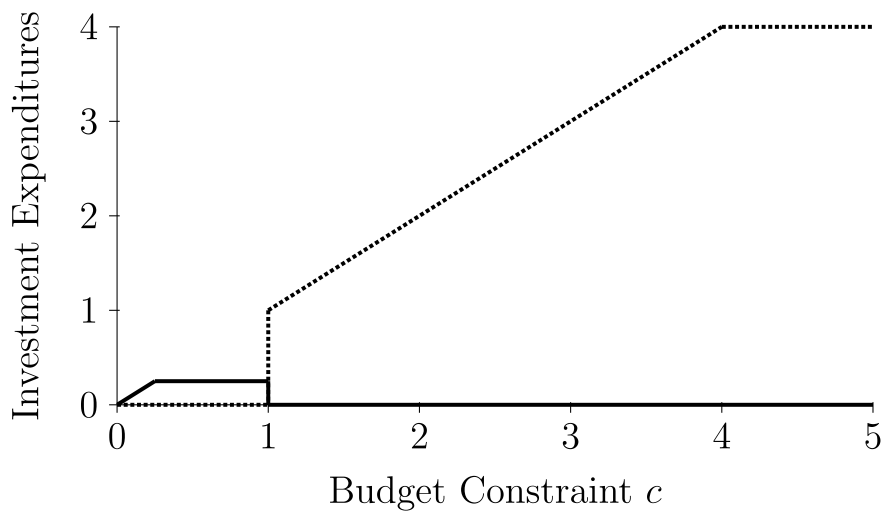

Figure 1: Investment expenditure in skill 1 (solid line) and 2 (dotted line) as a function of the budget constraint c when $A_1 = 1, A_2 = 4, \beta = 1, \delta S_1 = 2.75, S_2 = 0, \rho = 2, \sigma \approx 1$.

until $c = 4$. Beyond that, further investment is inefficient, and relaxing the constraint further has no effect on net output.

This result has a sensible explanation: even though production at any given job is concave in investment expenditure, the ability to choose jobs may make a worker's market production non-concave in investment expenditure. Therefore, constraints on investment expenditure can influence a worker's career without binding with equality. Intuitively, a plumber who would make a great doctor might not be able to afford the entirety of medical school. If this worker spends her budget on the half year of schooling she can afford, she will still effectively be a plumber with some medical skills; given that the return to medical skills is low in plumbing, this investment is not worthwhile.

We note that, in principle, it is possible to test for the existence of the type of credit constraint that our model highlights. In standard models, relaxing the budget constraint (increasing c in terms of the model) should have no effect on borrowing for anyone not borrowing up to the original credit limit. In standard models, if we present two samples with different credit limits $c_1 < c_2$, the cumulative density functions of the amount borrowed should, up to sampling error, be identical at levels of borrowing below c_1. In our model, some workers who would borrow strictly less than c_1 when that is the constraint, borrow more than c_1 when the constraint is relaxed. Thus, our model predicts that the cdf of borrowing under the higher constraint should merely first-order stochastically dominate the cdf of borrowing under the lower constraint. Experiments using different constraints could be conducted in a variety of settings to determine the importance of the phenomenon we discuss in the introduction.

3.2 Mobility Costs and Overinvestment

Suppose now that the worker's job choice problem is one of accepting an offer by a firm. If a multitude of firms (at least one at each job) offer wages prior to each period under perfect information about the worker's skills, the usual solution holds. But if there is a mobility cost m to be paid by the worker if she moves jobs between periods 1 and 2, things are different. The period-1 or *incumbent* firm can retain the worker by offering her the highest wage offered elsewhere, minus m, and will do so if this difference is less than the worker's productivity at the incumbent. The incumbent firm has local monopsony power. The monopsony rents will be distributed back to the worker as part of the period-1 wage, but there is still inefficiency.

In such a situation, if the worker does not move between periods, her investment decision is distorted. Instead of optimizing her net productivity at the job she actually will do, she maximizes the highest outside wage offer she receives - her outside option - minus the investment cost.[12]

This inefficiency can be ameliorated if firms have the ability to include training as part of the first-period offer: an investment in the worker's skills that the firm pays for. A first-period offer is therefore a triple (J, I^F, w) where J is the job, I^F is the firm's investment and w is the wage. The worker will then be able to augment this investment to $I^W \geq I^F$, paying the difference $C(I^W) - C(I^F)$. The worker's investment remains non-contractual: only the firm will be bound by the first-period offer to invest in a particular way. It is perhaps unsurprising that adding an additional dimension to offers improves efficiency, but the way in which this is accomplished reveals much about firms' incentives to manipulate human capital formation.

Denote the worker's investment best response function[13] mapping the firm's investment commitment to total investment by $I^W(\cdot)$, and by $I^*(J)$ the efficient investment for job J.

Proposition 5 *Assume $(\sigma - 1)(\rho - 1) > 1$. Suppose the worker with skill endowment $S >> 0$ stays in job J in both periods, but would not absent the mobility cost. Then the accepted first-period offer (J, I^F, w) 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

[12] This is akin to Konrad and Lommerud (2000) in which family members have an incentive to inefficiently over-invest in market human capital prior to intra-family bargaining in order to improve their outside options.

[13] Sometimes this will be a correspondence; if $(\rho - 1)(\sigma - 1) > 1$, it is guaranteed to be a function.

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. The reason for this is that at the appropriate level of investment in some skill for the current job, the direct effect of further investment in that skill on net production is only a second-order loss; whereas the efficiency gain from disincentivizing the worker's excessive investment elsewhere is a first-order effect.

The condition that $(\rho - 1)(\sigma - 1) > 1$ is sufficient, but far from necessary. It is not difficult to develop examples in which this condition is violated, but the result goes through. The essential requirements for the result to hold are that the worker treats investments in different skills as substitutes, and that the worker's return to investment expenditure in any skill is concave on a relevant set.

It is interesting to contrast our framework with the closely related 'Full-Competition' regime in Acemoglu and Pischke (1999b). In their model, as in ours, the firm can commit to a level of investment but cannot commit to a second-period wage. As a consequence, if, for any of a variety of reasons, the worker cannot recoup the value of any investment, the firm will commit to the investment. If, as in our model, there is no mobility, the firm will commit to the optimal level of investment. Because Acemoglu and Pischke model a world with only a homogeneous skill, outside opportunities that are never used in equilibrium do not distort investment. Distortion occurs in our model because there are multiple skills.

This model has a surprising testable implication. Suppose we observe a worker employed at a firm and offer free general training, and that this intervention takes place after the worker has joined the firm and that the new investment is in excess of I^W. If the worker can decide on the type of training received, her incentive is to increase her productivity at her best outside option. This increases the probability of mobility: if she would have moved previously, she will continue to move, and if she would not have, the increased investment may be sufficient to make moving optimal for her. In contrast, if the firm chooses the character of the training, its incentive is to use the additional investment to increase the worker's productivity at the incumbent firm relative to the best outside option. This may be sufficient to entice the worker to stay when she otherwise would not.

In a standard model with a single skill and no mobility, either party would use the investment to maximize the worker's productivity. In a Roy-style model with costly mobility, although the worker's choice of training could increase mobility, the firm could at best under-provide training to keep mobility constant; only in a multi-skilled model can the firm use general training to decrease mobility.

4 Continuous Time

To explore the dynamics of skill investment more fully, we port the model to continuous time. Unlike the two-period model, the continuous-time model does not lend itself to analytic results similar to the ones in previous sections.[14] However, its rich dynamics will allow us to simulate workers' adjustment to technological changes in production in a far more realistic way. Specifically, we will be able to investigate the effects of such shocks on the skill investment, career, and earning paths of workers of different ages and skill endowments.

4.1 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 vector is A and the worker chooses jobs $J(t)$ from the job set $\mathcal{J} = \{J \in \mathbb{R}_+^N | \sum_n J_n^\sigma \leq 1\}$ so that her time-t instantaneous production is $\Sigma_n A_n J_n(t) S_n(t)$. Skills depreciate at rate Δ, counterbalanced by investment $I(t)$, so that

$$\frac{d}{dt}S(t) = -\Delta S(t) + I(t). \tag{12}$$

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

$$\max_{J:[0,T]\to\mathcal{J},\ I:[0,T]\to\mathbb{R}_+^N} \int_0^\infty e^{-rt}\left[\Sigma_n A_n J_n(t) S_n(t) - \sum_n I_n(t)^\rho\right] dt \tag{13}$$

$$s.t.\ \frac{d}{dt}S(t) = -\Delta S(t) + I(t) \tag{14}$$

$$S(0) = S_0. \tag{15}$$

The worker chooses $I(t)$ and $J(t)$ optimally. However, as J does not influence the state variable, it is chosen according to (6). Thus, we can bypass job selection for the moment and reduce the problem to

[14] Although we do not prove these results analytically, many insights from the two-period model extend to this model. Since the convexity that drives the credit constraint result in the two-period model is present in continuous time, it is straightforward to provide examples in which credit constraints are important even though the worker does not borrow up to the constraint. Providing an example of overinvestment in the continuous-time model would be more challenging, in part because we would have to take a stand on the appropriate wage-setting mechanism in continuous time, which would take us far afield. Nevertheless, any mobility cost will lengthen the time the worker spends at a particular job. If the firm has some bargaining power, this creates the conditions for firm investment in the worker's skills and overinvestment.

$$\max_{I:[0,T]\to\mathbb{R}_+^N} \quad \int_0^\infty e^{-rt}\left[\left(\sum_n (A_n S_n(t))^{\frac{\sigma}{\sigma-1}}\right)^{\frac{\sigma-1}{\sigma}} - \sum_n I_n(t)^\rho\right] dt \tag{16}$$

$$s.t.\ \frac{d}{dt}S(t) = -\Delta S(t) + I(t) \tag{17}$$

$$S(0) = S_0. \tag{18}$$

We therefore construct the Hamiltonian

$$H = e^{-rt}\left[\left(\sum_n (A_n S_n(t))^{\frac{\sigma}{\sigma-1}}\right)^{\frac{\sigma-1}{\sigma}} - \sum_n I_n(t)^\rho\right] + \sum_n \mu_n(-\Delta S_n(t) + I_n(t)). \tag{19}$$

The solution is given by the N, one for each skill n, equations

$$\frac{A_n^{\frac{\sigma}{\sigma-1}}(S_n(t))^{\frac{1}{\sigma-1}}}{\left(\sum_n (A_n S_n(t))^{\frac{\sigma}{\sigma-1}}\right)^{\frac{1}{\sigma}}} + \rho(\rho-1)I_n(t)^{\rho-2}\frac{dI_n(t)}{dt} - (r+\Delta)\rho I_n(t)^{\rho-1} = 0 \tag{20}$$

along with the motion equations for skills

$$\frac{d}{dt}S(t) = -\Delta S(t) + I(t), \tag{21}$$

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

4.2 Ben-Porath Case

For some parameter values, a special case where all skills grow at the same rate obtains. In such a case the solution to (20) becomes

$$I_n(t) = \left(K_n \frac{1 - e^{(r+\Delta)(t-T)}}{\rho(r+\Delta)}\right)^{\frac{1}{\rho-1}} \tag{22}$$

where K_n is a constant. This leads to a constant ratio of investment in any two skills.

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, and the optimal investment path maintains that ratio. 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

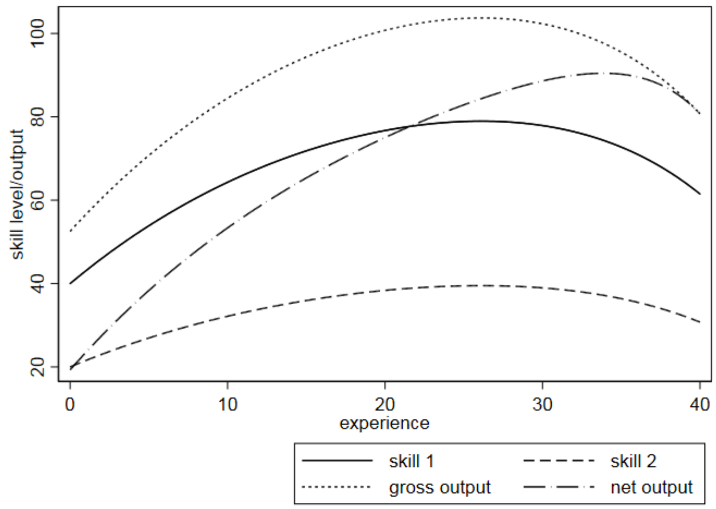

Figure 2: The 'Ben-Porath' case of proportional skill evolution. $A_1 = 2^{.2}$, $A_2 = 1$, $\rho = 2$, $\sigma = 2.5$, $S_0 = (40, 20)$, $\Delta = r = .05$

is the point at which the two are equal.

4.3 Jobs and Skills Over the Lifecycle

It is generally not possible to obtain a closed form solution for (20). We can, however, solve the system numerically for given values of A, Δ, r and S_0. To demonstrate the potential usefulness of this approach, we present two scenarios that we find particularly interesting.

First, we illustrate the case where the response to a shock is discontinuous. In this scenario, the worker enters the labor market with 10 units of each skill. Initially skill 1 is more valuable ($A_1 = 1.15$ and $A_2 = 1.00$). At some unanticipated point, the skill weights reverse. We have chosen parameters such that the worker will tend to move towards jobs that are intensive in their use of one skill.

Not surprisingly, when the worker is young, following the shock, she sharply shifts her investment towards skill 2. However, if she has been in the labor market for more than a few years, she has accumulated enough skill 1 that it is no longer beneficial for her to shift towards jobs that use skill 2 more intensively. As a consequence of the shock, she invests less heavily in skill 1 than she would have otherwise. She also invests somewhat more in skill 2, but the effect is so small that it would not be visible in the figure. The discontinuity arises from the horizon effect: as she ages the worker has less time in which to recoup her investment, but more importantly in this case, from the inertia effect: her accumulated investment in skill 1 makes it very costly to shift to a skill-2-intensive occupation.

Our second scenario features three skills which we refer to as Manual, Routine (Cognitive)

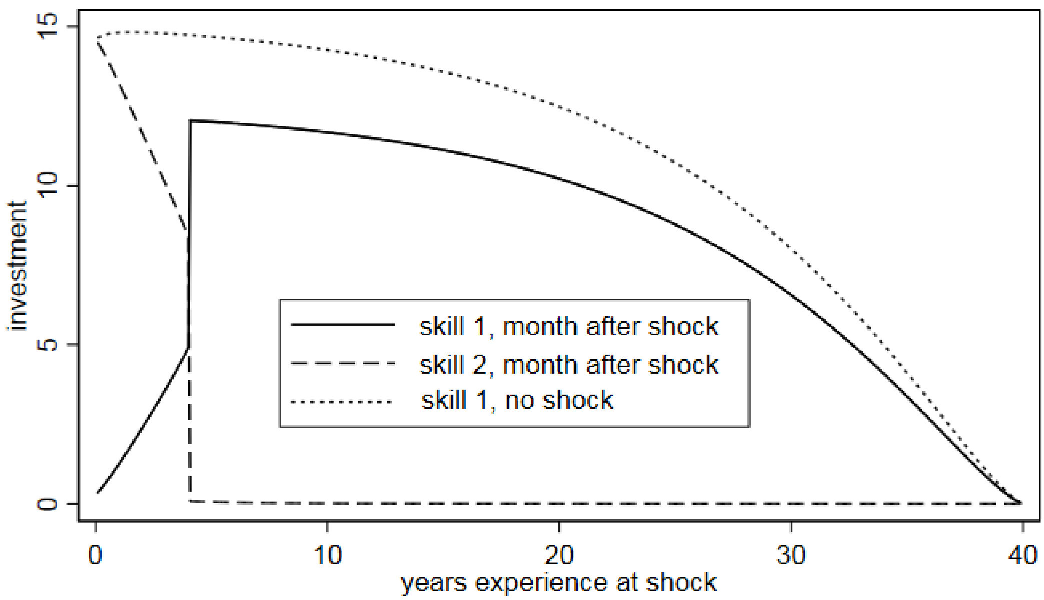

Figure 3: Initial Investment by Experience at Time of Shock
$\rho = 1.7$, $\sigma = 1.7$, $\Delta = r = .05$

and Abstract. Our chief example considers a worker subject to an unanticipated shock that increases the value of the Abstract skill while also decreasing the value of the Routine skill.

We consider an individual who arrives in the labor market with 25 units of each skill. Initially the Routine skill is the most valuable ($A_{Rout} = 1.2$); the first lies in the middle ($A_{Man} = 1.13$) and the third skill is the least valuable ($A_{Abs} = .8$) to the worker. The worker is assumed to be in the labor force for forty years, discounts future income using an interest rate of five percent which is also the rate at which each skill depreciates.

We consider an unanticipated shock that occurs in either the worker's 10th, 20th or 30th year in the market. The shock reduces A_{Rout} to .8 and increases A_{Abs} to 1.25 while leaving A_{Man} at 1.13.

While we view this exercise as an example rather than as a serious calibration exercise, we have chosen the parameters to be broadly consistent with reality. Quick conversations with some of the leaders in the field confirmed that the magnitude of the shocks we assume is broadly consistent with the post-1980s shock to the U.S. economy although, of necessity, we model the shock as instantaneous.[15] The remaining parameters were chosen to produce a wage profile broadly consistent with typical ordinary least squares estimates of the experience/earnings profile. If we regress log earnings (ln(gross output - investment cost)) on time and time squared measured in years, the coefficients are .100 and -.00162. This compares

[15]While it would be computationally messy, we could model the shock as a series of unexpected shocks playing out over a longer period. However, a more serious treatment would have to address how expectations formed following the initial shock, something that would take us far from the current model.

with OLS estimates for 1980 in Heckman, Loechner and Todd (2006) of .1255 and -.0022 for whites and .1075 and -.0016 for blacks. The OLS approximation to our data implies that earnings peak at 31 years of experience with earnings growth of approximately 154 log points at the peak. The Heckman, Loechner and Todd estimates imply a peak at about 29 years for whites and about 34 years for blacks, and earnings growth of 179 and 181 log points at the peaks. Using the simulated data rather than the quadratic approximation to those data, earnings peak at about 34 years of experience and are about 183 log points higher at the peak than initially.

Since in the example workers arrive in the market with similar amounts of the Routine and Abstract skills, the shock represents a mild form of positive shock for the youngest workers. However, given the initial technology, the worker will invest most heavily in the Routine skill before the shock. Therefore the adjustment path of experienced workers reflects both a seizure of new opportunities and a retrenchment from declining occupations.

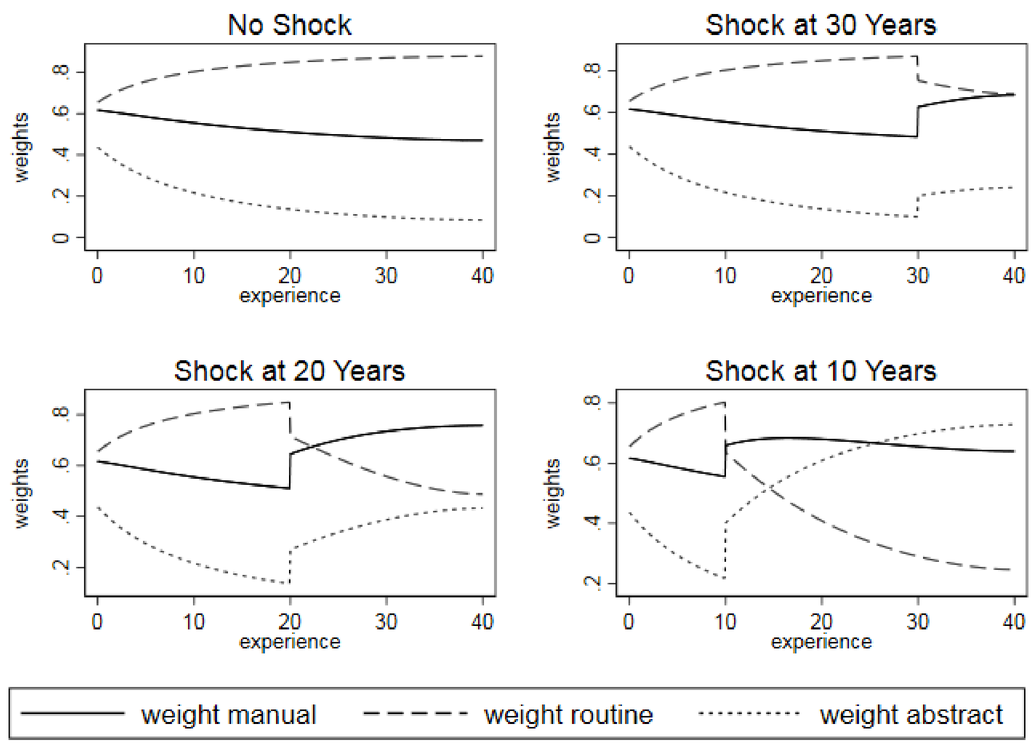

Figure 4: Job's Skill Weights by Experience at Time of Shock

Figure 4 shows the path of the worker's job's skill weights 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 the Routine skill.

We see that if the shock arrives when she has thirty years experience, she immediately mechanically (since A_{Rout} falls) chooses a slightly less Routine-intensive job. Overall, she adjusts very little. She continues to work in Routine-heavy jobs as she has accumulated a large stock of Routine skill even though the value of that stock has fallen by about a third, although she also shifts somewhat towards more manual-intensive jobs. Much of the increased weight on Abstract skill reflects the greatly increased value of that skill in all jobs rather than a shift towards investment in Abstract skill.

A shock at twenty years of experience has a more noticeable effect on career (job) choices. But because the worker's stock of Abstract skill has depreciated so much over twenty years, by the end of her career, she shifts towards Manual-heavy jobs. Again, much of the increased weight on the Abstract skill is mechanical.

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

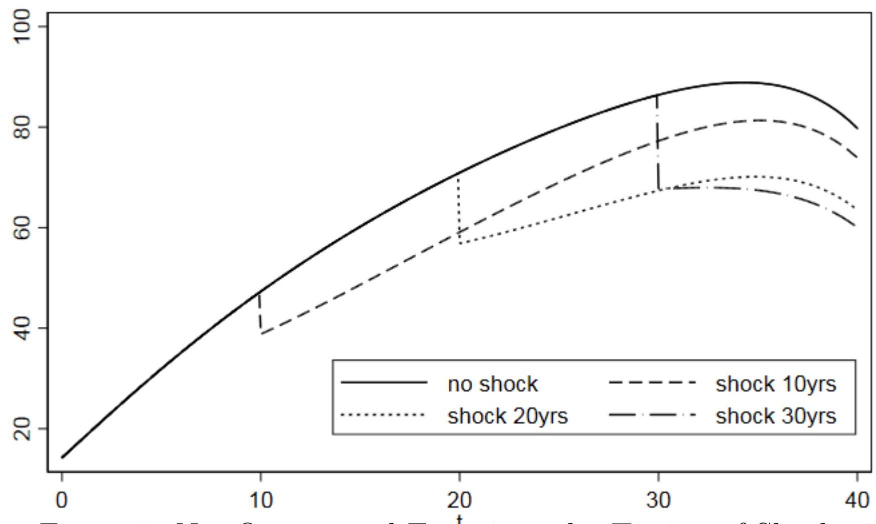

Figure 5: Net Output and Experience by Timing of Shock

Figure 5 shows net output over time. As prior to the shock, the worker invested most heavily in a skill 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 the Routine skill she has accumulated relative to the other skills. 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 every later experience level. The person shocked at 20 years of experience fares almost as badly in the last 10 years of work as the person shocked at 30 years.

In this example the shock is in a sense positive; a worker who begins her career just as the shock hits will earn 4.5 percent more over her lifetime than if she finished her career before the shock hit. 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. The significant point is that a positive shock can have a negative effect for a very long time. In Figure 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.

If the workers in our example 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 reduce 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. This implies that debt peaks at around age 35 and that savings turn positive at around age 49.

Here we find that a worker shocked at 30 years of experience must reduce her consumption by 22 percent relative to what she had anticipated. She had anticipated doing most of her saving for her retirement during the last ten years of her working life and is therefore hit sharply by the decline in her earning capacity over this period. Although both hold large and similar quantities of debt, workers shocked at 20 and 10 years of experience must reduce their consumption by 27 percent and 19 percent. Of course, the worker shocked at 10 years suffers this consumption loss over a longer period.

Perhaps the most striking aspect of the example is the length of time for which an ultimately positive shock can be negative. 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.

4.4 Why older workers adjust less

Table 1: Different skill paths in response to shock

Skill levels when shocked			Time	Twenty years after shock		
Manual	Routine	Abstract	remaining	Manual	Routine	Abstract
38.44	52.38	21.17	30	52.76	32.94	50.73
38.44	52.38	21.17	20	39.23	30.17	30.13
43.77	68.51	16.56	30	59.24	42.76	39.13
43.77	68.51	16.56	20	42.61	38.74	22.00

We now wish to explore *why* workers vary in their adjustment. In Table 1, we present four workers at the time the shock hits. Two have accumulated the level of skills that the baseline worker in the example has accumulated after ten years (the first two rows of the table), and two have the level of skills this worker accumulated after twenty years. Within each pair, we consider the investment decisions of such a worker with thirty (rows one and three) and twenty (rows two and four) work years remaining. The last three columns show the stock of skills for each worker twenty years after the shock.

Comparing the first and fourth rows shows the difference, after twenty years, in the skills of a worker shocked at ten and twenty years of experience. Relative to the latter, the former has substantially more Manual skill and, especially, Abstract skill, and less Routine skill despite the fact that the she had less Manual skill and only slightly more Abstract skill at the time the shock hit. As a consequence of both her greater accumulation of Routine skill and her shorter horizon, the worker shocked at twenty years ends up twenty years later with more of the Routine skill and less of the other skills. How much of this difference can be attributed to the fact that she has a shorter time horizon and how much to the fact that she is heavily invested in the Routine skill at the time the shock hits?

Comparing rows one and two casts light on the horizon effect. The worker with the shorter horizon ends up with about three-quarters as much of the Manual skill, a little more than half as much of the Abstract skill, and only slightly less of the Routine skill. The horizon effect causes less investment in all skills, but this effect is most noticeable for the positively shocked (Abstract) skill and least noticeable for the negatively shocked (Routine) skill.

By comparing rows one and three we can find the inertia effect. This effect generates a substantial reduction in the growth of the stock of the Abstract skill. The change in the final stock of the Manual skill is not greatly different from the difference in the initial stocks. The stock of Routine skill ends up substantially higher when the shock comes later but by noticeably less than the difference in the initial stocks. Note however, that due to depreciation the initial 16.1 difference in the stocks of Routine skill would have fallen to 5.9 had they invested at the same rate. The difference of 9.8 in their stocks after twenty years shows that the inertia effect actually leads the worker with a greater initial stock to invest more over the next twenty years. What remains once we have accounted for the two effects is the 'interaction effect', an adjustment when both the inertia and horizon effects apply at once.

Table 2 shows the responses in terms of job choices for workers with different initial stocks of skill and time remaining. Just prior to the shock, our worker who is shocked after ten years in the labor market is in a job that puts the most weight on the Routine skill. Twenty

Table 2: Job choice responses to shock

Job's weights when shocked			Time	Twenty years after shock		
Manual	Routine	Abstract	remaining	Manual	Routine	Abstract
.55	.80	.22	30	.66	.29	.70
.55	.80	.22	20	.70	.38	.60
.51	.85	.14	30	.75	.38	.55
.51	.85	.14	20	.76	.49	.43

years later her job puts the most weight on the Abstract skill, almost as much on the Manual skill and far less on Routine skill. In contrast, just before our worker with twenty years of experience suffers the shock, she is in an even more Routine-intensive job that puts almost no weight on the Abstract skill. Twenty years later, her job puts the most weight on the Manual skill, and roughly equal weights on the other two skills.

Again we can analyze this difference in terms of both the horizon and inertia effects. Both effects are roughly in the same direction: a shift towards Manual tasks and a moderation of the movement from Routine towards Abstract tasks. Strikingly for the Routine and Abstract tasks the changes from the first to the fourth row are close to the sum of the inertia and horizon effects. In contrast, the interaction of the two effects blunts the shift to the Manual task.

5 Discussion and Conclusion

We develop a model of the microfoundations of dynamic skill formation that provides both qualitative and quantitative insights. Nonconvexities that arise naturally in the model can create settings in which credit constraints affect behavior even though they do not bind with equality. Additionally, we explain why firms may invest in general skills: to manipulate worker investment incentives towards skills that the firm values and away from those that improve her outside option. This analysis draws on the insight of Lazear (2009) whose work is complementary to the story told here.

We extend the model to a tractable continuous-time setting. This allows us to investigate why similar workers of different ages react to shocks very differently. We show that, due to the front-loading of investment, large shocks, even if positive on net, can have long-lasting adverse effects on even relatively young workers. Although only an example, this should make us think very carefully about winners and losers and perhaps even the political economy issues.

Since Ricardo, arguments for trade and technological innovation have relied on compensating transfers. Our model suggests that 'losers' may be difficult to detect because they include not only those who continue using similar skills even though their value has declined, but also some workers displaced to jobs that use skills that have increased in value. Moreover, the importance of credit constraints for limiting transitions to better jobs may be hidden because workers who do not appear to be credit constrained are unable to afford the optimal set of new skills.

6 References

Acemoglu, Daron and Jörn-Steffen Pischke. 1999a. "Beyond Becker: Training in Imperfect Labor Markets." *Economic Journal* 109 (February): F112-42.

Acemoglu, Daron and Jörn-Steffen Pischke. 1999b. "The Structure of Wages and Investment in General Training." *Journal of Political Economy* 107 (June): 539-72.

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

Arora, Ashna. 2018. "Juvenile Crime and Anticipated Punishment." Working Paper no. 46 (January), Center for Development Economics and Policy, Columbia University, New York, NY.

Autor, David H. and David Dorn. 2009. "This Job is 'Getting Old': Measuring Changes in Job Opportunities using Occupational Age Structure." *American Economic Review: Papers and Proceedings* 99 (May): 45-51.

Autor, David H. and David Dorn. 2013. "The Growth of Low-Skill Service Jobs and the Polarization of the US Labor Market." *American Economic Review* 103 (August): 1553-97.

Ben-Porath, Yoram. 1967. "The Production of Human Capital and the Life Cycle of Earnings." *Journal of Political Economy* 75 (August): 352-65.

Bernhardt, Dan, and David Backus. 1990. "Borrowing Constraints, Occupational Choice, and Labor Supply." *Journal of Labor Economics* 8 (January): 145-73.

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

Cunha, Flavio and James J. Heckman. 2008. "Formulating, Identifying, and Estimating the Technology of Cognitive and Noncognitive Skill Formation." *Journal of Human Resources* 43 (Fall): 738-82.

Heckman, James J., Lance J. Loechner and Petra E. Todd. 2006. "Earnings Functions, Rates of Return and Treatment Effects: The Mincer Equation and Beyond." in Eric A. Hanushek and Finis Welch, eds., *Handbook of the Economics of Education, Volume 1*, New York and Amsterdam: Elsevier, 307-458.

Jovanovic, Boyan. 1979. "Firm-specific capital and turnover." Journal of Political Economy 87 (December): 1246-60.

Konrad, Kai A., and Kjell Erik Lommerud. 2000. "The bargaining family revisited." *Canadian Journal of Economics* 33 (April): 471-87.

Lang, Kevin and Paul Ruud. 1986. "Returns to Schooling, Implicit Discount Rates and Black-White Wage Differentials." *Review of Economics and Statistics* 68 (February): 41-7.

Lazear, Edward. 2009. "Firm-Specific Human Capital: A Skill-Weights Approach," *Journal of Political Economy* 117 (October): 914-40.

Lochner, Lance and Alexander Monge-Naranjo. 2012 "Credit Constraints in Education." *Annual Review of Economics* 4 (September): 225-56.

Mellor, Earl F. 1985. "Weekly Earnings in 1983: A Look at More than 200 Occupations." *Monthly Labor Review* 108, (January): 54-9.

Prada, Maria F., and Sergio S. Urzua. 2017. "One Size Does Not Fit All: Multiple Dimensions of Ability, College Attendance and Earnings." *Journal of Labor Economics* 35 (October): 953-91.

Sanders, Carl, and Christopher Taber. 2012. "Life-Cycle Wage Growth and Heterogeneous Human Capital." *Annual Review of Economics* 4.(September): 399-425.

Sviatschi, Maria Micaela. 2018. "Making a *Narco*: Childhood Exposure to Illegal Labor Markets and Criminal Life Paths." Manuscript, Department of Economics, Princeton University.

A Proofs

A.1 Proof of Proposition 1

Let $(I', J_2'), (I, J_2)$ solve the corresponding problems. Then, from optimality,

$$\beta\Sigma_m J_{2,m} A_m(\delta S_m + I_m) - C(I) \quad + \quad \beta\Sigma_m J'_{2,m} A_m(\delta S'_m + I'_m) - C(I') \tag{23}$$

$$\geq \beta\Sigma_m J_{2,m} A_m(\delta S'_m + I_m) - C(I) \quad + \quad \beta\Sigma_m J'_{2,m} A_m(\delta S_m + I'_m) - C(I') \tag{24}$$

which, cancelling terms, implies

$$\Sigma_m A_m(J_{2,m} - J'_{2,m})(S_m - S'_m) \geq 0.$$

which, recalling that for $m \neq n$, $S_m = S'_m$, becomes

$$(J_{2,n} - J'_{2,n})(S_n - S'_n) \geq 0.$$

Given that $S'_n > S_n$ by assumption, we have $J_{2,n} \leq J'_{2,n}$. If $J_{2,n} = J'_{2,n}$ then it follows that J'_2 must also solve the problem for endowment S, as incentives in dimensions other than n are unchanged; but J'_2 cannot solve the problem for both S and S' as they produce FOCs in dimension n that are different. Therefore $J'_{2,n} > J_{2,n}$. From that we deduce $I'_n > I_n$, as investment in skill n for a given second-period job $J^*_{2,n}$ is $I^*_n = [A_n J^*_{2,n} \rho^{-1}]^{\frac{1}{\rho-1}}$, which is increasing in $J^*_{2,n}$.

A.2 Proof of Proposition 2

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

$$\beta V(\delta S + I^*) - C(I^*) \geq \beta V(\delta S + I^{*\prime}) - C(I^{*\prime}) \tag{25}$$

$$\beta' V(\delta S + I^{*\prime}) - C(I^{*\prime}) \geq \beta' V(\delta S + I^*) - C(I^*) \tag{26}$$

so that, after some manipulation

$$(C(I^{*\prime}) - C(I^*))(\beta' - \beta) \geq 0 \tag{27}$$

$$C(I^*) \geq C(I^{*\prime}) \tag{28}$$

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 $I^{*\prime}$ improves on the objective over I^* in the β problem, so I^* is not a maximizer. Hence, it must be the case that $C(I^*) > C(I^{*\prime})$.

A.3 Proof of Proposition 3

Fix ρ, σ such that $(\rho - 1)(\sigma - 1) < 1$. WLOG let $A_1 = min_n A_n$; by assumption there's an n such that $A_1 < A_n$. Let $S = (1, 0, 0, ..., 0)$.

Let I^* be a function such that for all $\beta > 0$, $I^*(\beta)$ is an optimal investment. We claim I^* is not continuous.

First, we show that for high enough β, the worker does invest in skills other than 1. If the worker invests only in skill 1, she solves

$$\max_{I_1} \beta A_1[\delta + I_1] - I_1^{\rho} \tag{29}$$

for optimal investment

$$I_1 = \left[\frac{\beta A_1}{\rho}\right]^{\frac{1}{\rho-1}} \tag{30}$$

so that the value of the worker's problem is

$$\beta A_1 \left[\delta + \left(\frac{A_1 \beta}{\rho}\right)^{\frac{1}{\rho-1}}\right] - \left[\frac{\beta A_1}{\rho}\right]^{\frac{\rho}{\rho-1}} = \beta A_1 \delta + A_1^{\frac{\rho}{\rho-1}}(\rho \beta^{\frac{\rho}{\rho-1}} - 1)\rho^{-\frac{1}{\rho-1}} \tag{31}$$

whereas investing only in a skill n and choosing a job putting weight only on skill n similarly yields a value of

$$A_n^{\frac{\rho}{\rho-1}}(\rho \beta^{\frac{\rho}{\rho-1}} - 1)\rho^{-\frac{1}{\rho-1}}. \tag{32}$$

Investment in only skill 1 would therefore not be optimal if the value from investing in only skill n is higher; this is the case when

$$\beta A_1 \delta + A_1^{\frac{\rho}{\rho-1}}(\rho \beta^{\frac{\rho}{\rho-1}} - 1)\rho^{-\frac{1}{\rho-1}} < A_n^{\frac{\rho}{\rho-1}}(\rho \beta^{\frac{\rho}{\rho-1}} - 1)\rho^{-\frac{1}{\rho-1}} \tag{33}$$

which is necessarily true for high enough β as $\frac{\rho}{\rho-1} > 1$ and $A_n > A_1$. Therefore, there exists a $\beta \geq 0$ and $m \neq 1$ such that $I_m^*(\beta) \neq 0$. When $\beta = 0$, there is no future to invest in and $I^*(\beta) = 0$.

Thus, if I^* were continuous, a sequence $\beta_n \to \bar{\beta}$ would exist with the property that $0 \neq I_m^*(\beta_n) \to 0$. However, the derivative of the worker's objective at positive I_m with respect to I_m is

$$\beta \frac{A_m^{\frac{\sigma}{\sigma-1}} I_m^{\frac{1}{\sigma-1}}}{\left(\Sigma_k [A_k(\delta S_k + I_k)]^{\frac{\sigma}{\sigma-1}}\right)^{\frac{1}{\sigma}}} - \rho I_m^{\rho-1} \tag{34}$$

which is less than

$$\beta \frac{A_m^{\frac{\sigma}{\sigma-1}} I_m^{\frac{1}{\sigma-1}}}{(A_1 \delta)^{\frac{1}{\sigma-1}}} - \rho I_m^{\rho-1} \tag{35}$$

but as $(\rho-1)(\sigma-1) < 1$, we have $\rho - 1 < \frac{1}{\sigma-1}$, so the second expression must be negative for low enough positive I_m. Therefore, the first-order condition for positive I_m cannot be satisfied as $\beta_n \to \bar{\beta}$. Thus, I^* cannot be continuous.

A.4 Proof of Proposition 4

Take some $x > 0$ such that $x^{\frac{\rho}{\rho-1}}(1 - \frac{\rho}{\beta(\rho-1)}x^{\frac{(\sigma-1)(\rho-1)-1}{\rho-1}}) > 1$. As $(\sigma-1)(\rho-1) < 1$, this is possible. Then, assume an A such that there are n, m with $A_n \cdot x < A_m$; convene WLOG that $A_1 = min_l A_l$ and $A_2 = max_l A_l$. Then we can find $k > 1$ such that $\left(\frac{A_2}{A_1}\right)^{\frac{\rho}{\rho-1}}(1 - k\frac{\rho}{\beta(\rho-1)}\left(\frac{A_2}{A_1}\right)^{\frac{(\sigma-1)(\rho-1)-1}{\rho-1}}) > 1$. Set $c = \left(\frac{\beta A_1}{\rho}\right)^{\frac{\rho}{\rho-1}} k^\rho$ and $S = \left(\left(\frac{A_2}{A_1}\right)^\sigma c^{\frac{1}{\rho}}, 0, 0...\right)$.

First, consider the constrained problem. Suppose that the constraint c binds. Consider the problem of allocating c across the different skills; each $i > 1$ is allocated I_i and the remainder goes to skill 1. If the constraint binds, the investment decision must

$$\max \Pi_{bind} = \max_{I_{-1}} \left([A_1(S_1 + (c - \sum_{i>1} I_i^\rho)^{\frac{1}{\rho}})]^{\frac{\sigma}{\sigma-1}} + \sum_{i>1} (A_i I_i)^{\frac{\sigma}{\sigma-1}} \right)^{\frac{\sigma-1}{\sigma}}$$

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

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

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

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

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

$$\frac{\sigma}{\sigma-1} I_j^{\frac{1}{\sigma-1}} \left(A_j^{\frac{\sigma}{\sigma-1}} - A_2^{\frac{\sigma}{\sigma-1}} \right) \leq 0.$$

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

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

so that $I_{c,1}^* = \left(\frac{\beta A_1}{\rho}\right)^{\frac{1}{\rho-1}}$ for an expenditure of $\left(\frac{\beta A_1}{\rho}\right)^{\frac{\rho}{\rho-1}} < c$. 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 $\beta A_2 I_2^* - I_2^{*\rho} =$

$\left(\frac{\beta A_2}{\rho}\right)^{\frac{\rho}{\rho-1}}(\rho-1)$. We want to show this is greater than the constrained payoff, that is,

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

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

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

which we know to be true; therefore, the unconstrained problem yields strictly higher utility. So, it must be the case that any solution to the unconstrained problem must violate the constraint; thus, $C(I^*) > c > C(I_c^*)$.

A.5 Proof of Proposition 5

Step 1: At least one skill is overinvested in.
If for all n we have $I_n^W(I^F) \leq I_n^*(J)$ then from the solution to the static problem we have $J^*(\delta S + I^W(I^F)) \leq J$ and therefore either the constraint job constraint holds without equality: $\Sigma_k J_k^\sigma < 1$ and thus a strictly more productive job exists (a contradiction to optimality); or J is optimal in the second period in the absence of a mobility cost, which is not the case as (i) ex hypothesi the worker would move absent the cost and (ii) we know that the solution to the static job-choice problem is unique. Therefore there is an n for which $I_n^W(I^F) > I_n^*(J)$. There is at least one skill that is overinvested in.

Step 2: No skill is underinvested in.

Consider the worker's problem after accepting an offer. Given I^F, the worker chooses I^W to maximize her second period wage net of what she spends on investment:

$$\max_{I^W} \beta V(\delta S + I^W) - m - c(I^W) + c(I^F) \qquad \text{s.t. } I^W \geq I^F \tag{36}$$

which given the exogeneity of I^F and m is the same problem as

$$\max_{I^W}\left[\beta\left(\sum_n (A_n(\delta S_n + I_n^W))^{\frac{\sigma}{\sigma-1}}\right)^{\frac{\sigma-1}{\sigma}}\right] - \sum_n I_n^{W\rho} \quad \text{s.t. } I^W \geq I^F \tag{37}$$

The first order condition for I_n^W (when the n'th constraint does not bind with equality and thus $I_n^W > I_n^F$) is

$$\beta A_n^{\frac{\sigma}{\sigma-1}}(\delta S_n + I_n^W)^{\frac{1}{\sigma-1}} \left(\sum_m (A_m(\delta S_n + I_m^W))^{\frac{\sigma}{\sigma-1}} \right)^{-\frac{1}{\sigma}} - \rho I_n^{W\ \rho-1} = 0 \tag{38}$$

which can be rewritten as

$$\beta A_n^{\sigma} \frac{\delta S_n + I_n^W}{I_n^{W(\rho-1)(\sigma-1)}} = \rho V(\delta S + I^W) \tag{39}$$

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

Now we will show that if some skill is invested in to a level less than what is efficient at J, the initial contract could have been improved. If for some m such that $I_m^F = I_m^W(I^F)$ we had $\frac{dV(\delta S + I^W)}{dI_m^F} \leq 0$ (that is, if there is a constraint binding with equality that affects V negatively) then for any n such that $I_n^W(I^F) > I_n^F$, we would have $\frac{dI_n^W(I^F)}{dI_m^F} \geq 0$ and thus as $V(\delta S + I^W) = (\sum_k (A_k(\delta S_k + I_k^W))^{\frac{\sigma}{\sigma-1}})^{\frac{\sigma-1}{\sigma}}$, V would have to increase, a contradiction. Thus, we have shown that if the constraint in dimension n does not bind and $I_m^W(I^F) = I_m^F$ then $dI_n^W(I^F)/dI_m^F < 0$ $(\star)$.

Ex ante, as firms act perfectly competitively, the accepted period-1 contract maximizes the worker's lifetime utility subject to non-negative profits. That is, (J, I^F, w) must

$$\max_{J,I^F,w} \left[w + \beta V(\delta S + I^W) - c(I^F(I^W \right.$$

$$\text{s.t.} \sum_n J_n A_n (S_n + \beta(\delta S_n + I_n^W)) - \beta(V(\delta S + I^W(I^F)) - m) - w \quad \geq \quad 0$$

so that I^F in particular has to solve

$$\max_{I^F} \beta \sum_n J_n A_n (\delta S_n + I_n^W(I^F)) - \sum_n (I_n^W(I^F))^{\rho} \tag{42}$$

Suppose there is an n for which $I_n^W(I^F) < I_n^*(J)$, so that skill n is underinvested in. Then, setting $\forall m, I_m^{F\prime} := \max\{I_m^W, I_m^*(J)\}$, from $\star$ we have that $I^W(I^{F\prime}) = I^{F\prime}$. That is, as $I^{F\prime}$ produces a strictly higher V than $I^W(I^F)$, then when the contract is $(J, I^{F\prime}, w)$ the vector of constraints $I^W(I^{F\prime}) \geq I^{F\prime}$ must bind with equality. Therefore, as $I^{F\prime}$ improves the objective (42), we have a contradiction. Therefore, no skill is underinvested in: $I^W(I^F) \geq I^*(J)$. Combining this with Step 1, we have $I^W(I^F) \gneqq I^*(J)$.

Step 3: Every skill is overinvested in.

Now suppose there is a skill that is not overinvested in: by Step 2 this means that

$\exists n : I_n^W(I^F) = I_n^*(J)$. In Step 1, we have established that there is a skill m that is overinvested in: $I_m^W(I^F) > I_m^*(J)$. Two cases are of interest.

Case 1. Suppose the unconstrained first-order condition (39) holds for I_m^W. Then, consider the firm investment $\bar{I}^F = I^W(I^F)$. We have $I^W(\bar{I}^F) = I^W(I^W(I^F)) = I^W(I^F)$ so that since I^F is part of an optimal contract so is $\bar{I}^F$ (albeit with a compensating period-1 wage). We will consider increasing $\bar{I}_n^F$ to effect an increase in V and through it will implement a decrease in $\bar{I}_m^F$ without affecting $(\bar{I}_k)_{k\notin\{n,m\}}$.

We define the auxiliary function $\hat{I}_m^F(\hat{I}_n^F)$ implicitly by the worker's FOC in dimension n:

$$\beta A_m^\sigma \frac{\delta S_m + \hat{I}_m^F}{\hat{I}_m^{F(\rho-1)(\sigma-1)}} = \rho V\left((\delta S_n + \hat{I}_n^F), (\delta S_m + \hat{I}_m^F), (\delta S_k + \bar{I}_k^F)_{k\notin\{n,m\}}\right). \tag{43}$$

That is, $\hat{I}_m^F(\cdot)$ maps out, for each level of firm investment in skill n, the worker's optimal investment in skill m, when all other skill investment is held constant. This will allow us to vary skill m by varying skill n without violating the worker's optimal choice of skills.

As $S >> 0$, we have $\partial V/\partial \hat{I}_n^F > 0$ and $\partial V/\partial \hat{I}_m^F > 0$; furthermore, the left hand side of (43) is decreasing in $\hat{I}_m^F$ as $(\rho-1)(\sigma-1) > 1$. Therefore, we have $d\hat{I}_m^F(\hat{I}_n^F)/d\hat{I}_n^F < 0$. If all skills other than n and m are held constant, when the firm forces an increase in skill n, the worker decreases skill m.

Consider now perturbing the investment $\bar{I}^F$ by increasing $\bar{I}_n^F$ and lowering $\bar{I}_m^F$ along $\hat{I}_m^F(\cdot)$. As $V\left((\delta S_n + \hat{I}_n^F), (\delta S_m + \hat{I}_m^F), (\delta S_k + \bar{I}_k^F)_{k\notin\{n,m\}}\right) \geq V(\bar{I}^F)$ when $\hat{I}_n^F \geq \bar{I}_n^F$, we have that $I^W(\hat{I}_n^F, \hat{I}_m^F(\hat{I}_n^F), (\bar{I}_k)_{k\notin\{n,m\}}) = (\hat{I}_n^F, \hat{I}_m^F(\hat{I}_n^F), (\bar{I}_k)_{k\notin\{n,m\}})$.

Written solely in terms of $\hat{I}_n^F$ (and keeping constant skills other than n and m), the objective function (42) is

$$\beta J_n A_n(\delta S_n + \hat{I}_n^F) + \beta J_m A_m(\delta S_m + \hat{I}_m^F(\hat{I}_n^F)) + \beta \sum_{k\notin\{n,m\}} J_k A_k(\delta S_k + \bar{I}_k^F)$$
$$- C((\hat{I}_n^F, \hat{I}_m^F(\hat{I}_n^F), (\bar{I}_k)_{k\notin\{n,m\}}))$$

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

$$\beta A_n J_n + \beta A_m J_m \frac{d\hat{I}_m^F(\hat{I}_n^F)}{d\hat{I}_n^F} - \rho \hat{I}_n^{F\rho-1} - \frac{d\hat{I}_m^F(\hat{I}_n^F)}{d\hat{I}_n^F} \rho(\hat{I}_m^F(\hat{I}_n^F))^{\rho-1} \tag{44}$$

$$= (\beta A_n J_n - \rho \bar{I}_n^{F\rho-1}) + \frac{d\hat{I}_m^F(\hat{I}_n^F)}{d\hat{I}_n^F}(\beta A_m J_m - \rho(\hat{I}_m^F(\bar{I}_n^F))^{\rho-1}) \tag{45}$$

But by assumption $\bar{I}_n^F = I_n^*(J)$, so that $\beta A_n J_n - \rho \bar{I}_n^{F\rho-1} = 0$ and $\hat{I}_m^F(\bar{I}_n^F) = \bar{I}_m^F > I_m^*(J)$ so that $\beta A_m J_m - \rho \hat{I}_m^F(\bar{I}_n^F)^{\rho-1} < 0$. Furthermore, we have that $\frac{d\hat{I}_m^F(\hat{I}_n^F)}{d\hat{I}_n^F} < 0$. Therefore evaluated at $\bar{I}_n^F$, the restricted objective function's right derivative is positive. As a result, there exists a $\hat{I}_n^F > \bar{I}_n^F$ so that $(\hat{I}_n^F, \hat{I}_m^F(\hat{I}_n^F), (\bar{I}_k)_{k \notin \{n,m\}})$ improves the objective function (42) over the assumed maximizer I^F, a contradiction.

Case 2. Now suppose instead that the unconstrained first order condition (39) does not hold for any overinvested-in skill m. Consider again $\bar{I}^F = I^W(I^F)$, which is again optimal under the hypothesis that I^F is. For some overinvested-in skill m define $\hat{I}_m^F(\hat{I}_n^F)$ implicitly by
$V\Big((\delta S_n + \hat{I}_n^F), (\delta S_m + \hat{I}_m^F), (\delta S_k + \bar{I}_k^F)_{k \notin \{n,m\}}\Big) = V(\delta S + \bar{I}^F)$ when $\hat{I}_n^F \geq \bar{I}_n^F$ is small enough for a solution to exist. In other words, $\hat{I}_m^F$ adjusts to $\hat{I}_n^F$ so as to keep production at the optimal outside option job constant (even as the optimal outside job may change).

Then as V is constant along $(\hat{I}_n^F, I_m^F(\hat{I}_n^F))$ for $\hat{I}_n^F \geq \bar{I}_n^F$, skills $k \notin \{n, m\}$ stay constant. As V has strictly positive (as $S >> 0$) and continuous partials, $d\hat{I}_m^F(\hat{I}_n^F)/d\hat{I}_n^F > 0$; the rest of the argument follows as in Case 1.

B A Little Empirical Evidence

We examine workers who were in routine-cognitive intensive jobs in 1968 and in 1982 and examine their use of such skills fourteen years later. The IBM PC was introduced in mid-1981. A typist in 1968 was likely to be using an IBM Selectric typewriter although she (most probably) might have used a less sophisticated electric typewriter or even a manual typewriter. By 1982, it is very likely that she would have used an IBM Selectric although occasionally she might have moved on to an early wordprocessing system. In either case, her work would not have changed dramatically. And the job of the young typist in 1982 would not look that different from her counterpart in 1968. By 1996, the spread of the personal computer had dramatically reduced the role of typists. A similar story can be told for bookkeeping and other jobs that were routine-cognitive intensive.

To look at how employment of such workers changed, we use the Panel Study of Income Dynamics. We select the principal respondent and spouse, if any, who were employed and present in the sample either in both 1968 and 1982 or both 1982 and 1996 and who were age 20-49 at the beginning of the relevant period. We use the skill measures and crosswalk for 1970 occupations from Autor and Dorn (2013). We define a job as routine-cognitive intensive if it was in the top quartile of the use of routine-cognitive skills in 1982. The top quartile is measured by the distribution of routine-cognitive tasks in the unweighted sample.

We further require that their use of each of manual and abstract skills be below the average for the unweighted sample in 1982. Finally, we limit the sample to workers who were in such jobs at the beginning of the period (1968 or 1982). This left us with a sample of 162 individuals in 1968 and 219 in 1982 who were in the types of jobs that would be expected to be heavily affected by the technological revolution between 1982 and 1996.

For each sample, we then regressed the use of routine-cognitive skills in the later period (1982 or 1996) on age. The results are presented in table A.1. Comparing the two columns, from the constant terms we see that, relative to the earlier period, workers in the later period who started the period in routine-cognitive intensive jobs engaged in less routine-cognitive intensive work fourteen years later. Although for both periods, the slope coefficient is positive, indicating that older workers engaged in more routine-cognitive intensive work, the coefficient is only statistically significant in the later period. The point estimates suggest that workers who were less than 46 years old in 1982 reduced their use of routine-cognitive skills in 1996 relative to what similar workers in 1968 had done by 1982. The difference is statistically significant at the .05 level for each age in the 20-31 range.

Thus, as intuition and our model suggest, this small amount of evidence indicates that younger workers adjust more to shocks than do older workers. At the same time, we should not exaggerate this finding. The results for the two periods differ only at the .1 level, and the difference between the two slope coefficients does not reach significance at conventional levels.

Table A.1
Subsequent Routine-Cognitive Task Intensity Among Workers Initially in Routine-Cognitive Intensive Jobs

	1968-1982	1982-1996
Age	0.0276	0.0656**
	(0.0249)	(0.0260)
Constant	4.828***	3.092***
	(0.816)	(0.797)
Observations	162	219
R-squared	0.008	0.029

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1
PSID, Intensity measured 14 years later for individuals in routine-cognitive intensive jobs in 1968 or 1982

Ben-Porath meets Lazear: Online Appendix

Costas Cavounidis* and Kevin Lang†

December 19, 2018

1 The General Two-Period Model

In this section we generalize the basic two-period model. We reproduce several results in a more general setting.

There exist N different skills, and a worker is endowed with a skill vector $S \geq 0$ so that her ability in skill n is S_n.

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, non-singleton, 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 $\mathcal{J} \equiv \{J \in \mathbb{R}^N_+ | F(J) \leq 0\}$. We further assume that $\nabla^2 F(J)$ is positive definite when $\nabla F(J) >> 0$.

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

$$W(J, S) = (AJ)^T S \tag{1}$$

where A is a diagonal matrix. We can scale A and $\mathcal{J}$ so that A_n is to be interpreted as the maximum weight that any job puts on skill n.

We additionally assume that if some $J_n = 0$ and $F(J) = 0$ then $\partial F(J)/\partial J_n \leq 0$. In other words, if there is a job that does not use 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 tells us how different skills can be combined to produce.

*Department of Economics, University of Warwick. email: c.cavounidis@warwick.ac.uk
†Department of Economics, Boston University and NBER/IZA. email: lang@bu.edu.

1.1 Single job selection

A chooses a job from $\mathcal{J}$ to maximize her wage solves the constrained program

$$\max_{J\in\mathcal{J}} W(J,S) = \max_{J\in\mathcal{J}} (AJ)^T S. \qquad (P_1)$$

We can form the associated Lagrangian

$$\mathcal{L}_1 = (AJ)^T S - \lambda F(J). \qquad (L_1)$$

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

$$AS - \lambda \nabla F(J^*) = 0 \qquad (2)$$

$$F(J^*) = 0. \qquad (3)$$

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

1.1.1 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 A1 The weight that the optimal job places on a skill is increasing in that skill:

$$\frac{\partial J_n^*}{\partial S_n} > 0. \qquad (4)$$

Proof. From Fiacco (1976), we have that J^* is differentiable with respect to S. Taking a derivative of (2) with respect to S_n, we have

$$[0,...0,A_n,0,...,0]^T - \frac{\partial \lambda}{\partial S_n}\nabla F(J^*) - \lambda \nabla^2 F(J^*)\frac{\partial J^*}{\partial S_n} = 0. \qquad (5)$$

Differentiating (3) with respect to S_n, we obtain $\frac{\partial J^*}{\partial S_n}^T \nabla F(J^*) = 0$. Premultiplying (5) with $\frac{\partial J^*}{\partial S_n}^T$ and using that fact, we have

$$A_n \frac{\partial J_n^*}{\partial S_n} - \frac{\partial \lambda}{\partial S_n} \frac{\partial J^*}{\partial S_n}^T \nabla F(J^*) = \lambda \frac{\partial J^*}{\partial S_n}^T \nabla^2 F(J^*) \frac{\partial J^*}{\partial S_n} \tag{6}$$

$$\lambda^{-1} A_n \frac{\partial J_n^*}{\partial S_n} = \frac{\partial J^*}{\partial S_n}^T \nabla^2 F(J^*) \frac{\partial J^*}{\partial S_n} \tag{7}$$

As $S >> 0$, we have that $\nabla F(J^*) > 0$ 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 $\frac{\partial J_n^*}{\partial S_n} > 0$ as required. ■

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 (4) we can see that the value is strictly convex in any particular skill. As her nth 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.

1.2 Investment

The worker now lives for two periods, chooses a job from $\mathcal{J}$ for each (denoted J_1 and J_2) and invests in skills in the first period, while discounting payoffs in the second period by a factor of β.

Her skills 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 her chosen *investment vector* I.[1] Starting with skills S, the worker will have a skill vector $S' = \Delta S + I$ in the second period. That is, Δ_n is the fraction of the endowment in skill n that does not depreciate by period 2 due to aging. The rate of depreciation may differ among skills.

The cost of the investment I is $C(I)$, payable in period 1. 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 for each n, $\partial C(I)/\partial I_n$ is unbounded above.

The worker solves the problem

$$\max_{I \in \mathbb{R}_+^N,\ J_1, J_2 \in \mathcal{J}} J_2^T A S + \beta J_1^T A(\Delta S + I) - C(I) \tag{P_{12}}$$

[1] Diagonality of Δ 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.

For this problem, we form the Lagrangian

$$\mathcal{L}_{12} = J_1^T AS + \beta J_2^T A(\Delta S + I) - C(I) - \lambda F(J_2) - \mu F(J_1). \tag{L_{12}}$$

Clearly, the part related to J_1 is entirely separable from the rest and follows the discussion of job choice with exogenous skills above. The rest of the problem becomes

$$\max_{I \in \mathbb{R}_+^N,\ J_2 \in \mathcal{J}} \beta J_2^T A(\Delta S + I) - C(I) \tag{P_2}$$

$$\mathcal{L}_2 = \beta J_2^T A(\Delta S + I) - C(I) - \lambda F(J_2). \tag{L_2}$$

The Lagrangian (L_2) is solved with the first order conditions[2]

$$\begin{aligned} \beta A(\Delta S + I^*) - \lambda \nabla F(J_2^*) &= 0 && (8) \\ \beta J_2^* A - \nabla C(I^*) &= 0 && (9) \\ F(J_2^*) &= 0 && (10) \end{aligned}$$

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

$$H(S, I^*, J_2^*) = \begin{bmatrix} -\lambda \nabla^2 F(J_2^*) & \beta A \\ \beta A & -\nabla^2 C(I^*) \end{bmatrix} \tag{11}$$

1.2.1 Investment and Skill Persistence

Our assumptions that putting a small weight (rather than 0) on some skill affects the constraint little, plus that the cost of a little investment (rather than 0) in any skill is cheap, jointly lead to persistence in skills. We show that a worker will never abandon investment in a skill 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.

[2] Equation (9) has a solution as the additive separability of $C(I)$ across its components, the fact that $\nabla C(0) = 0$, and the continuity and unboundedness above of each $\partial C / \partial I_n$ give us that ∇C is a bijection.

Proposition A2 A worker always continues to invest in any skill she already possesses:

$$S_n > 0 \Rightarrow I_n > 0. \tag{12}$$

Proof. Substituting for J_2^* using (9) in (8) and (10) we have

$$\beta A(\Delta S + I^*) - \lambda \nabla F(\beta^{-1} A^{-1} \nabla C(I^*)) = 0 \tag{13}$$

$$F(\beta^{-1} A^{-1} \nabla C(I^*)) = 0 \tag{14}$$

From (13), $S_n > 0$ implies that $\frac{\partial}{\partial J_n} F(J) > 0$. From this and the assumption that $J_n = 0 \Rightarrow \frac{\partial}{\partial J_n} F(J) \leq 0$, we have that $\frac{\partial}{\partial I_n} C(I^*) > 0$; but as we've assumed that $\nabla C(0) = 0$ and that C is additively separable , it must be that $I_n^* > 0$ to satisfy $\frac{\partial}{\partial I_n} C(I^*) > 0$. ■

As all skills the worker has any initial ability with will be given weight in period 2, 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.

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. $J_2^* >> 0$, $I^* >> 0$. Then we have

Proposition A3 Investment is weakly increasing in the existing endowment of a skill:

$$\frac{\partial I_n^*}{\partial S_n} \geq 0. \tag{15}$$

Proof. First, we substitute for J_2^* using (9) in the block Hessian:

$$H = \begin{bmatrix} -\lambda \nabla^2 F(\beta^{-1} A^{-1} \nabla C(I^*)) & \beta A \\ \beta A & -\nabla^2 C(I^*) \end{bmatrix} \tag{16}$$

Differentiating (13) and (14) with respect to S_n, recalling that Δ and A are diagonal and suppressing functional arguments, we have

$$[0, .., 0, \beta A_n \Delta_n, 0, .., 0]^T + \beta A \frac{\partial I^*}{\partial S_n} - \frac{\partial \lambda}{\partial S_n} \nabla F - \lambda \beta^{-1} \nabla^2 F A^{-1} \nabla^2 C \frac{\partial I^*}{\partial S_n} = 0 \tag{17}$$

$$(\nabla F)^T A^{-1} \nabla^2 C \frac{\partial I^*}{\partial S_n} = 0. \tag{18}$$

Premultiplying (17) with $(\frac{\partial I^*}{\partial S_n})^T \nabla^2 C A^{-1}$ and recalling $\nabla^2 C$ is diagonal, we get

$$\beta \Delta_n \frac{\partial I_n^*}{\partial S_n} \frac{\partial^2 C}{\partial I_n^{*2}} + \beta \left(\frac{\partial I^*}{\partial S_n} \right)^T \nabla^2 C \frac{\partial I^*}{\partial S_n} = \lambda \beta^{-1} \left(\frac{\partial I^*}{\partial S_n} \right)^T \nabla^2 C A^{-1} \nabla^2 F A^{-1} \nabla^2 C \frac{\partial I^*}{\partial S_n} \tag{19}$$

$$\Delta_n \frac{\partial I_n^*}{\partial S_n} \frac{\partial^2 C}{\partial I_n^{*2}} = -\beta^{-2} \left(\frac{\partial I^*}{\partial S_n} \right)^T \nabla^2 C A^{-1} \Big[-\lambda \nabla^2 F + \beta^2 A (\nabla^2 C)^{-1} A \Big] A^{-1} \nabla^2 C \frac{\partial I^*}{\partial S_n} \tag{20}$$

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 (20) as a whole is nonnegative. As $\Delta_n > 0$ and $\frac{\partial^2 C}{\partial I_n^{*2}} > 0$, we thus have $\frac{\partial I_n^*}{\partial S_n} \geq 0$.
■

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 (4) and the fact the J_1-part of (P_{12}) is identical to (P_1), we also have

$$\frac{\partial J_{1,n}^*}{\partial S_n} \geq 0 \tag{21}$$

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.[3]

1.3 Comparative Statics

1.3.1 Depreciation

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

[3] 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.

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

1.3.2 Age

Proposition 2 from the main paper fully carries over to the general setting; the proof used there does not rely on the specifics of the functional forms.

2 References

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