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

Costas Cavounidis† and Kevin Lang‡

June 14, 2018

Abstract

We provide microfoundations for dynamic skill formation with a model of investment in multiple skills, and jobs placing different weights on skills. We show that credit constraints may affect investment even when not binding with equality. 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. In simulations, we show that a shock to skill productivity that benefits new workers but lowers one skill’s value can have large adverse and possibly discontinuous effects even on relatively young workers.

1 Introduction

Economists increasingly understand the evolution of jobs and wages through models in which the reduced value of routine cognitive skills has shifted workers into jobs requiring either abstract reasoning or nonroutine manual skills. Either implicitly or explicitly, these models assume a hierarchy of skills, with the most skilled workers working in jobs requiring abstract reasoning and the least skilled in manual labor, based on the principle of comparative advantage. In reality, most jobs require a variety of different skills. For instance, many jobs, including many medical and engineering jobs, kindergarten and elementary teaching, and some athletic positions require significant levels of both manual labor and abstract reasoning relative to their use of routine cognitive skills.

*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 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.
At the same time, the literature on job mobility treats workers as searching for or randomly finding better matches. Neal (1999), for example, models workers as first searching for a good career match, and then searching for a good firm match within that career. While undoubtedly, the evolution of careers is determined in part by happenstance, a great deal is not. Workers deliberately invest in packages of skills in order to progress in their careers.

This means 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. This is unsurprising but can only be addressed in an ad hoc fashion using existing models. Equally important, we would expect, and our model implies, that how the worker invests in response to a negative shock to routine cognitive work depends on her previously accumulated skills. Intuitively, we would expect those who already have relatively strong abstract skills to invest more in such skills and move to jobs making greater use of them. No current model captures this.

Similarly Sviatschi (2018) shows how early experience in the illegal side of the cocoa industry increases later activity in the illegal drug trade and does so by more than could be predicted by reduced school attendance. Again, a single skill model does not predict this. Rather the children build on their parents’ early investment in their knowledge of the illegal drug industry. The early investment builds on itself in ways that we make precise. Similarly Arora (2018) finds that decreasing punishment for crimes committed at a relatively young age increases criminality at older ages that do not experience the reduced punishment, consistent with a model of heterogeneous skills.

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 but drops the assumption that the sum of the weights must be one. We place this in the context of lifetime investment in the spirit of Ben-Porath (1967). Of course, allowing for heterogeneous skill investment might merely be a technical issue that tells us little that is new or surprising. But the model provides us with new insights.

First, we provide a novel explanation for why firms help workers invest in general skills.

\footnote{An earlier version of this paper (Cavounidis and Lang 2017) also included limited results for a much larger class of models.}
and why workers frequently invest in skills that they do not intend to use much. 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. As workers cannot commit to an investment plan, optimal contracts commit the firm to overinvest in skills useful at the current job, relative to the first best, in order to reduce 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 painters. 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. Under plausible assumptions, the internal rate of return on this investment relative to becoming a painter is in excess of 11 percent. Without question, there are many reasons that a worker might choose to become

---

2 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 Baseball 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.


5 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
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), indicating that credit constraints might, in fact, have been an important reason that workers did not go into plumbing. 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.\(^6\)

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. More generally, we 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 when there are three skills. Two effects 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.

---

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.

\(^6\)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.
The horizon effect captures why younger workers were more likely than older workers to exit from routine cognitive jobs following the decline in the value of routine skills (Autor and Dorn 2009). Assuming, as is plausible, that college-educated workers in routine jobs had a higher stock of abstract skills, the inertia effect explains why college-educated workers were more likely than high-school educated workers to shift from routine to abstract jobs when routine skills declined in value (Autor and Dorn 2009).

The horizon effect is also present in the Ben-Porath model. The inertia effect, however, cannot show up in a model with homogeneous human capital, and both effects can only be imperfectly discussed in the context of a two-period model. As investment is front-loaded, the inertia effect ‘kicks in’ early. In fact, we show that due to this front-loading of investment, 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 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 and also assumes that investment in a skill occurs only by diverting time using a skill to investing in that skill. This makes it impossible for a worker to invest in an entirely new skill and therefore less useful for thinking about shocks or credit constraints. Altonji (2010), in particular, emphasizes the need for a research agenda that recognizes that skill is multidimensional and that jobs differ in their requirements.
2 The Two-Period Model

The working paper version of this paper discusses a more general form of our model which we summarize in footnote 8. We will make occasional reference to results that we can prove more generally. In this paper, we restrict ourselves to what the working paper terms the diagonal model.

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.

The worker chooses a job $J \in \mathbb{R}^N_+$ which must satisfy

$$\sum_n J_n^\sigma \leq 1$$

(1)

where $\sigma > 1$. At job $J$, a worker with skills $S$ produces

$$\sum_n A_n J_n S_n$$

(2)

where $A >> 0$. Workers will choose their job to maximize production (or wages) given $S$. Optimal choice implies a job $J$ such that $\sum_n J_n^\sigma = 1$. $A$ is the productive efficiency of different skills, representing how the current technology uses each skill. This form gives $A_n$ 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 separation of tasks from the production technology.

Changes in $A$ can be thought of capturing changes over time in the value of a skill; individuals with different $A$s have different aptitudes. Various basketball skills are less

---

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

8 A worker is endowed with a skill vector $S \in \mathbb{R}^N_+$ so that her level of ability in skill $n$ is $S_n$. The worker chooses a job from $J \subseteq \mathbb{R}^N_+$, 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, convex, compact and can be described in terms of a strictly convex, smooth function $F: \mathbb{R}^N \rightarrow \mathbb{R}$ and the positive orthant so that $J \equiv \{J \in \mathbb{R}^N_+ | F(J) \leq 0 \}$. A worker with skill vector $S$ at job $J$ receives a wage $W(J,S) = (AJ)^T S$ where $A$ is a diagonal matrix. In the first period, the worker also chooses investment $I \geq 0$ and pays $C(I)$, where $C$ has a diagonal and positive definite Hessian. Skills from period 1 are carried over to period two by way of an $N \times N$ diagonal, positive definite non-depreciation matrix $\Delta \leq \mathbb{I}_{N \times N}$ so that starting with skills $S$, the worker has a skill vector $S' = \Delta S + I$ in the second period.
valuable for an unusually short worker than for one who is unusually tall. Similarly, knowing how to shoe horses is a skill that has declined in value even though the job persists.

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 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

$$\Sigma_n A_n J_n S_n - \lambda (\Sigma_n J_n^\sigma - 1).$$

Maximizing gives the first-order conditions

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

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

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

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

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

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

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

$$\sum A_n J_{1,n} S_n + \beta \sum A_n J_{2,n} (\delta S_n + I_n) - \sum I_n^\rho - \lambda (\sum J_{1,n}^\sigma - 1) - \mu (\sum J_{2,n}^\sigma - 1)$$

(8)

where $J_1$ and $J_2$ are the jobs in periods 1 and 2, $\beta$ is the discount factor and $\delta$ 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

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

(9)

which yields the first-order conditions for $I$:

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

(10)

and arrive at

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

(11)

For a corner solution with $I_n = 0$ to exist requires $S_n = 0$ and that $(\rho - 1)(\sigma - 1) \leq 1$.\(^9\)

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 is isomorphic to the Ben-Porath model of post-schooling investment.

\(^9\)To see this, notice that when $S_n = 0$, the left hand side of (11) goes to 0 slower 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.
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 a \( m \neq n \) s.t. \( S_{m} \neq 0 \), then \( I'_{n} > I_{n} \). Furthermore, the period-2 weight on skill \( n \) is increasing in the endowment of skill \( n \); \( J_{2,n}' > J_{2,n} \).

**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**.\(^{10}\)

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 even though the investment itself does not depreciate. Instead, because their initial skill depreciates, workers know they 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 \( \beta \) to change the ‘length’ of the second period. Increasing \( \beta \) (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 \( \beta \) and \( I'^* \) be a solution to the problem with discount factor \( \beta' \). Then \( C(I^*) > C(I'^*) \).

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.

\(^{10}\)Note that the proposition does not say that workers always specialize. The working paper derives conditions under which workers will tend to specialize and when workers (with the same \( A \) matrix) will all 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.
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 outlined in footnote 8. 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 $\beta$.

**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 $\beta$. 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 $I^* (\cdot)$ maps each $\beta$ into the set of optimal investments for that $\beta$, then $I^*$ does not admit a continuous singlevalued selection.\(^{11}\)

Informally, $\beta$ 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 $\beta$, 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.

### 3 Extensions of the Two-Period Model

#### 3.1 Credit Constraints

With a single skill, we could fully replicate the 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. However, our focus is on post-schooling investments and job choice. 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.

\(^{11}\)A singlevalued selection from a correspondence $F : X \rightrightarrows Y$ is a function $f : X \to Y$ such that $f(x) = y \implies y \in F(x)$. 

10
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\(^{12} \) 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; when skill productivity varies sufficiently, there are always endowments and budget constraints such that the budget constraint does not bind with equality but affects the optimum. Let \( I^*_c(\cdot) \) denote the function mapping the endowment \( S \) into optimal investment when the constraint is \( c \).

**Proposition 4** 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. That is, if \((\rho - 1)(\sigma - 1) < 1\) there is an \( x > 0 \) such that if there are \( n, m \) with \( A_n \cdot x < A_m \), then there exist \( S, c \) such that \( C(I^*_c(S)) < c \) and \( C(I^*(S)) > c \).

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

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

\(^{12}\)With the convexity inequality strict for all but parallel skill vectors.
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 until \( c = 4 \). Beyond that, further investment is inefficient, and relaxing the constraint further has no effect on net output.

### 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 (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.

This inefficiency can be ameliorated if firms have the ability to offer training as part of the first-period wage offer: an investment \( I^F \) in the worker’s skills, at a cost \( C(I^F) \) to the firm. 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** Suppose \((\sigma - 1)(\rho - 1) > 1\). Suppose the worker with skill endowment \( S >> 0 \) does not move in period 2, but would absent the mobility cost. Then the optimal contract \((J, I^F, w_0)\) satisfies \( I^W(I^F) >> I^*(J) \).

The intuition for this result is simple: to the worker, investments in different skills are substitutes. The worker’s incentives are to overinvest in certain skills relative to the current

\(^{13}\)Sometimes this will be a correspondence; if \((\rho - 1)(\sigma - 1) > 1\), it is guaranteed to be a function.
job’s weights in order to improve the outside option for bargaining purposes. Then, by increasing investment in other skills - those not overinvested in - the firm can dampen the worker’s incentives. The firm wants to commit to overinvest in these counterweight-skills, as at the appropriate level of investment for the current job, the direct effect of further investment on net production is a only a second-order loss, but the efficiency gain from reducing 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.\(^{14}\)

### 4 Continuous Time

To more fully explore the dynamics of skill investment, 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.\(^{15}\) 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 \(J = \{J \in \mathbb{R}^N_+ | \sum_n J_n^\sigma \leq 1\}\) so that her time-\(t\) instantaneous production is \(\sum_n A_n J_n(t) S_n(t)\). Skills depreciate at rate \(\Delta\), 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

---

\(^{14}\)A much weaker sufficient condition is that \(I^W(\cdot)\) is a continuous function on some relevant subset of its domain, but it is far harder to state in terms of primitives.

\(^{15}\)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, provided the firm has some bargaining power and the wage is influenced by the worker’s investment decision, any mobility cost will lengthen the time the worker spends at a particular job. This, in turn, creates the conditions for firm investment in the worker’s skills and overinvestment.
solving

\[
\max_{J: [0, T] \to J, \ I: [0, T] \to \mathbb{R}_+^N} \int_0^\infty e^{-rt} \left[ \sum_n A_n J_n(t) S_n(t) - \sum_n I_n(t)^\rho \right] dt
\]  
\text{s.t.} \quad \frac{d}{dt} S(t) = -\Delta S(t) + I(t) \quad \text{(14)} 
\quad S(0) = S_0. \quad \text{(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

\[
\max_{I: [0, T] \to \mathbb{R}_+^N} \int_0^\infty e^{-rt} \left[ \left( \sum_n (A_n S_n(t))^{\frac{\sigma-1}{\sigma}} \right)^{\frac{\sigma-1}{\sigma}} - \sum_n I_n(t)^\rho \right] dt
\]  
\text{s.t.} \quad \frac{d}{dt} S(t) = -\Delta S(t) + I(t) \quad \text{(17)} 
\quad S(0) = S_0. \quad \text{(18)}

We therefore construct the Hamiltonian

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

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

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

along with the motion equations for skills

\[
\frac{d}{dt} S(t) = -\Delta S(t) + I(t), \quad \text{(21)}
\]
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

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

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

![Figure 2: The ‘Ben-Porath’ case of proportional skill evolution.](image)

\( A_1 = 2^2, A_2 = 1, \rho = 2, \sigma = 2.5, S_0 = (20, 10) , \Delta = r = .05 \)

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 paths 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 is the point at which the two are equal.

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

![Figure 3: Initial Investment by Experience at Time of Shock](image)

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)
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 10 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. 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. If workers typically 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 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.

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. Much, but not all, of the increased weight on Abstract skill reflects the greatly increased value of that skill in all jobs rather than a large shift towards investment in Abstract skill.

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](image)

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 Manual skill. As a consequence, the individual shocked at 10 years of experience suffers an earlier but smaller output shock. Compared to a similar person suffering a shock at 20 years of experience, she has higher output at nearly every later experience level. When the shock hits the similar person at 30 years, the one shocked at 10 years has already recovered sufficiently to have higher net output. The person shocked at 20 years of experience fares almost as badly in the last 10 years of work as the person shocked at 30 years.

More generally, in this example in which the shock is in a sense positive, a worker who begins her career just as the shock hits will benefit. One who ends her career just as the shock hits will be unaffected. By continuity there will be a range of low experience levels at
which the effect of the shock will be positive. We expect, but have not shown, that the effect of the shock is U-shaped. 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. Here we find that a worker shocked at 20 years of experience must reduce her consumption by almost half (48 percent) relative to what she had anticipated. In contrast, workers shocked at 10 and 30 years of experience must reduce their consumption by 36 percent and 37 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. The same worker who spends her entire career after the shock will earn 6.4 percent more over her lifetime than if she finished her career before the shock hit. Yet even a worker who was only five years into her career never recovers from the shock and suffers a decline of about one-fifth in future consumption. This is because skill investment is extremely front-loaded to allow for longer exploitation time, so the loss is great even when the shock hits early. While ours is an example, not a calibration exercise, we find this duration and magnitude of the effect striking.

4.4 Why older workers adjust less

Table 1: Different skill paths in response to shock

<table>
<thead>
<tr>
<th>Skill levels when shocked</th>
<th>Time remaining</th>
<th>Twenty years after shock</th>
</tr>
</thead>
<tbody>
<tr>
<td>24.09 47.12 9.54</td>
<td>30</td>
<td>46.38 32.80 45.80</td>
</tr>
<tr>
<td>24.09 47.12 9.54</td>
<td>20</td>
<td>33.93 30.38 22.73</td>
</tr>
<tr>
<td>29.69 68.45 7.66</td>
<td>30</td>
<td>52.88 42.28 32.02</td>
</tr>
<tr>
<td>29.69 68.45 7.66</td>
<td>20</td>
<td>35.76 41.80 14.14</td>
</tr>
</tbody>
</table>

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 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, 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 21.3 difference in the stocks of Routine skill would have fallen to 7.8
had they invested at the same rate. The difference of 9.5 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. 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
years later her job puts the most weight on the Abstract 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.
Table 2: Job choice responses to shock

<table>
<thead>
<tr>
<th>Job’s weights when shocked</th>
<th>Time remaining</th>
<th>Twenty years after shock</th>
</tr>
</thead>
<tbody>
<tr>
<td>.43</td>
<td>.89</td>
<td>.12</td>
</tr>
<tr>
<td>.43</td>
<td>.89</td>
<td>.12</td>
</tr>
<tr>
<td>.34</td>
<td>.94</td>
<td>.06</td>
</tr>
<tr>
<td>.34</td>
<td>.94</td>
<td>.06</td>
</tr>
</tbody>
</table>

Twenty years later, her job puts the most weight on the Manual skill, and has not increased the weight on the Abstract skill all that much.

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 bargaining position. 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 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


A Proofs

A.1 Proof of Proposition 1

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

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

which, cancelling terms, implies

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

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 $\beta$ and $I''$ is one with $\beta'$. Then, optimality implies

\[ \beta V(\delta S + I^*) - C(I^*) \geq \beta V(\delta S + I'') - C(I'') \]  
\[ \beta' V(\delta S + I'') - C(I'') \geq \beta' V(\delta S + I^*) - C(I^*) \]

so that, after some manipulation

\[ (C(I'') - C(I^*)) (\beta' - \beta) \geq 0 \]  
\[ C(I^*) \geq C(I'') \]

Now, supposing $C(I^*) = C(I'')$ for contradiction, we have $V(\delta S + I^*) = V(\delta S + I'')$ or else one of the objective functions is improvable. Then, from the first order condition for $I''$ we have $\beta' \nabla V(\delta S + I'') = \nabla C(I'')$ and thus $\beta \nabla V(\delta S + I'') >> \nabla C(I'')$. This means that $I''$ improves on the objective over $I^*$ in the $\beta$ problem, so $I^*$ is not a maximizer. Hence, it must be the case that $C(I^*) > C(I'')$.

A.3 Proof of Proposition 3

Fix $\rho, \sigma$ 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 $\beta$, the worker does invest in skills other than 1. If the worker invests only in skill 1, she solves

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

for optimal investment

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

so that the value of the worker’s problem is

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

24
whereas investing only in a skill \( n \) and choosing a job putting weight only on skill \( n \) similarly yields a value of

\[
A_n^{\frac{\sigma}{\sigma-1}} (\rho \beta^{\frac{\sigma}{\sigma-1}} - 1) \rho^{-\frac{1}{\sigma-1}}.
\]  

(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

\[
\beta A_1 \delta + A_1^{\frac{\sigma}{\sigma-1}} (\rho \beta^{\frac{\sigma}{\sigma-1}} - 1) \rho^{-\frac{1}{\sigma-1}} < A_n^{\frac{\sigma}{\sigma-1}} (\rho \beta^{\frac{\sigma}{\sigma-1}} - 1) \rho^{-\frac{1}{\sigma-1}}
\]

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

Thus, if \( I_m^* \) 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

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

(34)

which is less than

\[
\beta \frac{A_m^{\frac{\sigma}{\sigma-1}} I_m^{\frac{1}{\sigma-1}}}{(A_1 \delta)^{\frac{1}{\sigma-1}}} - \rho I_m^{\rho-1}
\]

(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, \( I_m^* \) cannot be continuous.

A.4 Proof of Proposition 4

**Proof.** Pick some \( x > 0 \) such that \( x^{\frac{\rho}{\sigma}} (1 - \frac{\rho}{\beta(\rho-1)} x^{\frac{\sigma-1}{\rho}}} \) > 1. As \( (\sigma - 1)(\rho - 1) < 1 \), this is possible. Take any \( A \) 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 \( (A_2 A_1) \frac{\rho}{\beta(\rho-1)} (1 - k^{\frac{\rho}{\beta(\rho-1)}} (\frac{A_2}{A_1})^{\frac{(\sigma-1)(\rho-1)-1}{\sigma-1}}) > 1 \). Set \( c = (\frac{\beta A_1}{\rho})^{\frac{\rho}{\sigma}} k^\rho \) and \( S = (\frac{A_2}{A_1})^{\sigma} c^{\frac{1}{\sigma}} 0, 0... \). First, suppose that the constraint \( c \) binds. Consider the problem of allocating \( c \) across the different skills; each \( n > 1 \) is allocated \( I_n \) and the remainder goes to skill 1.

\[
\max_{I_{bind}} \left[ I_1(S_1 + (c - \sum_{m>1} I_m^{\frac{1}{\sigma}}))^{\frac{\sigma}{\sigma-1}} + \sum_{m>1} \left( A_m I_m \right)^{\frac{\sigma}{\sigma-1}} \right]^{\frac{\sigma-1}{\sigma}}
\]

25
then (ignoring the outer power) the first order condition with respect to $I_n$ is

$$\frac{\partial \Pi_{\text{bind}}}{\partial I_n} = \frac{\sigma}{\sigma - 1} \left( A_n^{\frac{\sigma}{\rho - 1}} I_n^{\frac{1}{\sigma - 1}} - A_1^{\frac{\sigma}{\rho - 1}} [S_1 + (c - \sum_{m>1} I_m^\rho) \frac{1}{\rho}]^{\frac{1}{\sigma - 1}} (c - \sum_{m>1} I_m^\rho) \frac{1-\rho}{\sigma} I_n^\rho \right)$$

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

$$\frac{\partial \Pi_{\text{bind}}}{\partial I_n} = \frac{\sigma}{\sigma - 1} I_n^{\frac{1}{\sigma - 1}} \left( A_n^{\frac{\sigma}{\rho - 1}} - A_1^{\frac{\sigma}{\rho - 1}} [S_1 + (c - \sum_{m>1} I_m^\rho) \frac{1}{\rho}]^{\frac{1}{\sigma - 1}} (c - \sum_{m>1} I_m^\rho) \frac{1-\rho}{\sigma} I_n^\rho \right)$$

$$< \frac{\sigma}{\sigma - 1} I_n^{\frac{1}{\sigma - 1}} \left( A_n^{\frac{\sigma}{\rho - 1}} - A_1^{\frac{\sigma}{\rho - 1}} S_1^{\frac{1}{\rho}} e^{\frac{1-\rho}{\rho} (\rho - 1)(\sigma - 1) - 1} \right) = \frac{\sigma}{\sigma - 1} I_n^{\frac{1}{\sigma - 1}} \left( A_n^{\frac{\sigma}{\rho - 1}} - A_1^{\frac{\sigma}{\rho - 1}} S_1^{\frac{1}{\rho}} e^{\frac{1}{\rho} (\sigma - 1) - 1} \right) =$$

$$\frac{\sigma}{\sigma - 1} I_n^{\frac{1}{\sigma - 1}} (A_n^{\frac{\sigma}{\rho - 1}} - A_2^{\frac{\sigma}{\rho - 1}}) \leq 0.$$

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

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

so that $I_1^* = \left( \frac{\beta A_1}{\rho} \right)^{\frac{1}{\rho - 1}}$ for an expenditure of $\left( \frac{\beta A_1}{\rho} \right)^{\frac{\sigma}{\rho - 1}} < c$. 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 $\beta A_2 I_2^\rho - I_2^\rho = \left( \frac{\beta A_2}{\rho} \right)^{\frac{\rho}{\rho - 1}} (\rho - 1)$. We want to show this is greater than the constrained payoff, that is,

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

which, after rearrangement and division by $\left( \frac{\beta A_1}{\rho} \right)^{\frac{\sigma}{\rho - 1}} (\rho - 1)$ 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, $C(I^*(S)) > c > C(I_2^*(S))$. ■
A.5 Proof of Proposition 5

Step 1: *J is not optimal in period 2 in the absence of a mobility cost.*

Suppose for contradiction that it is. Then the optimal investment in the absence of the mobility cost, $I^*$, is also optimal when there is a mobility cost as $I^W(0) = I^*$. From the optimality of $J$ with the mobility cost, we have

$$
\nabla F(J) \| A^d(S + \beta(\delta S + I^*))
$$

where $A^d$ denotes the diagonal matrix with diagonal elements $A$.

From the optimality of $J$ in period 2 without the mobility cost, we have

$$
\nabla F(J) \| A^d(\delta S + I^*)
$$

Combining the two, we have

$$
\nabla F(J) \| A^dS
$$

so that $J$ is also optimal in period 1 absent the mobility cost. But by assumption, the worker moves absent the mobility cost, a contradiction. Therefore $J$ cannot be optimal in period 2 absent the mobility cost.

Step 2: *No skill is underinvested in, and at least one is overinvested in.*

Given $I^F$, the worker chooses $I^W$ to maximize

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

(36)

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

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

(37)

The first order condition for $I^W_n$ (when the $n$’th constraint does not bind and thus $I^W_n > I^F_n$) is

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

(38)
which can be rewritten as

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

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 \(V\).

If for some \(m\) such that \(I_m^F = I_m^W(\bar{I}_m)\) we had \(\frac{dV(\delta S + I^W)}{dI_m^W} \leq 0\) then for any \(n\) such that \(I_n^W(\bar{I}_m) > I_n^F\), we would have \(\frac{dI_m^W(\bar{I}_m)}{dI_m^W} \geq 0\) and thus as \(V(\delta S + I^W) = (\sum_k (A_k(\delta S_k + I_k^W)))^{\frac{\sigma-1}{\sigma}}\), \(V\) would have to increase, a contradiction. Thus, we have shown that if \(I_n^W(\bar{I}_m) > I_n^F\) and \(I_m^W(\bar{I}_m) = I_m^F\) then \(dI_m^W(\bar{I}_m)/dI_m^F < 0\) (\(\star\)).

Ex ante, the contract maximizes

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

If for all \(n\) we have \(I_n^W(\bar{I}_m) \leq I_n^*(J)\) then \(J^*(\delta S + I^*(J)) \leq J\) and therefore either the constraint \(\Sigma_k I_k^* \leq 1\) does not bind and therefore a job where the worker is more productive both periods exists (a contradiction) or \(J\) is optimal in the second period in the absence of a moving cost, not the case by Step 1. Therefore there is an \(n\) for which \(I_n^W(\bar{I}_m) > I_n^*(J)\).

Suppose there is an \(n\) for which \(I_n^W(\bar{I}_m) < I_n^*(J)\). Then, setting \(\forall m, I_m^{F'} := \max\{I_m^W, I_m^*(J)\}\), from \(\star\) we have that \(I_n^W(I^F) = I_n^{F'}\). Therefore, \(I_n^{F'}\) improves the objective (40), a contradiction. So \(I_n^W(I^F) \geq I_n^*(J)\).

**Step 3: Every skill is overinvested in.**

Now suppose \(\exists n: I_n^W(I^F) = I_n^*(J)\). As \(I^W \geq I^*(J)\), there must be a \(m\) such that \(I_m^W(I^F) > I_m^*(J)\). Two cases are of interest.

**Case 1.** Suppose (39) holds for \(I_m^W\). Then, define \(I^F = I_m^W(\bar{I}_m)\); we have \(I_n^W(I^F) = I_n^W(I^F)\) and, as \(I^F\) is part of an optimal contract, so is \(I^F\) (albeit with a compensating period-1 wage).

We will consider increasing \(I_n^F\) to effect an increase in \(V\) and through it will implement a decrease in \(I_n^F\) without affecting \((\bar{I}_k)_{k \notin \{n,m\}}\).

We define the auxiliary function \(\bar{I}_m^F(\bar{I}_n^F)\) implicitly by

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

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 is decreasing in \(\hat{I}_m^F\) as \((\rho-1)(\sigma-1) > 1\). Therefore, we have \(d\bar{I}_m^F(\bar{I}_n^F)/d\bar{I}_n^F < 0\).
Consider now perturbing the optimal contract’s investment $\hat{I}^F$ by increasing $\hat{I}^F_n$ and lowering $\hat{I}^F_m$ along $\hat{I}^F_n(\cdot)$. As $V((\delta S_n + \hat{I}^F_n), (\delta S_m + \hat{I}^F_m), (\delta S_k + \hat{I}^F_k)_{k \notin \{n,m\}}) \geq V(\hat{I}^F)$ when $\hat{I}^F_n \geq \hat{I}^F_n$, we have that $I^W(\hat{I}^F_n, \hat{I}^F_m(\hat{I}^F_n), (\hat{I}_k)_{k \notin \{n,m\}}) = (\hat{I}^F_n, \hat{I}^F_m(\hat{I}^F_n), (\hat{I}_k)_{k \notin \{n,m\}})$.

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

$$\beta J^T A((\delta S_n + \hat{I}^F_n), (\delta S_m + \hat{I}^F_m(\hat{I}^F_n)), (\delta S_k + \hat{I}^F_k)_{k \notin \{n,m\}}) = C((\hat{I}^F_n, \hat{I}^F_m(\hat{I}^F_n), (\hat{I}_k)_{k \notin \{n,m\}})) \quad(42)$$

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

$$\beta A_n J_n + \beta A_m J_m \frac{d\hat{I}^F_n(\hat{I}^F_n)}{d\hat{I}^F_n} - \rho \hat{I}^F_n^{\rho - 1} - \frac{d\hat{I}^F_m(\hat{I}^F_n)}{d\hat{I}^F_n} \rho (\hat{I}^F_n(\hat{I}^F_n))^{\rho - 1}$$

$$= (\beta A_n J_n - \rho \hat{I}^F_n^{\rho - 1}) + \frac{d\hat{I}^F_m(\hat{I}^F_n)}{d\hat{I}^F_n} (\beta A_m J_m - \rho \hat{I}^F_m(\hat{I}^F_n))^{\rho - 1} \quad(43)$$

But by assumption $\hat{I}^F_n = I^*_n(J)$, so that $\beta A_n J_n - \rho \hat{I}^F_n^{\rho - 1} = 0$ and $\hat{I}^F_m(\hat{I}^F_n) = \hat{I}^F_m > I^*_m(J)$ so that $\beta A_m J_m - \rho \hat{I}^F_m(\hat{I}^F_n)^{\rho - 1} < 0$. Furthermore, we have that $\frac{d\hat{I}^F_m(\hat{I}^F_n)}{d\hat{I}^F_n} < 0$. Therefore evaluated at $\hat{I}^F_n$, the restricted objective function’s right derivative is positive. As a result, there exists a $\hat{I}^F_n > \hat{I}^F_n$ so that $(\hat{I}^F_n, \hat{I}^F_m(\hat{I}^F_n), (\hat{I}_k)_{k \notin \{n,m\}})$ improves the objective function (40) over the assumed maximizer $I^F$, a contradiction.

**Case 2.** Now suppose instead that (39) does not hold for any $I^W$. Consider again $\hat{I}^F = I^W(\hat{I}^F)$, which is again optimal under the hypothesis that $\hat{I}^F$ is. Define $\hat{I}^F_m(\hat{I}^F_n)$ by

$$V((\delta S_n + \hat{I}^F_n), (\delta S_m + \hat{I}^F_m), (\delta S_k + \hat{I}^F_k)_{k \notin \{n,m\}}) = V(\delta S + \hat{I}^F)$$

when $\hat{I}^F_n \geq \hat{I}^F_n$ is small enough for a solution to exist. In other words, $\hat{I}^F_n$ adjusts to $\hat{I}^F_n$ 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 $(\hat{I}^F_n, \hat{I}^F_m(\hat{I}^F_n))$ for $\hat{I}^F_n > \hat{I}^F_n$, skills $k \notin \{n,m\}$ stay constant. As $V$ has strictly positive (as $S >> 0$) and continuous partials, $d\hat{I}^F_n(\hat{I}^F_n)/d\hat{I}^F_n > 0$; 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

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Age</td>
<td>0.0276</td>
<td>0.0656**</td>
</tr>
<tr>
<td></td>
<td>(0.0249)</td>
<td>(0.0260)</td>
</tr>
<tr>
<td>Constant</td>
<td>4.828***</td>
<td>3.092***</td>
</tr>
<tr>
<td></td>
<td>(0.816)</td>
<td>(0.797)</td>
</tr>
<tr>
<td>Observations</td>
<td>162</td>
<td>219</td>
</tr>
<tr>
<td>R-squared</td>
<td>0.008</td>
<td>0.029</td>
</tr>
</tbody>
</table>

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