Kredler, M. (2014). Experience vs. obsolescence: A vintage-human-capital model. *Journal of Economic Theory*, 150, pp. 709–739

DOI: 10.1016/j.jet.2013.08.003

Experience vs. Obsolescence: A Vintage-Human-Capital Model*

Matthias Kredler†

6 May 2013

Abstract

I introduce endogenous human-capital accumulation into an infinite-horizon version of Chari & Hopenhayn's (1991) vintage-human-capital model. Returns to skill and tenure premia are highest in young vintages, where skill is scarcest and agents accumulate human capital fastest. As the vintage ages, the skill premium decreases and vanishes entirely upon vintage death. Workers run through cycles of human-capital accumulation: their wages rise as they accumulate skill, undergo downward pressure as the technology ages and finally drop sharply when the worker switches to a new technology. The results are in line with German linked employer-employee data: tenure premia are higher in young establishments, as well as in fast-growing industries, occupations and establishments. A calibration exercise suggests that human-capital accumulation is the most important determinant of workers' wage profiles, whereas changes to the price of skill and vintage productivity gains play a smaller quantitative role.

Keywords: vintage human capital, tenure-wage profiles
JEL codes: J01, E24

---

*This paper is part of my dissertation thesis at New York University. I thank Boyan Jovanovic, Jess Benhabib, Jonathan Eaton, Greg Kaplan, Silvana Melitsko, Gianluca Violante, two anonymous referees and participants at various seminars for discussion and comments.

†Departamento de Economía, Universidad Carlos III de Madrid, matthias.kredler@uc3m.es. Part of the research was funded by the Spanish Ministerio de Ciencia e Innovación, reference number SEJ2007-62908.

1 Introduction

Returns to skill vary substantially across industries, occupations and firms. This paper argues that this is what we should expect when skill is specific to technologies. The basic mechanism I propose is as follows. New technologies, in which skill is scarce, offer high returns to skill in order to provide incentives for rapid skill accumulation. In old technologies, however, there is an abundance of skilled workers, but firms in these technologies face problems filling vacancies at the entry level. Workers know that the technology is at risk of becoming obsolete and are thus reluctant to enter. In order to lure workers into these old technologies, firms have to compensate workers with higher entry wages.

To develop these ideas, I build a vintage-human-capital model with endogenous human-capital accumulation. As in Chari & Hopenhayn (1991), human capital is tied to a technology and is lost when the technology is phased out. In each vintage, different levels of human capital are used in production. Unlike in Chari & Hopenhayn's (1991) two-period overlapping-generations model, however, human-capital accumulation is endogenous and the possibly infinite lives of individuals allow for rich patterns in tenure-wage profiles (shown in figure 1).

In the model, workers run through cycles of skill formation. They enter into a vintage, accumulate skill there and finally re-locate to a new vintage once the technology becomes obsolete. If different skill levels are complementary, then workers enter all active vintages. This is true even for the oldest technologies because firms need to fill vacancies of low-skill workers. Since workers are ex-ante homogeneous, all technologies have to be equally attractive for workers at entry. This requires old technologies to pay higher entry wages in order to make up for the shorter duration of the career. This is apparent in the shortest earnings profiles in figure 1, which pertain to workers entering old technologies.

Skill accumulation is fastest in the newest vintages, in which skilled labor is scarce. For these workers earnings growth is fastest, as we see in the long profiles in figure 1. Entrants into frontier technologies have the lowest entry wages. They can reap the benefits from their skills over a long time, which makes these careers especially attractive. In equilibrium, entry into young technologies increases until the return of the career is equal to that of old technologies. Since skill is abundant in old technologies, human-capital accumulation is slowest and earnings profiles are flattest in old vintages.

We also see in figure 1 that many workers experience wage losses towards the end of their career. These are driven by *obsolescence*, the fact that the

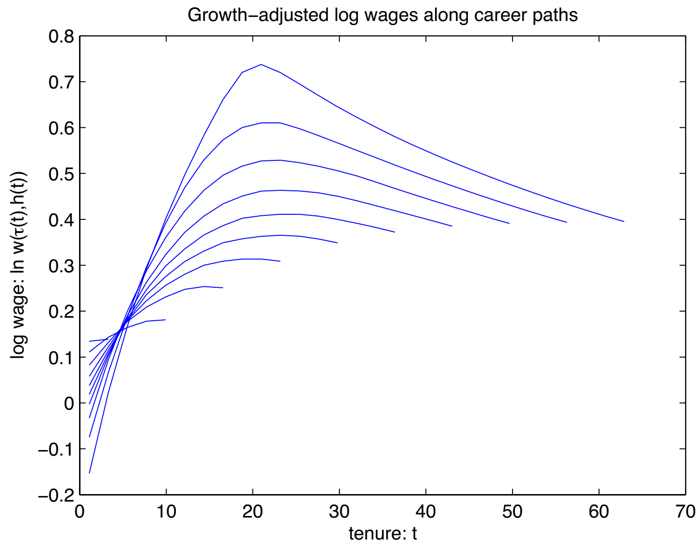

Figure 1: Tenure-earnings profiles over career

relative price of skill falls as the vintage ages. Upon re-locating to new technologies, workers again experience sharp drops in wages as they lose their vintage-specific human capital.

I find evidence supporting the model in a German matched employer-employee dataset when interpreting vintage age (in the model) as establishment age (in the data): young establishments have higher tenure premia, but pay lower average wages than old establishments. The model is also successful in predicting correlations of growth measures and the earnings structure: fast-growing industries, occupations and establishments display higher tenure premia than slow-growing ones, but pay wages on average.[1]

Finally, I calibrate the model to the cross-sectional tenure-earnings structure and the employment distribution across establishments. I find that there is complementarity between skills, but that this complementarity is weak. Workers' earnings growth is estimated to be due mainly to human-capital accumulation, whereas the relative scarcity of the skill and productivity gains in their technology play quantitatively minor roles. The calibration results indicate that an acceleration in the pace of technological change – as has been measured over recent decades by Cummins & Violante

[1] Only occupations are an exception to the latter statement: fast-growing occupations pay higher wages than slow-growing ones on average.

(2002) based on work by Gordon (1990) – leads to an intensification in skill accumulation and a rise in the premium on skill.

In relation to the previous literature, the model presented here is closest to Chari & Hopenhayn (1991), but differentiates itself by endogenous human-capital accumulation, workers' infinite life time and the resulting detailed predictions on tenure-wage profiles. In terms of predictions, a key difference is that workers in my setting experience wage losses both during their tenure in a vintage and upon switching between vintages, whereas workers always see their wage increase over time in their setting. Furthermore, since human-capital accumulation is exogenous in Chari & Hopenhayn's (1991), their setting cannot tell apart how much of wage growth is due to skill accumulation and which part is due to the scarcity of skill.

The paper is also related to the wider literature on vintage capital (see Boucekkine, de la Croix & Licandro, 2006, for an overview). Many issues, such as the incentives for the optimal scrapping time of a technology and the possibility of replacement echoes, are common to vintage capital and vintage human capital.

Similar to vintage models are ladder models of technology, in which all technologies are in principle available at all times but firms choose to push the frontier only little by little because they face high switching cost when jumping to technologies that are much more advanced than the one they are currently using. Parente (1994) studies a ladder model where agents face a trade-off between experience accumulation (following an exogenous learning curve) and obsolescence. In Violante's (2002) model, workers and machines of different vintages are matched in a frictional process. Again, workers accumulate skill according to a learning curve and experience skill losses that are increasing in the technological distance between machines. These human-capital losses induce wage losses upon job switches as in my model, but there is no obsolescence of skill: wages always increase during a worker's tenure in a vintage. Furthermore, these models do not predict that the premium on experience is higher in young technologies.

Nelson & Phelps (1966) have a view of human capital that is entirely different from, if not opposite to, vintage human capital. These authors posit that human capital facilitates the transmission and adoption of new ideas. Under this hypothesis, a high stock of human capital gives rise to frequent adoption of new technologies. Vintage-human-capital models take the opposite view: all skills are specific to a technology, so the adoption of new technologies destroys human capital. The Nelson-Phelps view suggests that an economy which adopts new technologies possesses high. The vintage-human-capital model in this paper, however, tells us that fast technological growth is disruptive in the sense that workers experience abrupt wage losses and an ensuing spell of steep wage growth when they re-locate to new vintages.[2]

Some theoretical analyses of the dynamics of organizations are also related to my model. Prescott & Boyd (1987) develop an overlapping-generations model of coalitions, where experienced and inexperienced workers face a trade-off between production of output and training of young workers. An important difference between their model and mine is that no reallocation of workers from obsolete technologies to new ones occurs in their model. In a more recent contribution, Garicano & Rossi-Hansberg (2008) model explicitly how tasks are shared within an organization and how organizations grow more complex over time. In my framework, however, technologies maintain the same structure over time; it is only the scarcity of different skills that changes.

An entirely different class of models that is able to generate increasing earnings profiles are search-and-matching models. Burdett & Coles (2003) show that firms in a frictional labor market optimally offer increasing wage schedules in order to prevent costly turnover. The predictions of the model presented here are different, however: in Burdett & Coles (2003) changes in *employer* are crucial for the determination of wages, whereas changes in the *technology* used by a worker are what matters in my setting.

The remainder of the paper is organized as follows: Section 2 presents the model, characterizes the competitive equilibrium and shows equivalence to a planner's problem. Section 3 presents quantitative results from a calibrated version of the model. Section 4 concludes and discusses potential further applications of the framework.

2 Model

2.1 Technology

Time is continuous. In every instant s, a new production technology (or *vintage*) arrives that is available to the agents in the economy for all $t \geq s$. I will either refer to the vintages by their birth date s or – especially in a stationary setting – identify them by their age $\tau \equiv t - s$. All vintages produce the same good y.

[2] Ljungqvist & Sargent (1998) use an ad-hoc specification of the idea that fast technological progress destroys human capital to argue that European labor-market institutions kept unemployment low in tranquil times but fail to do so in times of technological turbulence.

The production technology s uses labor inputs that are specific to this technology. The inputs are arranged on a hierarchy and indexed by $0 \leq h \leq 1$. The inputs on this ladder can be thought of as tasks that are increasing in difficulty; tasks with a higher index require more vintage-specific human capital. Section 2.2 will specify exactly how vintage-specific human capital is accumulated by workers.

The production function is supposed to capture the following notions: (i) newer vintages are more productive holding input ratios equal and (ii) the production function is complementary in its inputs. Specifically, I impose

$$Y(t,s) = e^{\gamma s}\tilde{Y}\big(n(t,s,\cdot)\big),$$

where $n(t,s,h)$ is the density function of workers at time t in vintage s with human capital h and $\tilde{Y}$ is a functional on the space of C^1 functions on $[0,1]$ with the following properties:

- Constant returns to scale (CRS): $\tilde{Y}(\lambda n) = \lambda \tilde{Y}(n)$.
- The Frechet derivative[3] $\tilde{w}(n)$ exists everywhere, is continuous in n and $\tilde{w}(n) > 0$ for all h, n.
- Weak concavity: $\tilde{Y}\big(\lambda n + (1-\lambda)n'\big) \geq \lambda\tilde{Y}(n) + (1-\lambda)\tilde{Y}(n')$ for all $0 \leq \lambda \leq 1$.

The first two properties imply that in a competitive setting, total wage payments exhaust production. An example for a functional that satisfies the above three conditions is the constant-elasticity-of-substitution (CES) aggregator

$$\tilde{Y}_{CES}[n(t,s,\cdot)] = \left[\int_0^1 f(h)n(t,s,h)^\rho dh,\right]^{1/\rho}, \tag{1}$$

where $\rho \leq 1$. The function $f(\cdot) \geq 0$ captures returns to human capital: I assume that returns to skill accumulation are non-negative ($f' \geq 0$) and weakly decreasing ($f'' \leq 0$). Total output in the economy at time t is given by $Y(t) = \int_{-\infty}^{t} Y(t,s)ds$.

At times, I will additionally invoke the following Inada condition:

Definition 2.1. *(Inada condition)* The production function is said to fulfill an Inada condition if $n(h) \to 0$ implies $\tilde{w}(h) \to \infty$ for all $h \in [0,1]$ and there

[3]Recall that the *Frechet derivative* is the generalization of the gradient vector from R^n to infinite-dimensional spaces. In this model, it is a wage function $\tilde{w} : [0,1] \to R^+$ which takes h as its argument. In the case of the CES aggregator in (1), it is given by the familiar $f(h)\tilde{Y}^{1-\rho}n(t,s,h)^{\rho-1}$.

is a unique element n^* on the interior of the unit simplex $\Delta = \{n : \int_h n = 1\}$ that maximizes output at $\bar{y} \equiv \max_{n \in \Delta} \tilde{Y}(n)$.

Note that n^* being the output-maximizing element of the unit simplex implies that marginal factor returns are equalized and that thus the wage schedule is constant, i.e. $\tilde{w}(n^*) = \bar{y}$. The CES aggregator in (1) above fulfills the Inada condition if and only if $\rho < 1$.

Competitive firms take wages for all labor inputs as given in each instant. Since the production technology is CRS, profits are zero for any t and s in equilibrium. Workers are paid their marginal product, so wages are $w(t, s, \cdot) = e^{\gamma s}\tilde{w}\big(n(t, s\cdot)\big)$.

2.2 Workers

There is a continuum of agents of mass one. Agents die at a constant rate δ. New agents are born at the same rate δ, keeping total population constant.

Agents have linear utility and discount the future at rate β, where $\beta + \delta > \gamma$. Each agent chooses a work life $\{s(t), h(t)\}_{0 \le t < \infty}$, which consists of a function $s(t)$ specifying the vintage the agent works in for all t and a function $h(t)$ specifying the task he performs at t in vintage $s(t)$. It is required that the vintage already exist at time t, i.e. $s(t) \le t$, and that $s(t)$ be a measurable function in t.[4]

As for human-capital accumulation $h(t)$, I require that a worker start her work life in position $h = 0$ when she enters the vintage; mathematically I impose that $h(\bar{t}) > 0$ only if there is an interval (a, b) around $\bar{t}$ such that $s(t) = \bar{t}$ for all $a < \bar{t} < b$.[5] There is no cost of switching between vintages. I will refer to a *career segment* (or short *career*) $l'(t)$ as the maximal open interval $(l'_0(t), l'_1(t))$ around an instant t that is entirely spent in one vintage.[6] If $l'_0(t) = t = l'_1(t)$, the career segment as an open interval is empty and we will not call this degenerate stay in a vintage a career segment. Since segments are open intervals and each of them contains a rational number, there can only be countably many of them in an agent's life.

To capture the notion that human-capital accumulation inside a vintage is costly, I require that the function h be differentiable on all segments and

[4] This specification allows for lives with more than countably many vintage changes; a relevant example for such a life is $s(t) = t$.

[5] This also means that a worker has to start at zero again even if he had worked in that vintage before but quit it at some point. This assumption is imposed for tractability; in equilibrium, workers would not want to return to vintages they have once left.

[6] Formally, define the end points as $l'_0(t) \equiv \inf\{a \le t : s(u) = s(t)\ \forall u \in [a, t]\}$ and $l'_1(t) = \sup\{b \ge t : s(u) = s(t)\ \forall u \in [t, b]\}$.

assume that the worker has to pay a flow cost $e^{\gamma s(t)} c\big(\dot{h}(t)\big)$ on segments, where $\dot{h}$ denotes the time derivative of h and c is a cost functional with the following properties:[7]

- Costless demotion: $c(\dot{h}) = 0$ if $\dot{h} \leq 0$.
- Convexity: $c'(\dot{h})$ is a continuous, strictly increasing function on $(0, \infty)$.
- Inada condition: $\lim_{\dot{h} \to \infty} c'(\dot{h}) = \infty$

An example that satisfies these properties is $c(\dot{h}) = \bar{c} \max\{\dot{h}, 0\}^2/2$, where $\bar{c} > 0$. No costs accrue for non-segments; observe that for any t that is not on a segment, we must have $h(t) = 0$. This cost may be interpreted as a psychic or monetary cost that the worker incurs when learning about the technology in his spare time or during unpaid overtime at work.

Each agent born at $t = 0$ enters the economy with some experience level h_0 for a vintage of age $s_0 \leq 0$, i.e. the first segment may start off with $h_0 \geq h(0) > 0$ if $s(0) = s_0$. An initial density $n_0(\tau, h)$ is given over these endowments, which constitutes the state of the economy at $t = 0$. Newborn workers enter without any endowment, i.e. $h(t) = 0$ for a worker born at $t > 0$.

To summarize, the agent's criterion for a given life l_t starting at t is

$$v(l_t) = \int_t^\infty e^{-(\beta+\delta)(u-t)} \Big[w\big(u, s_t(u), h_t(u)\big) - e^{\gamma s_t(u)} c\big(\dot{h}_t(u)\big) \Big] du,$$

where it is understood that $\dot{h} = 0$ on non-segments. The value function is defined as $V(t, s, h) = \sup_{l_t(t)=(s,h)} v(l_t)$, where the supremum is taken over all feasible lives starting with endowment (s, h). Since discounting is exponential, optimal policies are time-consistent and $V(t, s, h)$ also gives us the forward-looking value for any agent born before t who finds herself in position (s, h) at t.

[7]The cost of human-capital accumulation is growing at the pace of total factor productivity (TFP) to ensure stationarity of the economy. This specification entails that the costs of human-capital accumulation relative to productivity in a technology do not change. This is in line with models where workers have to set aside time from productive work in order to accumulate human capital; in such a setting, the opportunity cost of human-capital accumulation is given by the marginal productivity of devoting one's time to productive work instead of learning. The specification here is a modeling shortcut that avoids the explicit modeling of hours.

2.3 Stationary equilibrium

I will limit the discussion to densities n which have a collection of sets S_n^i in $X \equiv [0,\infty) \times [0,\infty) \times [0,1]$ as their support $S_n = \cup_i S_n^i$. I require the sets S_n^i to contain an open ball; n is assumed continuous and differentiable on each set S_n^i.[8]

For a stationary environment, I require that $n(t,s,h)$ depend only on the age of the vintage $\tau = t - s$, but not on time:

$$n(t,s,h) = n(s+\tau, s, h) = \bar{n}(\tau, h).$$

Stationarity immediately implies that wages and production grow at rate γ, i.e. $w(t,s,h) = e^{\gamma t}\bar{w}(\tau,h)$, $Y(t,s) = e^{\gamma t}\bar{Y}(\tau)$ and $Y(t) = e^{\gamma t}\bar{Y}$. By stationarity of the cost functional, also the value function grows at rate γ: $V(t,s,h) = e^{\gamma t}\bar{V}(\tau,h)$. From now on, we will only work with the stationary distribution; I thus drop the bar-notation and write simply $n(\tau,h)$, $w(\tau,h)$ and so forth.

Definition 2.2. A *stationary competitive equilibrium* is a stationary density $n(\tau,h)$, a measure μ on all possible work lives $l(t) = \{\tau(t), h(t)\}$ and a wage function $w(\tau,h)$ that is continuous on the interior of X such that:

- Compatibility of μ and n: for all Borel sets B in $\mathcal{R}^2$ and for all $u \geq 0$,[9]

$$\int_{t\leq u} e^{-\delta(u-t)} I\big\{\big(\tau_t(u), h_t(u)\big) \in B\big\} d\mu(l) = \int_B n(\tau,h) d\tau dh.$$

- Optimal labor demand: $n(\tau,\cdot) = \arg\max_{\tilde{n}} \big\{Y(\tilde{n}) - \int w(\tau,h)\tilde{n}(h)dh\big\}\ \forall\tau$.

- Optimal labor supply: let A be any set of lives such that $l_t \in A$ implies $v(l_t) < e^{\gamma t} V(\tau_t(t), h_t(t))$. Then A has measure zero under μ.

Note that this definition requires wages to be specified also for regions outside the support S_n of n. In such regions, equilibrium must specify a

[8] The Inada condition 2.1 will naturally lead to such non-degenerate sets S_n^i for the support. Only for the case of a linear production function (i.e. setting $\rho = 1$ in $\tilde{Y}_{CES}$) it will make sense to consider a more general class of sets for S_n, see section 2.6. Note that the specification here allows for densities that drop precipitously down to zero when a vintage dies — which is exactly what occurs in equilibrium. Also, note that feasibility requires that the neighborhoods be connected to points with $h = 0$ or $t = 0$.

[9] $I\{\cdot\}$ denotes the indicator function. The subscript l_t again refers to an agent born at $t \geq .0$. The simple multiplication of the indicator function by the survival function $e^{-\delta(u-t)}$ is valid since death is independent of workers' strategies.

wage schedule that makes workers (firms) choose labor demand (supply) equal to zero.

The economy is understood easiest by first deriving some general properties of equilibrium and by characterizing the worker's problem, which is done in the following subsections 2.4 and 2.5. Subsections 2.6 and 2.7 present sharper characterizations that are obtained when invoking more specific assumptions on the production function. Section 2.8 shows equivalence of competitive equilibrium to a planning solution, which which will give additional insights.

2.4 Properties of equilibrium

We will be looking for a value function $V \in C^1(X)$ that is consistent with a stationary equilibrium. I start to characterize the equilibrium by deriving some properties of the value function. Since workers can always drop down arbitrarily fast in the hierarchy at zero cost and the value function is continuous, we have:

Lemma 2.1. (Value function weakly increasing in h) *The value function $V(\tau, h)$ is weakly increasing in h for all fixed τ.*

Also, workers always have the option to start a new career immediately. So in any position, they must always be at least as well off as workers who start an optimal career.

Definition 2.3. Define the maximal value that can be attained by a career starter as $W = \max_\tau V(\tau, 0)$.

Lemma 2.2. (Value equal for all career starters) *We have $V(\tau, 0) = W$ for all τ and $V(\tau, h) \geq W$ for all (τ, h).*

We will now turn to characterizing the support of the equilibrium density S_n. First, observe that the Inada condition ensures that all rungs in the skill hierarchy must be filled if a vintage is in production:

Lemma 2.3. (All jobs filled in producing vintage) *If the Inada condition 2.1 holds, then $Y(\tau) > 0$ implies $(\tau, h) \in \bar{S_n}$.*[10]

This is a consequence of promotion costs being bounded for any position with $\tau > 0$ but wages going to infinity for empty slots in the skill ladder. Now and in the following, refer to appendix A.1.2 for formal proofs if these are not given in the text.

[10] $\bar{A}$ denotes the closure of a set A.

Another result that allows us to make some headway is that we do not have to consider the entire space of vintages $0 \leq \tau < \infty$, but can restrict ourselves to a finite interval $0 \leq \tau \leq T$:

Lemma 2.4. (Finite support of technologies) *There exists* $T < \infty$ *such that* $\int_0^1 n(\tau, h)dh = 0$ *for all* $\tau > T$.

The proof uses the argument that workers can always secure some positive wage in a frontier vintage without going through training, but that old vintages' productivity goes to zero relative to the frontier. The result is ultimately driven by the the fact that returns to learning are bounded but TFP growth is not.

Definition 2.4. Define the last vintage in production by $T^* \equiv \inf_\tau\{\tau : \int_0^1 n(\tau, h)dh = 0\}$. Note that $T^* < \infty$ is ensured by lemma 2.4.

In order to further characterize S_n, it will be useful to know more about the wage structure in the oldest technology. Consider the problem of a worker who optimizes his career with respect to the switching point $\bar{t}$ when he quits a vintage:

$$\max_{\bar{t}} \int_0^{\bar{t}} e^{-(\beta+\delta-\gamma)t} w[\tau(t), h(t)]dt + e^{-(\beta+\delta-\gamma)\bar{t}} W$$

Since w is continuous, differentiating with respect to τ yields that $\bar{t}$ can only be optimal if $w[\tau(\bar{t}), h(\bar{t})] = (\beta + \delta - \gamma)W$, where the right-hand side is the flow value of starting a new career. If the wage was still higher than that, the worker should stay in the vintage at least a bit more; if it was lower, quitting a bit earlier would make him better off. We summarize:

Lemma 2.5. (Final career wage) *At the end of any career segment* l_0*, wages tend to the flow value of starting a new career, i.e.* $\lim_{t \to l_1} w[\tau(t), h(t)] = (\beta + \delta - \gamma)W$.

Corollary 2.6. (Flat wage structure in oldest technology) *For all* $(T^*, h) \in S_n$*, we have* $w(T^*, h) = (\beta + \delta - \gamma)W$*. If the Inada condition 2.1 holds, this implies that vintages attain maximal productivity upon their death.*

For vintages $\tau > T^*$ that are out of production, the equilibrium definition 2.3 requires us to specify a wage structure that makes it undesirable for both workers and firms to use those vintages. There are many possible choices for w in this region; one of them is $w(\tau, h) = e^{-\gamma(t-T^*)/2} w(T^*, 0)$. Workers will strictly prefer W to any career beyond T^*, and firms would not break even for $\tau > T^*$ — even at optimal factor-input ratios, TFP decays faster with τ than the wage bill does. Also, continuity of w is ensured by this construction.

Reasoning along these lines shows that there cannot be any holes in the support of n along the τ-direction. If there were two vintages from which workers dropped out into new careers, then their wages in these two vintages would have to be equalized at the same value. But this is impossible since the older vintage has lower TFP.

Lemma 2.7. (No holes in vintage space) *Suppose the Inada condition 2.1 holds. Then, if both $Y(\tau_0) > 0$ and $Y(\tau_1) > 0$, also $Y(\tau) > 0$ for all $\tau_0 < \tau < \tau_1$.*

Lemma 2.7 together with lemma 2.3 implies that the closure of S_n must be a rectangle $[T_0, T^*] \times [0, 1]$ if the Inada condition 2.1 holds. Section 2.6 establishes that there cannot be holes in the τ-direction either when labor inputs are perfect substitutes. Arguments in subsections 2.6 and 2.7 will finally show that we must of course have $T_0 = 0$.

2.5 Recursive characterization: partial differential equations

We will now seek to further characterize equilibrium by studying the worker's behavior on career segments. The Hamilton-Jacobi-Bellman equation (HJB) for an interior point of a career segment is the following first-order partial differential equation (PDE):

$$(\beta + \delta - \gamma)V(\tau, h) = w(\tau, h) + V_\tau(\tau, h) + \max_{\dot{h}} \left\{ -c(\dot{h}) + \dot{h} V_h(\tau, h) \right\}, \quad (2)$$

where partial derivatives are denoted by subscripts. The equation says that the flow value of being in state (τ, h) equals the current wage plus the gains (or losses) from vintage aging and the gains from optimal human-capital accumulation.

The optimal career slope $\dot{h}$ depends on the marginal value of skill V_h and the cost of learning. The first-order condition (FOC) for $\dot{h}$ corresponding to the HJB (2) is

$$c'\big(\dot{h}(\tau, h)\big) = V_h(\tau, h), \quad (3)$$

where a unique solution for $\dot{h}$ is assured whenever $V_h > 0$ by the assumptions on c. Since c is convex, the FOC implies that greater value differentials in the hierarchy induce faster human-capital accumulation. Given the boundary condition $V(T^*, h) = W$ for all h, equations (2) and (3) together determine the optimal policies of an agent who takes the wage function w as given.

Sometimes, it will be convenient to work with the Euler equation, which tells us how the marginal value of human capital V_h changes along an optimal career path. Differentiate (2) with respect to h and use the envelope condition (3) to obtain

$$\frac{dV_h}{dt} = \dot{h}(V_h)_h + (V_h)_\tau = (\beta + \delta - \gamma)V_h - w_h, \tag{4}$$

where an agent's career is parameterized by time by t: we have $d\tau = dt$ and $dh = \dot{h}dt$. The dependence of the various functions on (τ, h) is suppressed for the sake of clarity. We can solve (4) as an ordinary differential equation in t along an agent's optimal career path and see that the marginal value of human capital equals the discounted integral of marginal wage gains over a career:

$$V_h(t) = \int_t^T e^{-(\beta+\delta-\gamma)(u-t)} w_h\big(\tau(u), h(u)\big)du, \tag{5}$$

where T is the end of the career segment and $V_h(T, h) = 0$ since $V(T, h) = W$ for all h, i.e. the marginal value of skill is zero at the end of a career. This suggests that the incentives for human-capital accumulation are strongest in the beginning of a career, and makes us expect that human-capital accumulation is decreasing over segments.

I now proceed to characterize how the density $n(\tau, h)$ evolves given the optimal local behavior of workers. Inside S_n, n must obey the following PDE:

$$n_\tau(\tau, h) + \dot{h}(\tau, h)n_h(\tau, h) = -\left[\delta + \dot{h}_h(\tau, h)\right] n(\tau, h), \tag{6}$$

where the notation $\dot{h}_h = \frac{\partial \dot{h}}{\partial h}$ is used. This PDE says the following: when following a worker's optimal career path, the density thins out at the death rate, δ, plus the divergence of the promotion paths, $\dot{h}_h$. Appendix A.4.1 provides a derivation of this equation.[11] For a given boundary condition $n(\tau, 0) = n_0(\tau)$ on $\tau \in [T_0, T^*]$ and given optimal policies $\dot{h}$, we may solve this PDE throughout S_n.

To summarize, the HJB (2) with its boundary conditions characterizes workers' optimal strategies given wages. Equation (6) tells us how the resulting decisions by workers in (3) translate into a density n (once we know

[11] The equation is the usual mass-transport equation for densities in a deterministic context; it may be seen as a special (non-stochastic) case of the Kolmogorov forward equation.

how new-borns enter vintages). Optimality of firms' decisions implies that wages w on the support of n are given by the Frechet derivative of $Y(\tau)$ with respect to $n(\tau, \cdot)$. Wages w then feed back again into workers' HJB.

Formally, we can say that any equilibrium must be associated with two functions n and V that are defined on a rectangle $[0, T^*] \times [0, 1]$. These functions (n, V) together with T^* must satisfy the system of PDEs given by (2) and (6). The boundary conditions for V are $V(T^*, h) = W$ for all $h \in [0, 1]$ and $V(\tau, 0) = W$ for all vintages $\tau \in [0, T^*]$ with positive entry, i.e. those τ for which $n(\tau, 0) > 0$. For n, we have the boundary condition $w(T^*, h) = (\beta + \delta - \gamma)W$ for all $h \in [0, 1]$.[12]

Note that there is no explicit boundary condition on $n(\tau, 0)$ at career entry (i.e. $h = 0$). However, this does not mean that n is irrelevant for the career-entry decision. Indeed, n must induce a wage structure that makes workers indifferent between entry into the different active vintages, which shows up as a boundary condition on $V(\tau, 0)$.

Note that this system of PDEs is non-standard in the following respects: first, wages are determined non-locally; they depend not only on the density in the immediate (τ, h)-neighborhood of the worker but also on h-levels in the same vintage that are far away from the worker. Second, the boundary location T^* and the boundary value W are unknowns. These complications make computation of equilibrium a challenging task since standard methods cannot be used.

2.6 Sub-case: perfect substitutes

The next two subsections will be concerned with the characterization of the wage structure and human-capital accumulation. Specifically, the following objects are of interest: the *skill premium* inside a given vintage (i.e. $w(\tau, \bar{h})/w(\tau, \underline{h})$ for $\bar{h} > \underline{h}$), the *(vintage-)tenure premium* (defined exactly as the skill premium, but conditioning on vintage tenure instead of h as the independent variable) and the intensity of *skill accumulation* $\dot{h}(\tau, h)$. I will also study how these objects depend on technological growth by comparing steady states for different values of γ, ceteris paribus.

We first turn our attention to the special case where different skill levels are perfect substitutes: take the CES-aggregator (1) with $\rho = 1$ as the production function. Then, wages are independent of the distribution of workers across the skill hierarchy: $w(\tau, h) = e^{-\gamma\tau} f(h)$. Throughout this

[12] The boundary conditions are implied by lemmas 2.2 and 2.5. W is given in definition 2.3.

section, we will assume a standard learning curve and require that $f'(h) > 0$ and $f''(h) < 0$.

With a linear production function, the problem essentially reduces to one of partial equilibrium: it is sufficient to solve a worker's problem, let every worker follow her optimal policy and collect the results in an equilibrium density n.[13] It is easy to show that the worker will then always switch to the newest vintage when she relocates, see lemma A.2 in the appendix for the formal statement. Also, by stationarity of the economy, it is optimal for the worker to complete identical cycles of human-capital accumulation over and over again.

We will now be concerned with the question how technological progress (in the form of a change in γ) affects agents' decisions. Using the solution to the Euler equation (5), we obtain

$$V_h(t) = \int_t^{T^*} e^{-(\beta+\delta)(u-t)} f'\big(h(u)\big)du, \tag{7}$$

where T^* is the optimal switching point to a new career. Since the wage gains from human-capital accumulation are decreasing in h by the concavity assumption on f, this entails that workers in a lower hierarchy position have stronger incentives to learn ceteris paribus. Another consequence is that a worker with a longer horizon accumulates human capital faster. Furthermore, equation (7) shows that technological growth is inessential for the inter-temporal incentives of human-capital accumulation while *inside* a vintage – it only affects the optimal switching point T^*

Now, define the *(vintage-) tenure premium* as the ratio of the wage of an tenure-t worker in a vintage to the wage of a career starter. In a stationary context, this equals $p_\gamma(t) = e^{-\gamma t} w(h(t))/w(h(0))$, where the optimal path h of course depends on γ through $T^*(\gamma)$. Using lemma A.3 (see appendix), one can then show the following:

Proposition 2.8. (Shorter horizon lowers tenure premia) *Assume* $\tilde{Y} = \tilde{Y}_{CES}$ *with* $\rho = 1$. *Suppose that* $\tilde{\gamma} \neq \gamma$, *but fix all other parameters. Then* $\tilde{T}^* > T^*$ *implies* $p_{\tilde{\gamma}}(t) > p_\gamma(t)$ *for all* $0 < t \leq T^*$.

This result says that with a linear production function, a shortening in the life span of technologies must always be accompanied by a decrease in the

[13] Note that typically all agents will follow the same path and thus n would be a *measure* that cannot be represented by a density function. However, this is unproblematic since the production function (1) and the PDE (6) still make sense, the former as a linear functional on measures and the latter in the weak sense.

tenure premium. Figure 2 illustrates the intuition for the result. Workers with a shorter planning horizon in their technology have fewer incentives to invest in technology-specific knowledge and thus learn at a slower pace, leading to a lower experience premium.

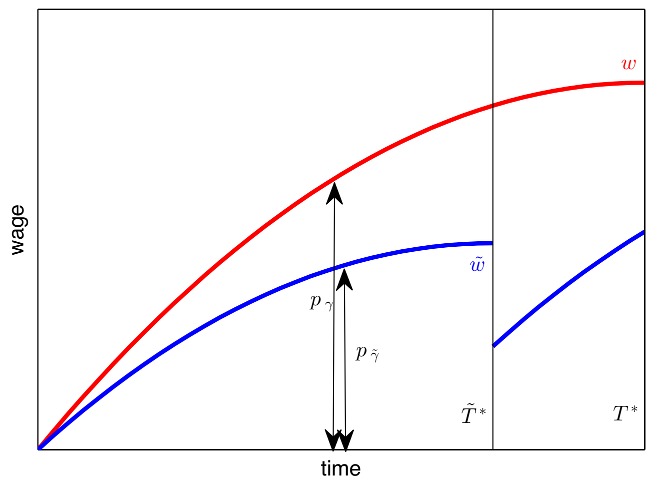

Figure 2: Horizon effect under substitutability

To show that faster technological growth leads to faster scrapping of technologies, define the discounted value of a career segment of length T by

$$K(T) = \int_0^T e^{-\rho t}\big[w\big(h_T(t)\big) - c\big(\dot{h}_T(t)\big)\big]\,dt, \tag{8}$$

where $h_T(\cdot)$ is the optimal skill-accumulation path for a career of length T.

Proposition 2.9. (Faster growth shortens careers and lowers tenure premia) *Assume* $\tilde{Y} = \tilde{Y}_{CES}$ *with* $\rho = 1$. *Then, if the function* $K(\cdot)$ *in (8) is twice differentiable, the following hold:* $T^*(\gamma)$ *is non-increasing in* γ *and strictly decreasing whenever* $T^*(\gamma) > 0$*; the tenure premium* $p_\gamma(t)$ *is decreasing in* γ *for any* $0 < t \leq T^*$.

One may find this result somewhat counter-intuitive: it says that technological growth leads to less learning-intensive careers. This is in the sense that for each given tenure, agents have accumulated less knowledge in the high-growth than in the low-growth world.[14] The following sections will

[14]Of course, agents in a high-growth world might still learn more in total since they switch to new vintages more often.

show that this result need not hold when different human-capital levels are complementary and how it can indeed be overturned.

2.7 Wage structure under complementarity

In this section I will further characterize the wage structure for the case where labor inputs of different skill levels are complementary and an Inada condition holds, as is the case for $\rho < 1$ in the CES production function. I start with the following observation:

Lemma 2.10. (Highest entry wage in vintage T^*) *If the Inada condition 2.1 holds, then* $w(T^*, 0) \geq w(\tau, 0)$ *for all* $0 \leq \tau < T^*$.

Intuitively, entry wages have to be lower in young technologies for the following reason: entering a new technology provides experience that will be valuable in the future. So, barring any offsetting wage differential, all workers would choose to enter new technologies. However, under the Inada condition some workers are also needed in low-skill tasks in the oldest technologies. In order for both entry options to be equally attractive, entry wages in young technologies have to be lower than in old technologies.

It turns out that on the top of the skill hierarchy, the converse is true:

Lemma 2.11. (Wage explosion for skilled in young technologies) *If the Inada condition 2.1 holds, then* $\lim_{\tau \to 0} w(\tau, 1) = \infty$ *and* $Y(\tau) > 0$ *for all* $\tau \in (0, T^*)$.

The intuition behind the result is that people with high skills in young technologies must have worked hard to acquire these skills. Thus, those workers have to be compensated by high wages. Note that this is ensured if only few people take such steep paths in young vintages. When scarce enough a factor, the skilled in young technologies can then earn unbounded returns.

Collecting the previous results yields:

Corollary 2.12. (Wage compression) *Suppose the Inada condition 2.1 holds. Then the wage difference between high-human-capital and low-human-capital workers is highest in the youngest vintages and lowest in the oldest vintages, i.e.* $w(\tau, 1) - w(\tau, 0) \to \infty$ *as* $\tau \to 0$ *and* $w(\tau, 1) - w(\tau, 0) \to 0$ *as* $\tau \to T^*$.

Intuitively, the wage structure is compressed over the life cycle of a vintage because skill becomes less scarce. It is easier to acquire skills over a long time than to master a technology that has barely been invented. With a view to tenure premia, note that wage compression opens the possibility that a technological acceleration leads to an increase in tenure premia. Since the wage structure is steeper in young technologies, a shortening of the vintage horizon T^* can send more workers into steep earnings paths and increase the average tenure premium.

Another consequence of the results above is the following:

Corollary 2.13. (Obsolescence/wage losses) *There is a positive measure of careers with $dw(\tau(t), h(t))/dt < 0$ for some t. Furthermore, agents who quit their vintage start their new career with a wage weakly lower than their last wage in the old career.*

Note that the first type of wage losses (those occurring during a career) cannot occur when skills are perfect substitutes. These wage losses are remarkable since they occur without depreciation of human capital — an assumption often invoked in Ben-Porath-type models in order to obtain downward-bending wage profiles for old workers. Here, agents do not lose any of their skill over their vintage career; the reason for the wage losses is that the relative price for skill falls over time, a phenomenon commonly referred to as *obsolescence*.

Finally, note that the second type of wage loss, which stems from the loss of vintage human capital due to a vintage change, is not due to an exogenous shock (an assumption sometimes made in human-capital models), but stems from an endogenous decision. The worker accepts a temporary wage loss in order to obtain skills in a new technology which pay off later in his work life.

2.8 The planner's problem

To conclude the section of theoretical results, I show that the competitive equilibrium characterized in the previous discussion is equivalent to the solution of the following planner's problem (a detailed formal discussion is provided in appendix A.4). Let the planner weigh the utility of an agent born at t with $e^{-\beta t}$. The planner's problem is to choose a density $n(t, s, h)$ to maximize

$$\int_0^\infty e^{-\beta t} \big(Y\big(n(t; \cdot)\big) - C(t)\big)\, dt,$$

where $C(t)$ denotes the aggregate cost of human-capital accumulation at t.

In the first step, I restrict the density n to have support up to a given vintage age T. In order to evaluate the planner's criterion, we first need to know the optimal human-capital-accumulation strategy for the planner to implement a given density n. It can be shown that the optimal accumulation strategy is such that agents' career paths inside a vintage never cross. This characterization is sufficient to derive an expression for total human-capital-accumulation costs in terms of n. Given this expression, one can then derive the planner's first-order conditions from a standard Lagrangian.

In the second step, I vary the maximal vintage age T to find the global optimum. Denote by T^* the vintage age that is associated with the global maximum of the planner's criterion. For this T^*, I show that the PDEs arising from the planner's first-order conditions are equivalent to system of PDEs implied by competitive equilibrium (as described in subsection 2.5).

The following proposition establishes that the global solution to the planner's problem is a competitive equilibrium and provides a partial converse of this statement:

Proposition 2.14. (Equivalence of planner's solution and competitive equilibrium) *The stationary (global) solution to the planner's problem with T^* is a competitive equilibrium (CE). Any stationary CE is also a solution to a planner's problem for some $T \leq T^*$. There is no CE with $T > T^*$.*

The planner's problem brings some insights that are not easily obtained in competitive equilibrium. For example, a concavity argument ensures uniqueness of the planner's solution and a decomposition of the planner's criterion gives an empirically observable upper bound on the total cost of human-capital accumulation in the economy, see appendix A.4 for details.

3 Calibration and quantitative results

In this section, I solve the model numerically and calibrate it to German matched employer-employee data. The goals of the calibration are the following.

First, by taking the model to the data we can see if some sign restrictions imposed by the model are borne out in the data. A key prediction of the model is that young technologies have higher premia on skill. This prediction is confirmed in the data, both when identifying young vintages by establishment age (subsection 3.3) and when identifying them by high growth (subsection 3.6).

Second, it is not clear a priori if the model can fit both the wage structure and the distribution of workers across vintages and skill levels. It turns out that the model achieves a good qualitative and a reasonable quantitative fit, but that it misses some specific quantitative features in the data.

Third, the calibration will tell us about the importance of different drivers of wage growth. The calibration indicates that earnings growth is largely accounted for by human-capital accumulation, as opposed to vintage productivity gains and the relative scarcity of skill. This result is due to the fact that the production function is estimated to be close to linear.

Finally, the calibration yields results on comparative statics that go beyond the analytical results derived so far. Most importantly, the calibration allows to sign the effect of an increase in γ on human-capital accumulation for the case $\rho < 1$, which was ambiguous in the theoretical analysis. A counterfactual exercise indicates that a technological acceleration leads to an intensification in skill accumulation and a rise in tenure premia, meaning that general-equilibrium effects more than outweigh the horizon effect.

3.1 Data

The model is calibrated to yearly data. I use the German IAB employment sample (1975-2001), a large employer-employee-matched panel from administrative sources, see appendix C for details. The panel has information on workers' earnings, hours and other characteristics. Each worker is matched to an establishment, for which we observe its founding date, the number of workers and industry classification, among others. Since the data set starts in 1977, the age of the establishment is not available for establishments founded before 1978. In view of this limitation, I group establishments into three age categories: young (founded in 1989 or later), medium-age (founded between 1978 and 1988) and old (founded before 1978).

3.2 Calibration

I consider a cross section of workers in 2001, assuming that the economy is in steady state. I identify vintage age in the model with the age of an establishment in the data. The underlying assumption is that new establishments incorporate the newest available technology.[15] For the production function, I use the CES aggregator defined in (1). To allow for a learning curve with diminishing returns, I choose the specification $f(h) = 1 + a_1 h + a_2 h^2$ to model returns to skill, where $f(0) = 1$ is chosen as a normalization. I impose $f'(h) \geq 0$ and $f''(h) \leq 0$ for all $h \in [0, 1]$, i.e. returns to skill are assumed to be positive and diminishing. For the cost of human-capital accumulation, I

[15]See subsection 3.6 for a discussion of this assumption and an alternative identification strategy.

choose a quadratic specification that is convenient because of its parsimony: $c(\dot{h}) = \bar{c}\max\{\dot{h}, 0\}^2/2$.

Finding a competitive equilibrium amounts to solving the system of PDEs and integral equations described in section 2.5. The wage equation is given by

$$w(\tau, h) = e^{-\gamma\tau} f(h) \left(\frac{\tilde{Y}(\tau)}{n(\tau, h)} \right)^{1-\rho}. \tag{9}$$

I propose a solution algorithm which uses a discretization scheme as in standard lattice methods and attacks the problem of endogeneity of the boundary values with an algorithm inspired by the way a real economy might oscillate around a steady state under some inertia. Appendix B provides the complete documentation of the algorithm.

Table 1 shows the calibrated parameters. A subset of parameters is chosen outside the model. The death rate δ is set to 0.092; this number is chosen to match the rate at which workers are displaced from establishments in the data (6.7% yearly) plus the rate at which they leave the labor force (2.5%, obtained from an expected labor-market participation of 40 years for a 20-year old[16]).[17] β is set to 0.015 to obtain a standard yearly discount rate of $\beta + 0.025 = 0.04$, where 0.025 is the exit rate from the labor force. Based on the results by Cummins & Violante (2002), I choose $\gamma = 0.005$ for vintage productivity growth. These authors report that in the second half of the 20th century, one third of GDP growth was due to increases in TFP and in the quality of capital. The rest came from increases in the quantity of labor, the quantity of capital and growth of human capital. I adopt the view that the increase in vintage productivity in the model equals the increase in TFP plus increases in the quality of capital. GDP growth in Germany was roughly 1.5% yearly in 1991-2011, from which I obtain $\gamma = 0.015/3 = 0.005$.[18]

The remaining four parameters $\rho = 0.89$, $\bar{c} = 2.2$, $a_1 = 1.5$ and $a_2 = 0.0$ were chosen to minimize the mean squared deviation of the ten model moments given in table 2 to their counterparts in the data. The rationale for the choice of the calibration moments is as follows.

---

[16] I take the expected retirement age of 60 years for Germans from the German association of retirement insurers, see www.deutsche-rentenversicherung.de.

[17] It is easy to see that a model where agents are displaced from a vintage with probability δ_p and die at rate δ_d yields the same allocations as a model with death rate $\delta = \delta_p + \delta_d$ — human-capital-accumulation decisions do not have any effect on a worker's life after a vintage-displacement shock.

[18] The German GDP data are from the German *Bundesbank*.

Table 1: Calibration parameters

Group	Parameter	Explanation	Value
chosen outside model	β	discount rate	0.015
	δ	exit rate from establishments	0.092
	γ	vintage productivity growth	0.005
chosen to match moments	ρ	CES production function	0.89
	$\bar{c}$	cost of learning	2.2
	a_1	slope of learning curve	1.5
	a_2	curvature of learning curve	0.0

The learning curve is specified as $f(h) = a_0 + a_1 h + a_2 h^2$, where $a_0 = 1$ is chosen as a normalization. The restrictions $f'(h) \geq 0$ and $f''(h) \leq 0$ for all $h \in [0, 1]$ are imposed, i.e. the permissible set is $\{(a_1, a_2) \in \mathbb{R}^2 : a_1 \geq 0, a_2 \leq 0 \text{ and } a_1 + 2a_2 \geq 0\}$.

In principle, the model has predictions on human-capital accumulation (h, $\dot{h}$), the distribution of workers across vintages (n), the wage structure over vintages and skill (w) and the output across vintages (Y). We cannot observe vintage human capital h in the data, which restricts the choice of calibration targets. Since labor is the only production input in the model economy, I treat $Y(\tau)$ in the model as a prediction on the wage bill of age-τ establishments, i.e. I exclude payments to the factor capital and entrepreneurial profits. Since hours worked are not available in the data, I only consider full-time workers and regard their earnings as equivalent to w in the model.

As a measure of central tendency for the wage bill, I include median log earnings by establishment category as a target. Since the model is homogeneous, the model results can be scaled by any constant. Thus, one normalization has to be chosen. I equate median earnings in old establishments in the model to those in the data and take the difference of median earnings in young and medium-age establishments to those in old establishments as the first two calibration targets.

I now turn to calibration targets 3 to 8, which capture the premium on skill in the different vintages. The skill premium is linked to the learning curve $f(h)$, as can be seen from the wage function (9). Note that the estimation of the learning curve $f(h)$ would be straightforward if we could observe workers' human capital h. Knowing h, one could obtain an estimate for the density n and then directly back out the function $f(h)$ from wage data using log-differences of equation (9). Given that h is not observ-

Table 2: Calibration targets

Moment	Establishments founded…	Data	Model
Difference of median log earnings to establishments founded before 1978	1978-1988	-0.13	-0.16
	1989-2000	-0.13	-0.45
5-year tenure premium	before 1978	0.28	0.20
	1978-1988	0.34	0.27
	1989-2000	0.41	0.36
10-year tenure premium	before 1978	0.35	0.40
	1978-1988	0.41	0.50
	1989-2000	0.47	0.64
Fraction of workers entering establishments	1978-1988	0.13	0.33
	1989-2000	0.55	0.56

The tenure premia are calculated from the predicted values of a censored regression (top-coded data) of log earnings on a quartic in establishment tenure within the respective group of establishments.

able, however, the best one can do is to keep track of workers' tenure in an establishment. Workers' human capital is then obtained from the model predictions on human-capital accumulation. In order to be able to identify both the slope and the curvature of the learning curve $f(h)$, I include both the 5- and the 10-year premium on establishment tenure by establishment category.

I use the raw establishment tenure premium, which I compute from the predicted values of regressions of log wages on a quartic in establishment tenure within each age category of establishments. I do not include labor-market experience, occupation or industry tenure in these regressions. These are highly correlated with establishment tenure and lead to problems in identifying the contribution of establishment tenure, a common problem in the empirical literature. I take the stand here that *all* human capital is specific to technologies, just as is the case in the model. The idea is that a worker's skills can become totally obsolete when there is a disruptive change to technology, even when the worker stays within her occupation or industry. An example would be a worker in a photo laboratory that changes from the physical development of films to digital technology, making the worker lose any competitive advantage over outside workers. The fact that other forms of human capital are not considered in the calibration is of course a limitation to my approach. Since the model in its current form does not allow for other forms of human capital, however, this is probably the best that can be done. I leave extensions in this direction to further research.

The last two calibration moments are related to the distribution n of workers across vintages. Since the model has all workers stay in their vintage until the vintage dies (or until they are exogenously displaced, as is assumed for the purpose of the calibration), the entry density into vintages together with the maximal vintage age T^* completely determine the number of workers with a given tenure in a vintage. I thus use only data on entry into establishments by age category to capture information about the distribution n.

3.3 Model fit and equilibrium properties

A grid search is used to minimize the mean squared distance of the model moments to the targets. Table 2 shows that that the model is qualitatively in line with the main features of the earnings distribution in the data: on average wages are higher in older vintages, but the tenure premium is lower in old vintages. These two facts also imply that entry wages are highest in old vintages. The model can also capture the fact that the bulk of entry occurs for establishments of young age. Vintages survive for $T^* = 65$ years for the calibrated parameters.[19]

The overall quantitative fit is reasonable, but there are some shortcomings. The model exaggerates the difference of median wages between young and old establishments. Furthermore, the model generates 5-year tenure premia that are slightly too low and 10-year tenure premia that are somewhat too high. Also, the model assigns too few entrants to young vintages. Finally, the model has one stark prediction that is likely not to be met in the data (although it cannot be rejected formally since establishment is right-censored in the data): the skill premium is zero in the dying vintage.

Figure 3 shows characteristics of the selected model. Since there is substantial entry into old vintages, the data reject a production function with

[19]Note that most workers do not stay in their vintage until its demise since they die or are displaced from the vintage. Since this is a perpetual-youth model, workers' age does not affect vintage choice upon entry. If workers had different time horizons, as is the case in reality, we would expect sorting by age, however: old workers should prefer older technologies since their horizon is too short to cash in on skill investments in young technologies. In the data, indeed I find that such sorting takes place: only 14% (48%) of workers in old establishments are below 30 years (40 years), whereas this number is 19% (57%) in medium-age and 21% (58%) in young establishments. The correlation between establishment age and worker age is 0.104, which is highly significant.

perfect substitutability of skills.[20] ρ is estimated to be close to unity, however, thus complementarities between skills are weak. The career lines in the upper-left panel show agents' equilibrium trajectories through the (τ, h)-space. The line starting in the lower left corner refers to workers who enter the youngest vintage, the line starting at $\tau = 60$ and $h = 0$ refers to a worker entering the vintage of age 60. We can see that there is entry into all vintages. Agents in young vintages make the hardest efforts to climb the skill hierarchy. This is in line with value differentials in the hierarchy being highest in these vintages, as the value function in the lower-left panel shows.[21]

The lower-right panel in figure 3 illustrates that wage compression is a process that happens all the way from the newest to the oldest vintages. The skill premium is highest in the youngest vintages and continuously shrinks as the vintage ages. The density function in the upper-right panel is in line with wages: workers with high experience in young technologies are relatively scarce, whereas skill becomes more abundant in old vintages as workers press up in the skill hierarchy from below.

The estimate $a_1 = 1.5$ implies that the learning curve has substantial slope. Together with the estimate $a_2 = 0.0$, it implies that the highest-skill workers are 2.5 times more productive than the lowest-skill workers if the number of workers was equalized across skill levels. As for the curvature of the learning curve, the best fit to the data is obtained at the zero lower bound for a_2, implying that the learning curve has no curvature. At first glance, this may seem at odds with the diminishing returns to establishment tenure that we see in the data: the 5-year tenure premium is more than half as high as the 10-year tenure premium for all establishment categories. However, note that the learning curve is only one of three forces that determine the returns to tenure in the model. As seen before, workers also slow down the rate of human-capital accumulation as they accumulate tenure in a vintage, and

[20] Recall that all workers enter the newest vintage if $\rho = 0$, see lemma A.2.1.

[21] A note is in order on why the trajectories in the upper-left panel of figure 3 do not have infinite slope although $\rho < 1$. Note that $\rho < 1$ implies that the Inada condition holds and that thus all positions on the skill ladder should be occupied in all vintages, as shown in lemma 2.3. This does not occur in the numerical implementation because the algorithm builds up the number of grid points one-by-one for the youngest vintages. This is done for reasons of computational stability. I have experimented with an algorithm that immediately starts with the full skill ladder for the youngest vintage, but this algorithm proved unstable. As the number of grid points is increased, the errors from leaving out grid points in the upper-left corner go to zero. This is due to the facts that (i) the density of workers goes to zero for high skills in the youngest vintages and (ii) no input is essential for production in the parameter range $\rho \in [0, 1]$. For a detailed discussion see appendix B.

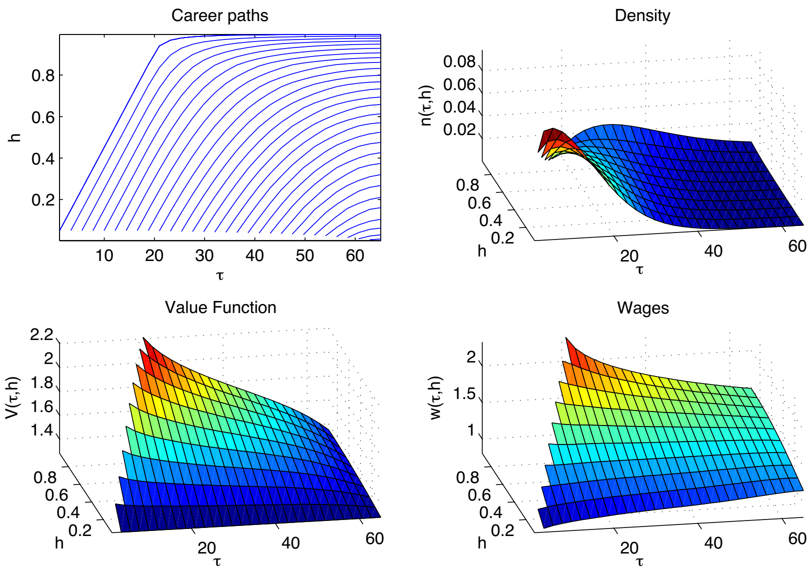

Figure 3: Equilibrium (I)

the relative scarcity of skill matters. We will later see that human-capital accumulation is the major force behind wage growth, and its deceleration over careers is sufficient to generate the diminishing returns to tenure we see in the data.

Figure 4 shows more variables of interest. In the upper-right panel, we see that entry into vintages is hump-shaped and strictly greater than zero even for the oldest vintages.[22] As apparent in figure 3, late entrants are compensated for learning the least useful skills by the highest entry wages. The upper-left panel illustrates that despite positive entry, total employment is decreasing in vintage age after some point because incumbent workers are displaced at a faster rate than new workers enter. A similar pattern is evident for output, see the lower-left panel.

The lower-right panel in figure 4 shows (average) labor productivity by vintage age. The pattern is reminiscent of the hump-shaped, back-loaded return profiles that are typical for organization-capital models (see Atkeson & Kehoe, 2005, for example). Young vintages are unproductive because they have an unbalanced mix of labor inputs; marginal returns to the different skill levels are far from equalized since high-skill labor is scarce. In older

[22]The entry density is given by $m(\tau) = n(\tau, 0)/\dot{h}(\tau, 0)$.

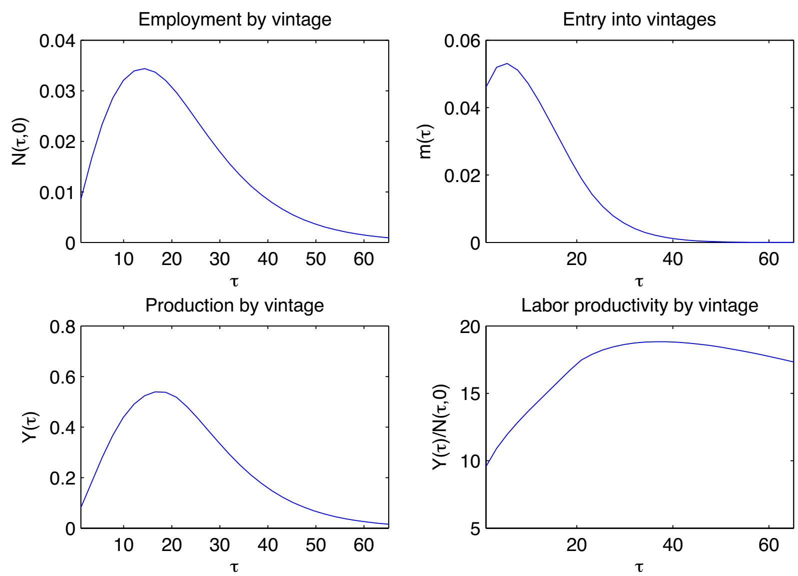

Figure 4: Equilibrium (II)

vintages, human-capital accumulation leads to gradual equalization of skill returns. However, the returns from skill-return equalization wear off over time and the negative TFP effect eventually dominates.[23]

How does the model perform when it comes to moments that were not targeted in the calibration? Table 3 compares measures of the unconditional wage distribution and the worker distribution from the model to the data. The model does quite a good job matching the middle of the wage distribution, but it does not generate enough dispersion in the tails. This is maybe not surprising since the model has no other source of heterogeneity than vintage human capital. As for the distribution of workers, we see that the model only slightly understates the number of workers in young establishments. However, it is hard for the model to generate the substantial numbers of workers in old establishments, as was already the case for entry. The most likely culprit here is that the model misses a selection mechanism as in firm-dynamics models: successful establishments grow faster, are likely to survive longer and thus retain more workers. Including such a mechanism

[23]When plotting a vintage's productivity over time, however, it is steadily increasing, flattening out towards the end. Recall that a vintage reaches maximal productivity only upon its death.

into the model thus looks like a promising avenue for further research.

Table 3: Non-targeted moments

Moment	Data	Model
Quantile differences in unconditional log-wage distribution		
10-50	-0.50	-0.37
25-50	-0.21	-0.24
75-50	0.24	0.19
86.25-50	0.48	0.27
Worker distribution: fraction of workers in establishments founded…		
before 1978	0.56	0.39
1978-1988	0.14	0.36
1989-2000	0.31	0.25

13.75% of wages are top-coded, I thus choose 86.25% instead of 90% as the highest quantile in the wage distribution.

3.4 Determinants of wages

Figure 1 shows wages following a cohort of vintage entrants over time as they accumulate human capital in their vintage. We see that workers entering different vintages have vastly different tenure-earnings profiles. The curve that extends farthest to the right refers to workers who enter the frontier technology; the shorter the curves become, the later the respective workers enter the vintage.

An interesting feature of the wage profiles generated by the model is that they have heterogeneous slopes and curvature. In Ben-Porath-type models, heterogeneity in shape is usually generated by assuming heterogeneous learning ability, see for example Guvenen & Kuruscu (2007) and Huggett, Ventura & Yaron (2006). In contrast to these models, agents here are ex-ante equal and all heterogeneity is ex-post. In fact, heterogeneity in earnings profiles is essential in order to give workers the incentives to enter *all* existing vintages and to ensure that all vintages have an efficient skill mix.

Another topic from the labor literature addressed by the model is "overtaking". Hause (1981) defines overtaking as the fact that two wage profiles with different slopes and the same present value have to intersect at a certain point. The model has precise predictions on when this overtaking point occurs for different pairs of agents in the economy. In the calibrated model, overtaking happens throughout the first decade of workers' careers, as is evident from figure 1.

To understand which forces are at work in workers' tenure-earnings profiles, it is useful to decompose wage growth into its different components. Divide both $n(\tau, h)$ and $\tilde{Y}(\tau)$ in equation (9) by the total number of workers $N(\tau, 0)$ in vintage τ. Then take logarithms and consider infinitesimal changes in the log-wage along a career $\{h(t), \tau(t)\}$, recalling that wages increase at rate γ over time in stationary equilibrium:

$$\left.\frac{d\ln w}{dt}\right|_{\tau(t),h(t)} = \dot{h}(\tau, h)\frac{f'(h)}{f(h)} + (1-\rho)\left(\frac{\partial \ln \frac{\tilde{Y}(\tau)}{N(\tau,0)}}{\partial \tau} - \frac{d\ln \frac{n[\tau(t),h(t)]}{N[\tau(t),0]}}{dt}\right). \quad (10)$$

The three terms on the right-hand side have a clear economic interpretation: I term them (from left to right) the *skill effect*, the *organization-capital effect* and the *relative-scarcity effect*.

The *skill effect* captures returns from human-capital accumulation. This is the only effect present when skills are perfect substitutes ($\rho = 1$). In figure 3 we see that human-capital accumulation $\dot{h}$ is always positive but decreasing over careers. Since the learning curve f was assumed to be increasing and concave, this means that the skill effect is always positive but wears off over the course of a career.

The innovation in the model presented here with respect to the models in the literature lies in the terms that are switched on when lowering ρ below unity. These are stemming from relative factor supply. The second term is proportional to the growth of labor productivity $\tilde{Y}(\tau)/N(\tau, 0)$ in the vintage. This term is positive for all vintages in the calibration.[24] It represents the gains from joint learning and the equalization of factor returns, which are positive throughout the vintage's lifetime. I call this term the *organization-capital effect*.

Finally, the *relative-scarcity effect* is key for understanding why earnings profiles are decreasing for most workers towards the end of their careers. This third term is related to the relative abundance of workers of a given skill, $\frac{n[\tau(t),h(t)]}{N[\tau(t),0]}$. As a vintage ages and more agents enter it, once-scarce skills become more common as workers from the lower ranks press up in the skill hierarchy. A real-world example for this might be an HTML-programmer whose skills commanded high returns when the internet was in its infancy but saw his wages dwindle as more and more other programmers learned HTML and his knowledge became less scarce. Proposition 2.13 shows that

[24] Recall that $\tilde{Y}(\tau) = e^{\gamma\tau}Y(\tau)$.

if $\rho < 1$ this effect must override the other two for some workers and lead to earnings losses.

Table 4 decomposes wage growth in the calibrated economy according to equation (10), averaging over all workers in the vintage-age categories from before. We see that the skill effect is quantitatively most important. The other two effects are less pronounced because ρ is estimated to be close to unity. Especially workers in young technologies accumulate skill fast and experience strong wage growth through the skill effect. The organization-capital effect is weaker than the skill effect and wears off as the vintage ages. Finally, we see that the relative-scarcity effect indeed becomes negative for workers in old technologies as their skills become obsolete.

Table 4: Wage-growth decomposition

Establishments founded...	Skill	Organization capital	Relative scarcity
before 1978	0.021	0.001	−0.005
1978-1988	0.045	0.003	0.000
1989-2000	0.056	0.004	0.010

Effects in wage-growth decomposition from equation (10) by age category of establishments.

3.5 Technological acceleration

I now return to the question on how a technological acceleration impacts the economy. To do this, I increase γ from 0.005 to 0.006 and keep the other parameters at their values from the baseline calibration.

Higher frontier productivity growth gives incentive to abandon old vintages earlier; the maximal vintage age T^* drops from 66 to 63 years. More workers enter young technologies: entry into young vintages is up to 62% from 56%.

As for human-capital accumulation, the theoretical analysis indicated that an increase in γ had two competing effects: a shorter vintage horizon discourages skill accumulation; however, faster productivity growth lures more workers into learning-intensive young vintages. Indeed, I find that different effects dominate for different kinds of workers. In medium-age and old vintages there is less human-capital accumulation; the skill effect in table 4 goes slightly down to 0.044 and 0.019, respectively. In the young vintages, however, the average skill effect rises to 0.059, since more workers enter the frontier technologies. On the whole, the rise in learning in the young vintages dominates and there is more human-capital accumulation in the economy. Tenure premia follow the same pattern: the 5-year tenure premium increases to 0.37 for young vintages, stays the same for medium-age vintages and decreases to 0.19 for old vintages. The results for the 10-year tenure premium are similar.

There is almost no change in the organization-capital and relative-scarcity effects. The gap in median log wages from young vintages versus old ones decreases to 0.43, but remains almost the same for medium-age vintages. This is because young technologies have now a larger TFP advantage over old ones. For the same reason, the right side of the hump in the cross-vintage productivity profile in figure 4 becomes steeper.

3.6 More evidence

Note that the identification strategy pursued so far has relied on the assumption that new establishments, and only new establishments, operate the frontier technology. For large firms, this assumption is reasonable enough: new plants should usually incorporate the latest available technology. Old plants are more likely to be stuck with old technologies due to switching costs, but some of them will also upgrade to the newest vintage. For small, one-establishment firms, the case is less clear-cut. There are surely many new, small firms that are not on the technological frontier. Also, new firms have been shown to have high exit rates, which is at odds with young vintages having the longest life expectancy in the model. We would still expect small entrants to operate newer technologies than incumbents on average, but identifying vintages by establishment age may be probematic in quantitative terms.

In view of these shortcomings, I will now turn to the implications the theory has on vintage growth. Suppose that a certain industry has plants that are predominantly of new vintages. Then the model makes us expect higher tenure premia and lower average wages in this industry than in an industry mainly comprised of old-vintage plants. The same should be true for occupations or even establishments: occupations/establishments which make heavier use of new vintages should display higher tenure premia and lower mean wages. Also, note from figure 4 that young vintages grow strongly in both employment and output, whereas older vintages are contracting; so industries/occupations/establishments (IND/OCC/EST) with predominantly new vintages grow faster according to the model. To summarize, we would expect high tenure premia and low mean wages in fast-growing

IND/OCC/EST.

Table 5 shows the results of a regression of cross-sectional earnings on workers' (IND/OCC/EST-)tenure, IND/OCC/EST growth and quadratic and interaction terms of the two. The interaction terms are all positive, as predicted by the model: faster growing IND/OCC/EST have higher tenure premia. The coefficients on growth (g) have the expected negative sign for IND and EST, but a positive sign for OCC: faster-growing IND/OCC/EST have lower mean wages. So qualitatively, the regression results support the model's predictions.[25]

Table 5: Regressions on growth and earnings structure

Category	N	*const*	*ten*	$ten^2/100$	g	$g^2/100$	$ten \times g$
Industry	208,986	4.930	.055	-.141	**-.113**	.065	**.023**
Occupation	209,721	4.923	.060	-.164	**.189**	-.221	**.027**
Establishment	152,383	5.043	.039	-.098	**-.005**	.007	**.006**
Model: vintage			.060	-.078	**-.225**	.007	**.036**

Censored regression (top-coded data) of log earnings on a second-order polynomial in IND/OCC/EST tenure and IND/OCC/EST growth, both referring to the respective category. IND/OCC/EST-growth g is calculated as $\ln M(2000) - \ln M(1995)$, where M is the number of workers in the respective category. Tenure and earnings data are from 2000. All coefficients – except for the one on g for EST – are significantly different from 0 at the 1%-level.

The magnitude of the effects is considerable: an IND (OCC, EST) that grows by 1% on a yearly basis has a 5-year tenure premium that is 0.58 (0.67, 0.14) percentage points higher than that of a stagnant one. In an IND (OCC, EST) that grows by one standard deviation faster than a stagnant one, the 5-year tenure premium is 2.38 (2.42, 1.92) percentage points higher than in stagnant one.[26]

[25] These empirical patterns are in line with Michelacci & Quadrini's (2004) results from Finnish matched-employer-employee data. They find that in fast-growing firms returns to tenure are highest and starting wages are lowest. Their model explains this phenomenon by financial constraints that are especially severe for fast-growing firms, inducing firms to "borrow" from their workers by offering back-loaded tenure-earnings profiles.

[26] The x-year tenure premium in an IND with growth g is $tp_x(g) = \beta_{ten}x + \beta_{ten^2}x^2/100 + \beta_{ten/ind}xg$. Thus, the difference in the x-year premium between an industry with growth g_1 and g_0 is $\Delta tp_x = \beta_{ten,ind}x(g_1 - g_0)$. Setting $x = 5$, $g_0 = 0$ and $g_1 = 0.05$ (note that g is calculated over a 5-year period in the data, so yearly growth of 0.01 translates to $g = 0.05$) we obtain 0.0058 for IND. The standard deviations for g are 0.206 (IND), 0.181 (OCC) and 0.514 (EST), which yield the second set of numbers.

The last line in the table shows the regression results from the calibrated model. The coefficients are not too far away from their estimated counterparts, but the model somewhat over-predicts the interaction term between tenure and growth, i.e. it generates too steep tenure-earnings profiles in the fast-growing vintages. Furthermore, the coefficient on growth is lower than in the data, meaning that the model exaggerates the differences in average wages between fast-growing and slow-growing sectors. This problem was already mentioned when discussing the fit of the baseline calibration.

4 Conclusions

This paper has studied a model of vintage-human-capital accumulation that matches key facts on the tenure-earnings distribution in a German data set. It provides a promising avenue for understanding the systematic variation in the earnings structure across establishments, industries and occupations. In the following, some potential applications of the framework are briefly discussed.

A first proposed application is a macroeconomic one: the model relates the rate of embodied technological growth to the earnings structure, both at the industry and the economy-wide level. Previous versions of the paper had focused on this point, arguing that the steepening of age-earnings profiles and the concomitant rise in cross-sectional and time-series variance of earnings in many industrialized countries over the last decades could have been caused by a technological acceleration.

A second aspect worth mentioning, which has only been touched upon in the previous discussion, is the productivity profile of a vintage over time (see the lower-right panel of figure 4). It displays the typical back-loaded shape that is often posited in an ad-hoc fashion for organization capital (see Atkeson & Kehoe, 2005, for example). In fact, the model presented here can be construed as a micro-foundation for the way an organization increases its productivity over time and how it shares these productivity gains among its members.

Finally, one could study the riskiness of human capital and technology choice for workers by introducing a stochastic component into the framework.

References

Atkeson, A. & Kehoe, P. J. (2005), 'Modeling and measuring organization capital', *Journal of Political Economy* **113**, 1026–1053.

Boucekkine, R., de la Croix, D. & Licandro, O. (2006), Vintage capital, Economics Working Papers ECO2006/8, European University Institute.

Burdett, K. & Coles, M. (2003), 'Equilibrium wage-tenure contracts', *Econometrica* **71**(5), 1377–1404.

Chari, V. V. & Hopenhayn, H. (1991), 'Vintage human capital, growth, and the diffusion of new technology', *Journal of Political Economy* **99**(6), 1142–65.

Cummins, J. G. & Violante, G. L. (2002), Investment-specific technical change in the US (1947-2000): Measurement and macroeconomic consequences, Finance and Economics Discussion Series 2002-10, Board of Governors of the Federal Reserve System (U.S.).

Drews, N., Hamann, S., Köhler, M., Krug, G., Wübbeke, C. & Autorengemeinschaft 'ITM-Benutzerhandbücher' (2006), 'Variablen der schwach anonymisierten version der iab-beschäftigten-stichprobe 1975-2001', *Benutzerhandbuch des Forschungsdatenzentrum des Instituts für Arbeitsmarkt- und Berufsforschung* .

Garicano, L. & Rossi-Hansberg, E. (2008), 'Organizing growth', *Working Paper* .

Gordon, R. J. (1990), *The Measurement of Durable Goods Prices*, University of Chicago Press, Chicago.

Guvenen, F. & Kuruscu, B. (2007), 'A quantitative analysis of the evolution of the U.S. wage distribution: 1970-2000', *Working Paper* .

Hause, J. C. (1981), 'The fine structure of earnings and the on-the-job training hypothesis', *Econometrica* **49**(1), 277.

Huggett, M., Ventura, G. & Yaron, A. (2006), 'Human capital and earnings distribution dynamics', *Journal of Monetary Economics* **53**, 265–290.

Kushner, H. & Dupuis, P. (1992), *Numerical Methods for Stochastic Control Problems in Continuous Time*, Springer Verlag, Berlin and New York.

Ljungqvist, L. & Sargent, T. J. (1998), 'The european unemployment dilemma', *Journal of Political Economy* **106**(3), 514–550.

Luenberger, D. G. (1973), *Optimization Using Vector Space Methods*, AddisonWesley, New York, NY, USA.

Michelacci, C. & Quadrini, V. (2004), Financial markets and wages, 2004 Meeting Papers 116, Society for Economic Dynamics.

Nelson, R. R. & Phelps, E. S. (1966), 'Investment in humans, technological diffusion, and economic growth', *American Economic Review* **56**(1), 69–75.

Parente, S. L. (1994), 'Technology adoption, learning-by-doing, and economic growth', *Journal of Economic Theory* **63**(2), 346–369.

Prescott, E. C. & Boyd, J. H. (1987), 'Dynamic coalitions: Engines of growth', *The American Economic Review* **77**(2), 63–67.

Rudin, W. (1973), *Functional Analysis*, McGraw-Hill, New York, NY, USA.

Violante, G. L. (2002), 'Technological acceleration, skill transferability, and the rise in residual inequality', *Quarterly Journal of Economics* **117**(1), 297–338.

A Proofs

A.1 General setting

This subsection contains the proofs for the statements in section 2.4, where no additional assumptions are imposed on the production function.

A.1.1 Bounded resources

Lemma A.1. (Bounded resources) *There is a uniform bound $\bar{y} < \infty$ on $\tilde{Y}(n)$, $\int n \leq 1$. Thus, resources in the economy are bounded for each fixed t.*

Proof. Let $\Delta = \{n : \int n = 1\}$ be the unit simplex. By weak concavity of $\tilde{Y}(\cdot)$, the set $B = \{(n, Y) : n \in \Delta, Y \leq \tilde{Y}(n)\}$ is convex and has non-empty interior. Now, fix some interior point $\bar{n} \in \Delta$, say $\bar{n}(\cdot) = 1$. By the separating-hyperplane theorem, there is a bounded linear functional f on Δ such that $Y(n) \leq f(n)$ for all $n \in B$; in other words, all points in B must be in the half-space below the hyperplane $\{(n, Y) : n \in \Delta, f(n) = \tilde{Y}(\bar{n})$. Since f is bounded, we must have $\tilde{Y}(n) \leq M||n|| = M\}$ for all $n \in \Delta$ for some $M < \infty$ (the norm of f), where we use the norm $||n|| = \int |n|$ for the functions n. □

A.1.2 Proof of lemma 2.3: all jobs filled in producing vintage

Proof. $Y(\bar{\tau}) > 0$ implies that some open ball $B_\epsilon(\bar{\tau}, \bar{h})$ lies in the support of n for some $\bar{h} \in (0,1)$. If there was some h' such that $(\bar{\tau}, h')$ did not lie in the closure of n's support, then there would be a ball $B_{\epsilon'}(\bar{\tau}, h')$ with $\epsilon' \leq \epsilon$ in which wages must be infinity — if not, firms should optimally choose to employ some workers there. But then, any career segment passing through $B_{\epsilon'}(\bar{\tau}, h')$ would yield infinite wages yet could be reached with a finite cost, implying that $W = \infty$. This is clearly impossible since resources in the economy are bounded, see lemma A.1. □

A.1.3 Proof for lemma 2.4: finite support of technologies

Proof. Since there exists τ such that $w(\tau, 0) > 0$ (by the assumptions on $\tilde{Y}$ and $\tilde{w}$), there is a strictly positive flow value $\varepsilon > 0$ that a worker can secure by working continuously in $(\tau, 0)$. Now, we will argue that in very old vintages, this value cannot be provided to workers since TFP eventually goes below any positive bound.

Now, fix some old vintage S. Note that in equilibrium, the value of every career segment l' (which may be of finite or infinite length, and where we cut off parts in vintages younger than S) spent in vintages above S must exceed the value of working for ε — if not, the worker should certainly replace the segment by ϵ:

$$\tilde{v}(l') \equiv \int_{l_0}^{l_1} e^{(\gamma-\beta-\delta)t} w(t - s'(t), h'(t))dt \geq \int_{l_0}^{l_1} e^{(\gamma-\beta-\delta)t} \epsilon dt$$

The inequality must hold since since l' also includes non-negative human-capital-accumulation costs.

Now, observe that the value of all discounted career segments in vintages older than S has to be lower than total discounted wages and thus production in those vintages. Integrate the above inequality over all career segments of type l' in the economy:

$$\int_{\text{all } l'} \tilde{v}(l') \leq \int_0^\infty e^{(\gamma-\beta-\delta)t} \int_{-\infty}^{t-S} \int_0^1 n(t,s,h) w(t,s,h) dh ds dt \leq$$
$$\leq \bar{y} e^{-\gamma S} \int_0^\infty e^{(\gamma-\beta-\delta)t} \left(\int_{s,h} n(t,s,h) \right) dt,$$

where in the last step I used that the upper bound on production for vintages even older than S is at most $e^{-\gamma S}\bar{y}$ for some $\bar{y} < \infty$, see lemma A.1) for a proof.

On the other hand, we know that each agent must weakly prefer working in an old vintage to working for ϵ — again, integrating up over all segments we get:

$$\int_{\text{all } l'} \tilde{v}(l') \geq \epsilon \int e^{(\gamma-\beta-\delta)t} \left(\int_{s,h} n(t,s,h) \right) dt$$

But combining the above inequalities yields a contradiction: by choosing S large enough, we can make $e^{-\gamma S}\bar{y} < \epsilon$, making it impossible that very old vintages provide enough value to be attractive to workers. □

A.1.4 Proof of lemma 2.7: no holes in vintage space

Proof. Suppose there was some $\tau' \in (\tau_0, \tau_1)$ for which $Y(\tau') = 0$. Then there must be a positive measure of career segments ending on $[\tau_0, \tau')$ and the final wages of these segments must be equalized, which implies that all agents leave the vintage at once for some $\tau_e = \sup\{\tau : Y(\tau > 0)\}$ and that $w(\tau_e, h) = e^{-\gamma\tau_e}\bar{y} = (\beta + \delta - \gamma)W$ for all h. But this contradicts the fact that $w(T^*, h) = e^{-\gamma T^*}\bar{y} = (\beta + \delta - \gamma)W$ since $T^* \geq \tau_1 > \tau'$. □

A.2 Substitutability in production function: $\rho = 1$

This subsection contains the proofs for the statements in section 2.7, throughout which the production function is assumed to be linear.

A.2.1 Always enter newest vintage

Lemma A.2. (Always enter newest vintage under substitutability) *If $\tilde{Y} = \tilde{Y}_{CES}$ with $\rho = 1$, then for any optimal life $h(t) = 0$ implies $\tau(t) = 0$ at the beginning of segments and $\tau(t) = 0$ almost everywhere on non-segments.*

Proof. Suppose the worker chose a career segment with $s(t_1) > t_1$ on $t \in [t_1, t_2)$. Then this career is strictly dominated by choosing the same career in $s(t_1) = t_1$. Obviously, the same holds true for choosing $s(t) > t$ and $h(t) = 0$ on non-segments of positive measure. □

A.2.2 Paths cross at most once

Lemma A.3. (Paths cross at most once) *Assume $\tilde{Y} = \tilde{Y}_{CES}$ with $\rho = 1$. Suppose that $\tilde{\gamma} \neq \gamma$, but fix all other parameters. Let $h(\cdot)$ and $g(\cdot)$ be optimal careers given productivity growth γ and $\tilde{\gamma}$, respectively. Then we have:*

$$h(t) \geq g(t) \text{ and } \dot{h}(t) < \dot{g}(t) \;\Rightarrow\; h(s) > g(s) \text{ for all } s < t.$$

Proof. Suppose that the paths crossed again and denote by s the first crossing point, i.e. $s = \max_{u<t}\{u : h(u) \leq g(u)\}$. Together with $h(t) \geq g(t)$ this implies

$$h(t) - h(s) \geq g(t) - g(s) \Rightarrow \int_s^t \dot{h}(u)du \geq \int_s^t \dot{g}(u)du, \tag{11}$$

i.e. h must grow by at least as much as g over the interval to end up above g. By the assumption on the wage function, $w_h(h)$ is a decreasing function in h. Using the FOC (3), this implies that for all $u \in (s, t)$, we have

$$\begin{aligned} c'\big(\dot{h}(u)\big) = \int_u^t e^{-\beta v} w_h[h(v)]dv + e^{-\tilde{\beta}(t-u)} c'\big(\dot{h}(t)\big) < \\ < \int_u^t e^{-\tilde{\beta} v} w_h[g(v)]dv + e^{-\tilde{\beta}(t-u)} c'\big(\dot{g}(t)\big) = c'\big(\dot{g}(u)\big) \end{aligned}$$

since by assumption $\dot{h}(t) < \dot{g}(t)$ and $w_h[h(v)] \leq w_h[g(v)]$ point-wise; the inequality follows from c' being increasing. This again implies $\dot{h}(u) < \dot{g}(u)$ for all u, which in turn contradicts (11). □

A.2.3 Proof of proposition 2.8: shorter horizon lowers tenure premia ($\rho = 1$)

Proof. Without loss of generality, take two optimal career segments $\tilde{h}$ and h in vintage $s = 0$ starting with $\tilde{h}(0) = h(0) = 0$ and $\tilde{T}^* > T^*$. Now suppose that $\tilde{h}(T^*) \leq h(T^*)$. First, note that $\dot{h}(T^*) = 0$ but $\dot{\tilde{h}}(T^*) > 0$ by equation (7) and the fact that $c'(\dot{h}) = V_h$. By lemma A.3, the two paths cannot cross again for any $0 \leq t > T^*$. But this is a contradiction to $h(0) = \tilde{h}(0) = 0$. By the same argument, the two paths cannot intersect at any other point $0 < t < T^*(\gamma)$. So we must have $\tilde{h}(t) \geq h(t)$ and so $w(\tilde{h}(t)) \geq w(h(t))$, which implies the desired result. □

A.2.4 Proof of proposition 2.9: faster growth shortens careers and lowers tenure premia

Proof. Let $Z(\gamma)$ be the value of being an inexperienced worker at time 0 given vintage productivity growth γ. The worker's problem is then to choose the switching time T when to leave the vintage to maximize

$$\tilde{Z}(\gamma, T) = K(T) + e^{-(\beta-\gamma)T} Z(\gamma),$$

where $K(\cdot)$ is given in (8). Invoking the assumption that $K(\cdot)$ is twice differentiable, the derivatives are computed as

$$\tilde{Z}_T(\gamma, T) = K'(T) - (\beta - \gamma)e^{-(\beta-\gamma)T} Z(\gamma) \tag{12}$$

$$\tilde{Z}_{TT}(\gamma, T) = K''(T) + (\beta - \gamma)^2 e^{-(\beta-\gamma)T} Z(\gamma), \tag{13}$$

The FOC for the optimal career length $T^*(\gamma)$ is $\tilde{Z}_T(\gamma, T^*(\gamma)) = 0$, the SOC is $\tilde{Z}_{TT}(\gamma, T^*(\gamma)) < 0$.

I will now state the problem in slightly different terms, which will enable us to derive how $Z^*(\gamma) \equiv Z(\gamma, T^*(\gamma))$ changes as γ changes. Note that since the worker's problem is recursive, we can write his value as $\tilde{Z}(\gamma, T) = K(T)/(1 - e^{-(\beta-\gamma)T})$. $T^*(\gamma)$ maximizes the function $\tilde{Z}(\gamma, T)$ for a given γ — indeed, the first-order conditions yield just the same result as in the problem above when maximizing $\tilde{Z}(\gamma, \cdot)$. But the formulation here is much more handy to see what happens to the agent's value when we change γ:

$$\frac{\partial Z^*(\gamma)}{\partial \gamma} = \left.\frac{dZ(\gamma, T^*(\gamma))}{d\gamma}\right|_{\gamma, T^*(\gamma)} = \frac{e^{-(\beta+\gamma)T^*}}{1 - e^{-(\beta-\gamma)T^*}} T^* Z(\gamma, T^*)$$

where the envelope condition $\tilde{Z}_T(\gamma, T^*(\gamma)) = 0$ is used.

Now, re-state the first-order condition for $T^*(\gamma)$ from (12):

$$K'\big(T^*(\gamma)\big) = (\beta-\gamma)Z^*(\gamma)e^{-(\beta-\gamma)T^*(\gamma)}$$

Take the total derivative of this equation with respect to γ and use (13) to obtain

$$\frac{dT^*}{d\gamma}\underbrace{\tilde{Z}_{TT}\big(\gamma, T^*(\gamma)\big)}_{<0 \text{ by SOC (13)}} = \underbrace{(\beta-\gamma)e^{-(\beta-\gamma)T^*(\gamma)}Z^*(\gamma)}_{>0}\Bigg[\underbrace{\frac{T^*(\gamma)}{1-e^{-(\beta-\gamma)T^*(\gamma)}} - \frac{1}{\beta-\gamma}}_{\equiv\Phi_\gamma(T^*)}\Bigg] \tag{14}$$

We see that if T^* is large, also Φ_γ grows large, implying that also the effect $dT^*/d\gamma$ is negative and large in absolute value. When taking $T^* \to 0$ and using L'Hopital's rule, one finds that $\Phi_\gamma \to 0$, implying that the effects on T^* become very small.

The derivative of Φ_γ in T^* is

$$\Phi'_\gamma(T^*) = \frac{1 - \frac{1+(\beta-\gamma)T^*}{e^{(\beta-\gamma)T^*}}}{(1-e^{-(\beta-\gamma)T^*})^2}.$$

Note that in the numerator, $1 + (\beta-\gamma)T^*$ is nothing but the first-order Taylor expansion of the function $e^{(\beta-\gamma)T^*}$ in T^* around 0, which always stays below the function itself since the exponential function is convex. This implies that the fraction in the numerator is always smaller than one, implying that Φ_γ is globally increasing. This in turn implies $\Phi_\gamma > 0$ (recall that $\lim_{T^*\to 0}\Phi(T^*) = 0$), which tells us we have $dT^*/d\gamma < 0$ for all $\gamma > 0$.[27]

There may exist values of γ where $T^*(\gamma) = 0$; in this case, the statements in the proposition are trivial.

Finally, since $T^*(\gamma)$ is a decreasing function it follows from proposition 2.8 that $p_\gamma(t)$ is decreasing in γ for any fixed $t > 0$. □

A.3 Complementarity in production ($\rho < 1$)

This subsection contains the proofs for the statements in section 2.7, throughout which an Inada condition is invoked on the production function.

A.3.1 Proof of lemma 2.10: vintage T^* has highest entry wage

Proof. By corollary 2.6, $w(T^*)/(\beta+\delta-\gamma) = W$, i.e. always working in the oldest vintage as an unskilled worker is an optimal strategy. Suppose $w(\tau, 0) > w(T^*, 0)$ for some τ. Then always working in position $(\tau, 0)$ would give value $(\beta+\delta-\gamma)w(\tau,0) > W$, which contradicts W being the maximal attainable value for a career starter. □

[27] Note that these calculations fail to provide us with any upper bound on $dT^*/d\gamma$, so in principle this change can be arbitrarily large.

A.3.2 Proof of lemma 2.11: wage explosion for the skilled in young technologies

Proof. First, I show that the cost $C(\Delta h, \Delta t)$ of accumulating human capital Δh in a time interval Δt goes to infinity for fixed Δh when letting $\Delta t \to 0$. By Jensen's inequality, the minimal cost of accumulating Δh within Δt is by setting a constant $\dot{h} = \Delta h/\Delta t$ throughout Δt. Then $C(\Delta h, \Delta t) \geq c(\Delta h/\Delta t)\Delta t \to \infty$ if $\Delta t \to 0$ by our assumption on $c(\cdot)$.

By lemma 2.7, S_n must be a rectangle $(T_0, T^*) \times [0, 1]$. Now, suppose there was no singularity for w in the upper left corner and $w(T_0, 1) < \infty$. Then, by continuity of w, for each ϵ there is a ball $B_\delta(T_0, 1)$ in which wages deviate not more than ϵ from $w(T_0, 1)$. So the parts of any career segment contained in $B_\delta(T_0, 1)$ yield bounded wage payments. But we can definitely find a sequence of careers for which learning costs inside the ball exceed any bound. To see this, set $\Delta h = \delta$, take a sequence $\Delta\tau \to T_0$ and note that the cost of reaching $(T_0 + \Delta\tau, 1)$ inside B must go to infinity. Note also that a positive measure of workers must take such paths since no region is empty by lemma 2.3. But then, those workers cannot behave optimally and should change their h-path through the ball B, which is inconsistent with equilibrium.

This also implies that $T_0 = 0$. If this was not the case, then workers with careers in $B_\delta(T_0, 1)$ should reach those by choosing flatter careers entering at $\tau = 0$, which would imply that those careers could achieve unbounded value by the above argumentation. This contradicts $W < \infty$. □

A.3.3 Proof for corollary 2.12: wage compression

Proof. The first statement follows from $w(\tau, 0) \leq w(T^*, 0)$ for all $\tau < T^*$ (see lemma 2.10) and $w(\epsilon, 1) \to \infty$ (see lemma 2.11). The second statement follows from corollary 2.6. □

A.3.4 Proof for corollary 2.13: obsolescence/wage losses

Proof. The first statement follows from the reasoning laid out in lemma 2.11: there is a positive measure of agents with high human capital $h \in [1 - \epsilon, 1]$ in young vintages $\tau \in (0, \epsilon]$ with a high wage $w(\tau, h) > M$, M large, which must experience wage losses once they leave the high-wage region. The second statement is an obvious consequence of lemma 2.10 and corollary 2.6. □

A.4 Planner's problem: formal analysis

Consider a social planner who weighs the utility of an agent born at t with $e^{-\beta t}$. Since it costs the planner $e^{-\delta(u-t)}$ units of time-u output to supply one unit to each surviving member of a cohort born at t and since utility is linear for all agents, it is easy to see that the planner's criterion is then to choose a function $n(t, s, h)$ (which we require again to be C^1 on a given support S_n) to maximize

$$J(n) = \int_0^\infty e^{-\beta t} \left(Y\left(n(t;\cdot)\right) - C(t)\right) dt,$$

where $C(t)$ denotes the aggregate cost of human-capital accumulation at t. First, I will derive an expression for $C(t)$ given the optimal strategy to implement a given density n.

A.4.1 Optimal promotion strategy

It turns out that the optimal promotion strategy is such that agents' career paths inside a vintage never cross. Intuitively, if a positive measure of agents crossed each other's way, then one could improve upon the strategy by maintaining the ordering inside the vintage, making agents go shorter paths and hence lowering total cost for the planner.

Lemma A.4. (No-crossing measure is optimal) *For a given density $n(t, s, h)$, it is optimal for the planner not to let career paths cross when implementing the density. This means that the planner makes workers follow paths $h(t+u)$ for any given t, any vintage s and any $u \in (0, T-t)$ such that $N[t+u, s, h(t+u)] = \exp(-\delta u)N[t, s, h(t)]$.*

Proof. I will proceed constructively to engineer the optimal measure on life paths by a discrete approximation procedure. Cut time and vintages into intervals of length $2^{-k}T^*$ for $k = 1, 2, \ldots$ to obtain grids $\{t_i^{(k)}\}_{i=1}^\infty$ and $\{s_i^{(k)}\}_{i=1}^{N_s}$. For human capital, slice such that the points $\{h_i^{(k)}\}_{i=1}^{N_s}$ yield intervals of length 2^{-k}. Approximate every path by connecting the middle of the interval $[h_i, h_{i+1}]$ it passes through at t_i for $t = 0, 2^{-k}, \ldots$ with straight lines. For every given measure μ on lives, summing up the costs over all possible promotion paths weighted by the densities induced by the measure μ gives us an approximation $C_k(\mu)$ for the total cost of human-capital accumulation for this μ.

Now, we will construct a lower bound C_k^* on this cost for a fixed iteration k in the algorithm. Note that it is enough to consider the task of moving workers between t_i and t_{i+1} for each point in time. It does not matter how we combine these path segments sequentially later, any such combination must obviously yield the same value.

Without loss of generality, consider the case $k = 1$ for $t_1 = 0$ and $t_2 = 1$ for the vintage $s = 1$ (note that the case $s = 0$ is trivial). The claim is that it cannot be optimal to choose a promotion scheme under which the paths of a positive measure of agents cross. Suppose we chose a promotion scheme under which a positive measure of agents crossed, i.e. a measure $\bar{\epsilon}$ went from $\bar{h}_0$ to h_1 and a measure $\epsilon > 0$ from h_0 to $\bar{h}_1$, where all the mentioned h-levels are center points of the approximation grid, and where $\bar{h}_j > h_j$. Now, set $\epsilon' = \min\{\epsilon, \bar{\epsilon}\}$ and consider the alternative of moving ϵ' agents from $\bar{h}_0$ to $\bar{h}_1$ and the measure ϵ' from h_0 to h_1. This would dominate the original allocation because of the following argument: take z to be the intersection of the lines $\bar{h}_0$ to h_1 and h_0 to $\bar{h}_1$. Then, clearly the process of sending everybody to z but then exchanging the flows to keep workers positions in the hierarchy fixed is just as cheap as the original policy. However, notice that this new policy must be weakly inferior to sending workers on the direct line $\bar{h}_0$ to $\bar{h}_1$ and h_0 to h_1, since this is the cost-minimizing strategy by Jensen's inequality.

Also, notice that there always exists a policy which does not make any worker flows cross: first, fill the uppermost interval at $t = 1$ with the uppermost workers from $t = 0$; proceed by filling the second interval with the uppermost workers left at $t = 0$ after the first step, and so forth. It is also clear that any process that does not follow these rules must make some workers cross and that any such process can be rendered into the proposed no-crossing algorithm by a finite number of improving operations; this shows that the no-crossing mechanism is optimal for a fixed k.

Obviously, the values C_k^* converge to the value of implementing the no-crossing measure μ_{nc}. Now, observe that no other measure μ' can yield a cost strictly lower than this: if we approximate μ' by the above scheme, by the above argument it must be that $C_k(\mu') \geq C_k(\mu_{nc})$. This precludes $C(\mu') = \lim_{k\to\infty} C_k(\mu') < \lim_{k\to\infty} C_k^* = C^*$.

It remains to prove that the lines of the no-crossing measure follow the proposed law. By the algorithm above, it is clear that an agent who at t had $N(t, s, h)$ workers above himself (position h) in vintage s and survives until $t + u$ will have $\exp(-\delta u)N(t, s, h)$ workers above himself at $t+u$ if none of the other workers crosses his path. This proves the second claim of the statement. □

In the following, it will prove useful to work with the anti-cdf $N(t, s, h) \equiv \int_h^1 n(t, s, \tilde{h})d\tilde{h}$. In a scheme where agents' paths do not cross, this function must decrease at the death rate δ when we evaluate it along an agent's path staying in a fixed vintage s. A first-order approximation following a career line $\{h(t), \tau(t)\}$ yields:

$$N_t(t, s, h) + \dot{h}(t, s, h)N_h(t, s, h) = -\delta N(t, s, h), \tag{15}$$

where we note that $N_h = -n$. Taking the h-derivative of the above and imposing stationarity yields the PDE for the evolution of n, which we already know from competitive equilibrium, see equation (6).

Re-arranging equation (15) gives us an expression for the career slope $\dot{h}$ that the planner should choose given that she wants to implement a given n:

$$\dot{h}(t, s, h) = \frac{N_t(t, s, h) + \delta N(t, s, h)}{n(t, s, h)}. \tag{16}$$

In order to aggregate costs over all agents, we have to weigh the cost of $\dot{h}$ by the mass of agents across the (t, s, h)-space and obtain $C(t) = \int_{s,h} n(t, s, h)c[\dot{h}(t, s, h)]$.

A.4.2 The planner's first-order conditions

The strategy to obtain the first-order conditions (FOCs) for the planner's problem is as follows: I will first allow the planner to to choose any – possibly time-varying – density $n(t, s, h)$. I then look for a stationary distribution which solves this unrestricted problem. This ensures that the planner would not want to deviate from the stationary density $n(\tau, h)$ although she could do so. I will first restrict S_n to the entire rectangle below a maximal vintage age T and then let T vary to find the optimal support T^*.

It turns out that it is useful to introduce the variable $u(t,s,h) \equiv n_t(t,s,h)$ and connect it to the functions n, N and N_t with equality constraints. The Lagrangian is then[28]

$$
\begin{aligned}
\mathcal{L} = \int_0^\infty e^{-\beta t} \Bigg[\int_{t-T}^{t} Y(t,s) - e^{\gamma s} \bigg(\int_0^1 c\big[\dot{h}(t,s,h)\big] n(t,s,h) dh \bigg) ds \Bigg] dt + \\
+ \int_{t,s,h} e^{-(\beta-\gamma)t} \Bigg[\nu(t,s,h) \bigg(\dot{h} - \frac{\dot{N} + \delta N}{n} \bigg) + \\
+ \lambda(t,s,h) \bigg(n_0(s,h) + \int_0^t u(\tilde{t},s,h) d\tilde{t} - n(t,s,h) \bigg) + \\
+ \eta(t,s,h) \bigg(\dot{N}(t,s,h) - \int_h^1 u(t,s,\tilde{h}) d\tilde{h} \bigg) + \\
+ \xi(t,s,h) \bigg(N(t,s,h) - \int_h^1 n(t,s,\tilde{h}) d\tilde{h} \bigg) + \\
+ \mu(t) \bigg(1 - \int_{t-T}^{t} \int_0^1 n(t,s,h) dh ds \bigg) dt \Bigg],
\end{aligned}
$$

where the Lagrange multipliers are scaled by $e^{-(\beta-\gamma)t}$ to render them stationary. The set of constraints linked to the multipliers ν is taken from equation (16). The constraints connected to μ enforce that total population not exceed the bound 1. The rest of the constraints link the various variables related to the density n.

The FOC with respect to $\dot{N}(t,s,h)$, $\dot{h}(t,s,h)$ and $N(t,s,h)$ immediately tell us that η is the marginal cost of human-capital accumulation, and that ν and ξ are closely linked to η:

$$
\begin{aligned}
\eta(t,s,h) &= e^{-\gamma\tau} c'\big(\dot{h}(t,s,h)\big) \\
\nu(t,s,h) &= e^{-\gamma\tau} c'\big(\dot{h}(t,s,h)\big) n(t,s,h) \\
\xi(t,s,h) &= \delta e^{-\gamma\tau} c'\big(\dot{h}(t,s,h)\big).
\end{aligned}
\tag{17}
$$

Using these equalities, the FOC with respect to $n(t,s,h)$ becomes

$$
\begin{aligned}
\lambda(t,s,h) =& w(t,s,h) - e^{-\gamma\tau} c\big(\dot{h}(t,s,h)\big) + e^{-\gamma\tau} \dot{h}(t,s,h) c'\big(\dot{h}(t,s,h)\big) - \\
& - \mu(t) - \delta \int_0^h \eta(t,s,\tilde{h}) d\tilde{h}.
\end{aligned}
\tag{18}
$$

[28]See Luenberger (1973) for necessary conditions of constrained-optimization problems in infinite-dimensional spaces.

where we recognize in the terms involving $c(\cdot)$ the Hamiltonian from the value function (2) in the worker's problem. The last remaining derivative is the one with respect to to $u(t,s,h)$, which will prove crucial to obtain the PDE that is equivalent to the HJB (2):

$$\int_{\tau}^{T} e^{-(\beta-\gamma)(\tilde{\tau}-\tau)}\lambda(\tilde{\tau},h)d\tilde{\tau} = \int_{0}^{h} \eta(\tau,\tilde{h})d\tilde{h}. \tag{19}$$

At a stationary solution, we require that the density fulfill $n(t,s,h) = \bar{n}(\tau,h)$. As a consequence wages grow at rate γ: $w(t,s,h) = e^{\gamma t}\bar{w}(\tau,h)$. The Lagrange multipliers must also be time-independent, i.e. $\nu(t,s,h) = \bar{\nu}(\tau,h)$, $\mu(t) = \bar{\mu}$ and so forth. Again, I drop the bar-notation in the following.

When substituting the expressions for the Lagrange multipliers (17) and (18) into (19) and imposing stationarity, one obtains

$$\begin{aligned}\int_{\tau}^{T} e^{-(\beta-\gamma)(\tilde{\tau}-\tau)} \Big[w(\tilde{\tau},h) - e^{-\gamma\tilde{\tau}}c\big(\dot{h}(\tilde{\tau},h)\big) + \dot{h}(\tilde{\tau},h)c'\big(\dot{h}(\tilde{\tau},h)\big) - \mu - \\ -\delta\int_{0}^{h}\eta(\tilde{\tau},\tilde{h})d\tilde{h}\Big]d\tilde{\tau} = \int_{0}^{h} e^{-\gamma\tau}c'\big(\dot{h}(\tau,\tilde{h})\big)d\tilde{h} \equiv \Lambda(\tau,h).\end{aligned} \tag{20}$$

We will now see that $\Lambda(\tau,h)$ is an "excess-value function": it tells us what the value of an agent to the planner in position (τ,h) is *in excess* of the unconditional value μ of an additional unskilled agent.

Directly from (20), we can get the following insights: First, when $\tau \to T$, the left-hand side and with it the marginal cost of human-capital accumulation $c'(\dot{h})$, and hence $\dot{h}$ itself, go to zero. This says that one should not accumulate human capital anymore just before the vintage shuts down, which also implies that $w(T,h)$ must be weakly increasing in h by non-negativity of the multipliers η. Second,when we let $h \to 0$, the right-hand side of (20) goes to zero and we see that $\lambda(\tau,0) = 0$ for all τ. This says that for all entry jobs the value function must be equalized. Third, when we let both $\tau \to T$ and $h \to 0$ and use the insights from above, we obtain $w(T,0) = \mu$. This says that $w(T,0)$ is the reference wage of the economy: it does not provide any valuable experience, so it has to be just as attractive per se as any other career (in flow terms).

Now, take the derivatives of Λ in (20) in both dimensions to see how this excess-value function behaves on the interior:

$$\Lambda_h(\tau,h) = e^{-\gamma\tau}c'\big(\dot{h}(\tau,h)\big) \tag{21}$$

$$-\Lambda_\tau(\tau,h) = w(\tau,h) - e^{-\gamma\tau}c(\dot{h}) + e^{-\gamma\tau}\dot{h}c'(\dot{h}) - \mu - (\beta+\delta-\gamma)\Lambda(\tau,h). \tag{22}$$

When adding an agent's value at the start of a career segment $W = \mu/(\beta+\delta-\gamma)$ to Λ by defining $V = \Lambda + W$, we obtain

$$-V_\tau(\tau,h) = w(\tau,h) - e^{-\gamma\tau}c(\dot{h}) + \dot{h}V_h - (\beta+\delta-\gamma)V(\tau,h), \tag{23}$$

where we use $V_h = \Lambda_h = c'(\dot{h})$. When imposing the boundary conditions $V(\tau,0) = V(T,h) = 0$ for all τ and for all h, this system is the same as the agent's HJB (2) and its boundary conditions in the decentralized problem. The proof for proposition 2.14 in section A.4.6 will discuss equivalence of the planner's problem to the competitive equilibrium more carefully; before, it is useful to analyze the effects of variations in T.

A.4.3 Uniqueness

A fundamental concavity argument allows us to establish uniqueness of the planner's solution:

Proposition A.5. (Solution to planner's problem is unique) *If $\tilde{Y}(n_{t,s}(\cdot))$ is strictly concave in $n_{t,s}(h)$*[29]*, then $J(n)$ is strictly concave in n and there is at most one density $n(t,s,h)$ that maximizes $J(n)$.*

Proof. Suppose there were two maximizers n_1 and n_2. Clearly, a convex combination $n_\lambda = \lambda n_1 + (1-\lambda)n_2$ would also be feasible. Implementing n_λ in terms of promotion costs would be at least as cheap as implementing λn_1 and $(1-\lambda)n_2$ separately and adding up the costs. Output, however, will be strictly larger for each fixed pair (t,s) by the concavity assumption on $\tilde{Y}(\cdot)$, which implies the desired result. □

It is worthwhile to note that this argument does not hinge on the assumption of n being continuous or differentiable, nor on any restriction on S_n.

If Y is not strictly concave, matters are slightly more complicated. Take the example from subsection 2.6 with a linear production function: uniqueness of the planner's problem depends on uniqueness of the partial-equilibrium solution for the agent. If the agent's problem has a unique solution for any starting value of h, then the solution to the planner's problem is unique.

Existence of equilibrium is not a problem computationally, but could not be established formally without making an equicontinuity assumption on the function space for n; see the following section for a discussion.

A.4.4 Existence of solution to the planner's problem

In order to reap the benefits of compactness, we may restrict ourselves to seek a maximand n in the planner's problem that satisfies the following conditions: we re-parameterize the density from $n(t,s,h)$ to $n(t,\tau,h)$, which ensures that the partial derivative $n_t \to 0$ everywhere as $t \to \infty$ for any n that converges to a stationary distribution. Then, compactify the t-dimension using an increasing concave transform that maps $[0,\infty) \to [0,1)$ and define $\lim_{t\to\infty} n(t,\tau,h)$ as $\tilde{n}(1,\cdot)$. We then impose a Lipschitz condition uniformly on the entire family of $\tilde{n}$ in which we look for a maximand (This essentially means that the modulus of continuity for the original n becomes always stricter in the t-direction as t increases; the "wiggling" in n has to become smaller as t grows).

[29] For the CES case, this is equivalent to assuming $\rho < 1$.

If we further assume that n is point-wise bounded – which is unproblematic – , equicontinuity allows us to employ the Arzela-Ascoli theorem which tells that such a family of functions $\tilde{n}$ is a compact set; see Rudin (1973) for a statement of the theorem. The computational exercises indicate that indeed the optimizer n^* satisfies a Lipschitz condition; decreasing the grid size to allow for always steeper functions n does not significantly alter the solution after some point. However, it is hard to prove that the solution really satisfies such a Lipschitz condition.

A.4.5 Varying T

So far, we had fixed the maximal vintage age T and imposed it on the planner; we will now be concerned with varying T and finding the optimal T^* under the assumption that $\tilde{Y}$ is strictly concave. By the concavity argument in lemma A.5, there is at most one T^* for which the planner's criterion is maximized. An argument analogous to the proof for 2.4 shows that $T^* < \infty$. However, it is very hard to further characterize T^*. Computationally, it may be found by finding the optimal n for each fixed T and then pick the value T^* that yields the highest value to the planner. The following discussion describes regularities and problems that arose in this process.

First, for $T < T^*$, the simulations usually yield the wage structure is not flat in the last vintage yet. In this case, an argument along the lines of corollary 2.6 shows that it is preferable for the planner to extend the vintage horizon T marginally; marginal productivities for different h-levels are not aligned yet and there is room for further gains through human-capital accumulation.

Second, for $T > T^*$ computational problems may arise because of the following issue: the problem of finding the optimal n given T will usually not have a maximand in the space of continuous differentiable functions. To see this, suppose there was such a maximand $n^*(T)$. Since $J(n^*(T^*)) > J(n^*(T))$, by concavity also $J(n_\lambda) > J(n^*(T))$ where we define $n_\lambda = \lambda n^*(T^*) + (1-\lambda)n^*(T)$ for any $\lambda \in (0,1)$. In turn, any n_λ may be approximated arbitrarily well by any continuous, differentiable n with support until T. So there is a sequence of densities for which J converges to the global optimum, but the global optimum is not in the space we are considering since its support only extends to $T^* < T$ and is discontinuous at this point.

A.4.6 Proof for proposition 2.14: equivalence of CE and planner's problem

Proof. I will first show that the global solution to the planner's problem constitutes a CE. Set wages $w(\tau, h) = \partial Y(\tau)/\partial n(\tau, h)$ for $\tau \leq T^*$ and $w(\tau, h) = w(T^*, 0) = \mu$ for all $\tau > T^*$, all h. This implies that firms optimally choose not to produce for $\tau > T^*$ since even the cost-minimizing input combination leads to losses. For $\tau \leq T^*$, $n(\tau, h)$ is an optimal input choice and profits are zero. For agents, the HJB (23) and its boundary conditions imply that any career segment which fulfills $\dot{h} = V_h$ everywhere is an optimal strategy with starting value μ. This weakly dominates any career segment in vintages (T^*, ∞). One may then insert agents into careers to engineer the entry density $n(\tau, 0)$ since agents are indifferent between all careers. Equation (6) ensures that the density n reproduces itself given the optimal decisions of agents.

Second, I prove that any CE is a solution to the planner's problem for some $T \leq T^*$. To start, note that the worker's HJB (2) and the corresponding optimal policy (3) in competitive equilibrium have their exact counterparts in equations (23) and (22) for the planner's problem. Equation (20) follows by integrating from the boundary over τ and h, which in turn is equivalent to (19). Since the first-order conditions (17) and (18) can be used to define the Lagrange multipliers, equation (19) already ensures that all first-order conditions for the Lagrangian hold for any competitive equilibrium.

This means that any competitive equilibrium is a stationary point of the Lagrangian.[30] However, there can be at most one stationary point for a given T since J is a concave function and the set of permissible n is convex. Hence this stationary point must be the global maximum of the planner's problem corresponding to the T induced by the respective CE. As the discussion in A.4.5 showed, no such maximizer exists for $T > T^*$, which means that there cannot be any CE with $T > T^*$. □

It is hard to formally rule out competitive equilibria with $T < T^*$. If there is such a CE, then it must be that $\mu_T > \mu_{T^*}$ since these multipliers equal wages in the last vintage. This seems to suggest that $J_T > J_{T^*}$, which would be a contradiction to T^* being associated with a global maximizer. However, as the discussion in A.4.7 will show, J also includes the excess value for agents already born at $t = 0$ starting with $h(0) > 0$, which is not comprised in the multiplier μ.[31]

A.4.7 Decomposing the planner's criterion

Thinking along the lines of the planner's problem also proves useful in assessing vintage productivities. First, note that we can decompose the planner's criterion by integrating over the single agents' values:

$$J = \frac{w(T,0)}{\beta + \delta - \gamma} + \underbrace{\int_{\tau,h} \Lambda(\tau, h)}_{\equiv \bar{\Lambda}} + \int_0^\infty e^{-(\beta-\gamma)t} \delta \frac{w(T,0)}{\beta + \delta - \gamma} dt = \frac{w(T,0)}{\beta - \gamma} + \bar{\Lambda}$$

where the first equality decomposes the value for the measure one of agents alive at $t = 0$ according to $V = W + \Lambda$ and uses the fact that $\Lambda = 0$ for all agents born later. We can juxtapose this decomposition and the decomposition of J into

[30] A stationary point is defined as a point where the Frechet-derivative is zero in all directions, see Luenberger (1973) — this is the equivalent to the gradient being zero in $\mathbb{R}^n$.

[31] In the numerical exercises, however, enforcing $T < T^*$ always led to an increasing wage structure at T which is not compatible with a CE according to corollary 2.6.

production and promotion costs:

$$Y - C = (\beta - \gamma)J = w(T, 0) + (\beta - \gamma)\Lambda,$$

where we write $Y = Y(0)$ and $C = C(0)$. Since $\Lambda \geq 0$ and $C \geq 0$, this equation says that labor productivity in the last vintage $w(T, 0)$ is lower than average labor productivity Y in the overall economy. Furthermore, it gives us upper bounds for both C and Λ that can be empirically assessed by observing productivity in dying vintages and average productivity in the economy:

$$C \leq Y - w(T, 0), \qquad \Lambda \leq \frac{Y - w(T, 0)}{\beta - \gamma}.$$

B Computational algorithm

The following method discretizes the state space into a finite number of vintages and a finite number of rungs in the skill ladder. The algorithm can be interpreted as introducing a random element to skill accumulation (see Kushner & Dupuis, 1992, on the approximation of continuous-time models by discrete-time Markov chains). The algorithm is not only useful compute an approximation to the equilibrium, but also to form some intuition about the value function, agent's paths and other objects of the (continuous) model. The death probability δ is set to zero to simplify the exposition; of course, all arguments presented here also apply to the case $\delta > 0$. Throughout, we only consider stationary allocations, i.e. variables depend only on (τ, h) but not on t.

First, construct a discrete grid on the rectangle $(0 \leq \tau \leq T, 0 \leq h \leq 1$ as follows: divide the vintages into S sub-intervals (of equal size $\Delta\tau$) and the experience levels into h sub-intervals (of equal size Δh). The center points of these intervals are denoted by $\{\tau_i\}_{i=1}^{T}$ and $\{h_j\}_{j=1}^{S}$.

To approximate skill approximation choices $\dot{h}$, we linearly interpolate the value function between adjacent cells. If the grid is such that workers climb less than Δh over a time interval of $\Delta\tau$ in all cells, than linear interpolation is equivalent to the following "stochastic careers": set the probability $p(\tau_i, h_j)$ that the agent moves one box up (from h_j in vintage τ_i to h_{j+1} in vintage τ_{i+1}, that is) such that the expected slope of his career equals $\dot{h}(\tau_i, h_j)$, but that it does not exceed one:

$$p(\tau_i, h_j) = \min\left\{\dot{h}(\tau_i, h_j)\frac{\Delta\tau}{\Delta h}, 1\right\}$$

This means that in order to be able to replicate very steep slopes in this fashion, we need to make the slope $\Delta h/\Delta\tau$ become successively greater as k grows. I will make the following limiting argument: if we have an infinite sequence of discrete approximations as described above, choose the number of grid points as $S_k = kS_0$ and $H_k = k^{3/2}H_0$ (the reason for this choice will become clear later). Now, since the number of grid points for the hierarchy grows faster than the number of grid points for vintages, the maximal possible slope $\Delta h_k/\Delta\tau_k$ will grow to infinity, so any slope $\dot{h}$ will be covered from some k on, and all points in the upper-left corner of the rectangle will be reached by some mass from some k on. Of course, for each given grid size there still might be some cells in which the bound 1 is reached.

Consider now how the density of workers evolves on the grid:

$$n(\tau_{i+1}, h_j) = \big[1 - p(\tau_i, h_j)\big] n(\tau_i, h_j) + p(\tau_i, h_{j-1}) n(\tau_i, h_{j-1}).$$

Now, introduce the (upward-) difference operators $\Delta_h f(\tau_i, h_j) = f(\tau_i, h_{j+1}) - f(\tau_i, h_j)$ and $\Delta_\tau f(\tau_i, h_j) = f(\tau_{i+1}, h_j) - f(\tau_i, h_j)$ for arbitrary functions $f(\cdot, \cdot)$. Then we can re-write the above as

$$\begin{aligned} \Delta_\tau n(\tau, h) = &- \Delta_h \Big[n(\tau, h-1) p(\tau, h-1) \Big] = -n(\tau, h-1) \Delta_h p(\tau, h-1) \\ &- p(\tau, h-1) \Delta_h n(\tau, h-1) - \Delta_h n(\tau, h-1) \Delta_h p(\tau, h-1). \end{aligned}$$

Note that the last term on the right-hand side will become small compared to the others when the grid becomes very fine. In the limit, the equation becomes equivalent to the mass-transport PDE (6) that describes the behavior of the density n.

Production in a vintage $\tilde{Y}$ (excluding the TFP term $e^{-\gamma \tau_i}$) is calculated as

$$\tilde{Y}(\tau_i) = \Big[\sum \left(f(\tau_j) n(\tau_i, h_j) \right)^\rho \Delta h \Big]^{1/\rho},$$

where again the function f is evaluated in the middle of the corresponding box (τ_i, h_j). This expression converges to $\tilde{Y}(n(\tau_i, \cdot))$ (under mild conditions) for a given function $n(\cdot)$ as $\delta h \to 0$.

The discrete counterpart for wages is

$$w(\tau_i, h_j) = \exp[-\gamma \tau_i] f_j \left(\frac{\tilde{Y}(\tau_i)}{n(\tau_i, h_j)} \right)^{1-\rho}. \tag{24}$$

Note that this gives the wage rate *per unit of time*. If we want to calculate the counterpart to wage payments over time a worker spends inside a box (τ_i, h_j), of course we have to multiply this wage rate by $\Delta \tau$.

The discrete counterpart of the value function is

$$\begin{aligned} V(\tau_i, h_j) &= w(\tau_i, h_j) \Delta \tau + e^{-(\beta + \delta - \gamma) \Delta \tau} V(\tau_{i+1}, h_j) + \qquad (25) \\ &= \max_{\dot{h}} \Big\{ - \frac{c}{2} \dot{h}^2 \Delta \tau + \underbrace{\dot{h} \frac{\Delta \tau}{\Delta h}}_{=p} e^{-(\beta + \delta - \gamma) \Delta \tau} \Delta_h V(\tau_{i+1}, h_j) \Big\}. \end{aligned}$$

Since agents only move upward in equilibrium, we take the upward-difference to approximate the h-derive of V in the spirit of upwind-differencing.

Solving for the optimal policy gives us

$$\dot{h}^*(\tau_i, h_j) = \frac{e^{-(\beta + \delta - \gamma) \Delta \tau}}{c} \frac{\Delta_h V(\tau_{i+1}, h_j)}{\Delta h}, \tag{26}$$

which converges to the optimal policy from the agent's first-order condition in the continuous case. Plugging back in, we obtain the Bellman equation

$$V(\tau_i, h_j) = w(\tau_i, h_j)\Delta\tau + e^{-(\beta+\delta-\gamma)\Delta\tau} V(\tau_{i+1}, h_j) + e^{-2(\beta+\delta-\gamma)\Delta\tau} \frac{1}{c} \left(\frac{\Delta_h V(\tau_{i+1}, h_j)}{\Delta h} \right)^2 \Delta\tau.$$

When dividing this equation by $\Delta\tau$ and taking the limit as $\Delta\tau \to 0$, we obtain the Hamilton-Jacobi-Bellman equation (HJB) (2) for the continuous case.

I solve the system for a given rectangle with length T using an algorithm that is inspired by how a real economy might converge to a steady state under adaptive expectations, assuming some inertia in agents' actions. Given a distribution of agents n_k (where k indexes the iterations of the algorithm) over the grid, one can calculate the resulting wages from (24). Using the fact that the marginal value of skill is zero when the vintage dies (i.e. $\Delta_h V^{(k)}(\tau_{T+1}, h_j = 0$ for all j), we can back out the value function recursively going from τ_T back to τ_1 using (25), which also yields optimal policies $\dot{h}^{*(k)}$ from (26).

As for the promotion efforts $\dot{h}$, we now mix some of the optimal policies into the existing ones: $\dot{h}^{(k+1)} = \alpha \dot{h}^k + (1-\alpha)\dot{h}^{*(k)}$. As for the entry decisions, I send more mass into the starting points with higher value and less mass into those with higher value. Since wages are inversely related to the density, this algorithm drives the system towards an equilibrium if the tuning parameters are chosen right. Further work is required to prove that this algorithm is indeed a contraction.

To find T^*, the vintage horizon that is optimal from the planner's point of view, I vary T and find a density n_T by the algorithm above. I then choose T^* as the horizon T that maximizes the planner's criterion described in the beginning of section 2.8.

The algorithm performs well for a 30-by-10 grid on (τ, h)-space. Increasing the mesh size from this point on leads to almost identical numerical results but has a large computational cost.

The complete *Matlab* code used in the calibration and more detailed documentation are available from the author upon request.

C Data

This study uses the weakly anonymous IAB Employment Sample (years 1975-2001). Data access was provided via on-site use at the Research Data Center (FDZ) of the German Federal Employment Agency (BA) at the Institute for Employment Research (IAB) and remote data access, see Drews, Hamann, Köhler, Krug, Wübbeke & Autorengemeinschaft 'ITM-Benutzerhandbücher' (2006) for an excellent documentation of the IAB Employment Sample. The data set is a 2% random sample of all Germans covered by the mandatory public unemployment-insurance (UI) scheme. The sample does not include tenured public-sector employees and the self-employed; these groups are not overwhelmingly large so that the data set can be regarded as fairly representative of the German labor market.

Every individual holding a job that fell under the UI scheme for at least several weeks at any point of the period 1975-2001 was at the same 2% risk of being sampled. For every sampled individual, all available employment spells were collected and included in the data set.

Available worker characteristics are gender, age and a measure of education[32]. For each employment spell of the worker, we observe pre-tax earnings, a 3-digit occupation code and have an identifier of the employer's establishment. For establishments, we have a 3-digit industry classification and some information that was obtained from aggregates over the original set of administrative data before the 2%-sample was drawn. These data include 3-digit-level industry classification, number of employees in the establishment in the respective year, and the first and last date between 1975 and 2001 in which the establishment hired a worker subject to UI contributions. I take the first date that the establishment hired a worker as the establishment's founding date.

As is common in the literature, I restrict the sample to males who work full time and are between 20 and 61 years old. For consistency reasons, only workers born in former West Germany are considered. The following paragraphs provide a more detailed description of the data and the exclusion restrictions.

I only consider spells coming from the *BeH*, the database for work relationships – all spells stemming from *LeH*, the database for UI payments, are automatically excluded from the sample. Also, I consider only full-time employees (`stib`< 8). Furthermore, all spells that are marked as "geringfügige Beschäftigung" (tax-exempt part-time employment) are dropped. Note that apprentices and interns are *included* in the sample. This is a deliberate choice; since these employees constitute arguably a considerable fraction of the labor force that has no job-specific skills yet, it would not be desirable to discard this information in a study on human-capital accumulation.

Also, I only consider individuals whose first employment is with an establishment located in former West Germany. This is done since reunification in 1990 was a major disruption for the careers of East Germans; it is not clear how this event affected these workers' human-capital-accumulation choices. Furthermore, some quality checks are performed on the data: spells of individuals for whom more than one full-time job is declared are discarded. Also, spells with unreasonably low daily earnings are deleted (below 7 Euros in 2000 Euros per day).

Earnings are adjusted for inflation using the consumer price index for West Germany provided by the *Bundesbank* (the German Central Bank).

Stata programs and documentation on how the moments in section 3 were obtained in detail are available from the author upon request.

[32]This education measure is only filled for employment contracts where education information is necessary to determine UI contributions or benefits, so its information content is limited.