Pyramids, Diamonds, and Oscillations: AI and the Structure of Internal Labor Markets

Guido Friebel* Yao Huang† Jin Li‡ Soumitra Shukla§
Andrew Zhang¶

April 10, 2026

Abstract

AI is changing workplaces: it leads to higher individual productivity and faster learning. We model the effects of these changes on internal labor markets. We show that for productivity shocks, firms preserve the span in the long run: a pyramid remains a pyramid. In the short run, however, firms freeze junior hiring, temporarily transforming the pyramid toward a diamond. In the transition to the new steady state, firms may oscillate between pyramids and diamonds before settling. For learning shocks, the long-run span falls, potentially permanently shifting the firm from a pyramid to a diamond. These fluctuations create inequality between junior cohorts and affect firm value when human capital is firm-specific.

Keywords: Span of control, knowledge work, labor markets, firm-specific human capital

*Goethe University Frankfurt. Email: gfriebel@wiwi.uni-frankfurt.de
†The University of Hong Kong. Email: yaohuang@connect.hku.hk
‡The University of Hong Kong. Email: jli1@hku.hk
§Harvard Business School & The Burning Glass Institute. Email: sshukla@hbs.edu
¶Goethe University Frankfurt. Email: Y.Zhang@econ.uni-frankfurt.de

1 Introduction

On-the-job training and learning by doing are among the most important determinants of human capital accumulation. As Arrow (1962) argued, learning is the product of experience and can only take place through the attempt to solve problems. While schools and universities supply conceptual foundations, workers develop expertise by doing. The internal labor markets of firms (Doeringer and Piore, 1971; Baker, Gibbs and Holmstrom, 1994*a*) provide the structure in which this progression takes place: junior workers build human capital through the tasks they perform on the job, and they gradually advance to more important positions (Gibbons and Waldman, 1999*a*, 2004; Waldman, 2013).

As AI automates many of the tasks that junior workers have traditionally performed, such as collecting data and drafting routine documents, firms across industries are starting to hire fewer entry-level workers[1]. This trend suggests that the traditional pyramid is giving way to a diamond: fewer juniors at the base, a bulge of experienced workers in the middle. PwC's chief AI officer predicts that "human-AI collaboration could boost productivity and speed by 50 percent" and that "what emerges will be a shift of the traditional labor pyramid to a diamond shape" (Priest, 2025).

Yet, if the tasks that once served as the bottom rungs of the career ladder are disappearing, what happens to the pipeline that produces tomorrow's senior workers? Forward-looking executives have warned that cutting junior hiring is dangerously short-sighted. As AWS CEO Matt Garman put it, the idea of firing juniors because AI can do their jobs is "the dumbest thing I've ever heard. . . how's that going to work when ten years in the future you have no one that has learned anything" (Sharwood, 2025).

At the heart of the matter lies a single tension: junior positions serve a dual purpose. In the short run, juniors perform tasks and generate output. In the long run, they are the training ground through which the firm replenishes its senior workforce. A firm that cuts junior hiring captures immediate efficiency gains but may be undermining its own future.

The reduction in junior hiring thus raises a set of interconnected questions that go well beyond the immediate cost savings. What happens to the structure of the internal labor market: does the pyramid permanently become a diamond, or is this a transient phenomenon? What happens to the accumulation of human capital in the economy when

[1] Brynjolfsson, Chandar and Chen (2025) document a 16 percent decline in early-career hiring in AI-exposed fields, while demand for senior workers remains stable. Lichtinger and Hosseini Maasoum (2025), using résumé data covering 62 million U.S. workers across 285,000 firms, show that generative AI adoption sharply reduced junior employment. Similar patterns appear in freelance markets (Hui, Reshef and Zhou, 2024). Adding nuance to this pattern, Humlum and Vestergaard (2026) observe that while early-career employment declines in Denmark, actual adopting firms primarily integrate the technology through internal task reorganization rather than reducing headcounts.

fewer workers pass through the entry-level positions where skills are built? How are workers' careers and wages affected? And what happens to the firm's own profitability over time, especially when experienced workers carry knowledge that cannot easily be replaced from outside?

In this paper, we develop a model of internal labor markets to address these questions. A CEO allocates a fixed set of tasks to senior and junior workers, where seniors are more productive because they have accumulated human capital. Juniors learn on the job and, with some probability, acquire the skills needed to become seniors. The firm must therefore manage the intertemporal tradeoff between current production and the future supply of experienced workers. The setup is deliberately simple, but it captures the essential features of pyramidal organizations in which cohorts of entry-level workers sustain a talent pipeline into senior ranks. A key feature of the model is that the ratio of juniors to seniors is determined endogenously rather than fixed at one-to-one. This flexibility is what enables us to study the shape of the internal labor market and how that shape responds to AI.

We distinguish three channels through which AI may affect the workplace: it can raise the productivity of senior workers, raise the productivity of junior workers, or accelerate the rate at which juniors learn and qualify for senior positions. These are distinct shocks, and they lead to different answers. The paper has three main contributions.

First, if the main effect of AI is on productivity, whether of senior or junior workers, firms may change in size, but their internal structure does not change. Perhaps surprisingly, this holds even though the short-run disruption to junior hiring can be severe. In a nutshell, pyramids remain pyramids in the long run. The wage ratio between seniors and juniors does shift, however: when seniors become more productive, the wage gap widens; when juniors become more productive, it narrows.

Second, if the main effect of AI is on how fast juniors learn and can become seniors, a permanent structural change occurs. Pyramids become diamonds. This is the only type of AI shock that permanently reshapes the organizational hierarchy.

Third, we characterize the transition dynamics following both types of shock. When the shock is large enough, the transition is not smooth. Firms oscillate between pyramid and diamond shapes before settling into the new steady state: an initial hiring freeze starves the talent pipeline, triggering a future surge in junior hiring, which in turn produces a larger supply of seniors, and so on. These oscillations have consequences both for the firm and for workers. They temporarily erode the stock of firm-specific human capital and reduce firm value. They also create a 'lost cohort' of juniors who enter during a hiring freeze and face sharply reduced opportunities to work and accumulate skills, while the cohort that follows benefits from the resulting senior shortage.

We extend the model to allow for firm-specific human capital, where internally promoted seniors are more productive than external hires. In this setting, firms hire exclusively from within, which makes the talent pipeline even more consequential: when it breaks down, the firm cannot simply plug the gap from the outside market, and its value falls. An additional finding emerges for the wage structure. When AI accelerates learning, the wage ratio between seniors and juniors compresses. Faster learning raises the value of the junior layer—each junior now generates firm-specific capital more quickly—which narrows the premium that seniors command.

These results have implications for management and talent management more broadly (Friebel and Raith, 2026). While the forward-looking managers in our model contract the junior layer to capture immediate efficiency gains, doing so can disrupt the firm's internal supply of talent and reduce the firm's future value. Therefore, when the stability of future talent supply is important, a CEO may need to smooth these fluctuations by managing the talent pipeline.

It is useful to stress the policy implications for inequality between junior cohorts. For different cohorts of juniors, the employment and learning opportunities can vary significantly. Juniors entering during a period of excess seniors face reduced opportunities to be employed and acquire skills. This "lost cohort," however, may also create a shortage of seniors in the future. This shortage benefits the next cohort by increasing their opportunities to work and learn.

The ingredients of the model are inspired by a large literature on internal labor markets, a concept first introduced by Doeringer and Piore (1971). Their seminal work triggered a thriving theoretical and empirical literature, summarized by Gibbons (1997), Gibbons and Waldman (1999*a*), Lazear and Oyer (2013), and Waldman (2013). A number of empirical papers, most notably Baker, Gibbs and Holmstrom (1994*a*,*b*), provide detailed case studies of a firm indicating that internal labor markets were conducive to human capital acquisition. These papers and related theoretical contributions[2] also shed light on incentive issues that we currently abstract from. Pastorino (2024) investigates the importance of human capital accumulation and learning about employees' abilities within one firm. Recent contributions underline the importance of internal labor markets despite increasing competitive pressure (Huitfeldt et al., 2023; Osterman, 2024).

What this literature leaves open, however, is how the shape of the hierarchy itself responds to a technology shock. We take the pyramid structure not as a primitive but as an

[2] See Lazear and Rosen (1981); Rosen (1986); Malcomson (1984); MacLeod and Malcomson (1988); Prendergast (1993); Gibbons and Waldman (1999*b*); Zabojnik and Bernhardt (2001); Waldman (2003); Kräkel and Schöttner (2012); Auriol, Friebel and Von Bieberstein (2016); Ke, Li and Powell (2018); Bianchi et al. (2023); Friebel and Raith (2026).

equilibrium outcome, and ask how it changes when AI arrives. This allows for identifying new shapes of hierarchies that practitioners are discussing but economics so far has had little to say about.

How AI affects the development of human capital has recently been modeled by Garicano and Rayo (2025) and Ide (2026). Both papers adapt the apprenticeship idea first brought and modeled by Garicano and Rayo (2017) and Fudenberg and Rayo (2019), focusing on how AI changes the intensity of human capital accumulation at the individual level.

We share their interest in the interaction between AI and learning, but depart from them in a crucial respect. By bringing profit-maximizing firms explicitly into the picture, operating in a labor market, we can speak to questions those models cannot: not only how much each worker learns, but how many trained workers the economy produces over time. The firm's intertemporal hiring problem — hire juniors today who become seniors tomorrow — is central to our analysis and has no counterpart in those frameworks.

Our work is also related to the literature on vacancy chains and slot constraints in internal labor markets (Simon, 1951; White, 1970; Beckmann, 1978; Stewman and Konda, 1983; Rosenbaum, 1984; Ke, Li and Powell, 2018). These structural features create career spillovers, where the progression of junior workers relies on the departure of seniors, a phenomenon documented empirically by Friebel and Panova (2008) and Bianchi et al. (2023).

That literature establishes that cohort fates are linked through the pipeline. What it does not consider is how a technology shock propagates through that pipeline and generates oscillating cohort inequality of the kind we characterize here.

The production function of the main model is one in which CEOs need to get z tasks done by delegating to seniors and juniors depending on their productivity, which is enhanced by AI. We show that the main results are robust to richer production technologies. When seniors and juniors are complements rather than perfect substitutes — as in a Cobb-Douglas specification — the long-run span remains pinned by human capital flow rates alone, and the pipeline dynamics persist. The Cobb-Douglas case does, however, enrich the cohort inequality result: because the wage premium covaries with the span during the transition, cohorts face a trade-off between ease of entry and future reward, generating alternating cycles of scarce jobs with high future pay and abundant jobs with low future pay.

It should be noted though that the basic production function is not without charm because it is grounded in the emerging empirical literature that shows that more tasks can be done by AI in a given time and leads to faster learning. Generative AI significantly increases worker productivity, particularly for novice and lower-skilled individuals, allowing less experienced workers to rapidly close performance gaps with top performers (Noy and Zhang, 2023; Brynjolfsson, Li and Raymond, 2025; Chen et al., 2025; Humlum and Vestergaard, 2025). This internal augmentation may translate into external displacement; as AI automates entry-level tasks, firms may be freezing hiring for junior roles. Beyond labor demand, reliance on AI introduces risks to human capital formation and creative diversity. While AI tools can help workers move down the experience curve faster in supported environments (Brynjolfsson, Li and Raymond, 2025), unfettered access in educational settings can act as a "crutch," improving immediate performance but significantly degrading unassisted learning and exam scores once the tool is removed (Bastani et al., 2025).

2 Model Setup

In this section, we develop a dynamic general equilibrium model. The economy is populated by CEOs, junior workers, and senior workers. Time is discrete and indexed by $t = 0, 1, 2, \ldots$. All agents are risk-neutral and discount the future at $\delta \in (0, 1)$.

Firms and Production. There is a continuum of CEOs with total mass M. In every period t, each CEO has z tasks to complete. Each completed task yields a revenue of y. To execute these tasks, the CEO i organizes a firm by hiring two types of workers: juniors (l) and seniors (h). We henceforth use "CEO i" and "firm i" interchangeably. A junior worker can complete γ^l tasks per period, and a senior worker can complete γ^h tasks, where $0 < \gamma^l < \gamma^h$.

Let $n^l_{i,t}$ and $n^h_{i,t}$ denote the number of juniors and seniors employed by firm i in period t, and let $z^l_{i,t} \geq 0$ and $z^h_{i,t} \geq 0$ denote the tasks allocated to each group. The firm's production is constrained by the capacity of its workforce and the total number of tasks:

$$\begin{aligned} \gamma^l n^l_{i,t} &\geq z^l_{i,t}, \\ \gamma^h n^h_{i,t} &\geq z^h_{i,t}, \\ z^l_{i,t} + z^h_{i,t} &\leq z. \end{aligned} \tag{CC}$$

Workers and Human Capital Accumulation. There is a sufficiently large supply of potential junior workers with an outside option $\underline{v} > 0$. Once employed, junior workers accumulate human capital by doing tasks. Specifically, a junior who is allocated tasks learns the skills and becomes a senior in the subsequent period with an exogenous probability $q \in (0, 1]$. Let $\tilde{q}_{i,t} = q \frac{z^l_{i,t}}{\gamma^l n^l_{i,t}}$ denote the probability that this happens, which depends on the number of tasks allocated to juniors. At the end of each period, junior and senior workers exit the market at probabilities $d_l \in (0,1)$ and $d_h \in (0,1)$. Exiting workers receive their outside option $\underline{v}$.

Let $\{w_t^l, w_t^h\}_{t=0}^{\infty}$ denote the sequence of market wages. Given these wages, we can write the present discounted values of being a junior and a senior, v_t^l and v_t^h, at the beginning of period t as:

$$v_t^l = w_t^l + \delta(1-d_l)\left[\tilde{q}_{i,t} v_{t+1}^h + (1-\tilde{q}_{i,t}) v_{t+1}^l\right] + \delta d_l \underline{v}, \tag{VF^l}$$

$$v_t^h = w_t^h + \delta(1-d_h) v_{t+1}^h + \delta d_h \underline{v}. \tag{VF^h}$$

Timing. Figure 1 shows the timeline of the stage game. At the beginning of period t, firms decide the hiring or firing of workers. Then firms allocate tasks between senior and junior workers. After that, production happens and the output is realized. During the production process, junior workers acquire skills stochastically. Finally, some workers leave, and those remaining juniors who have acquired skills become seniors.

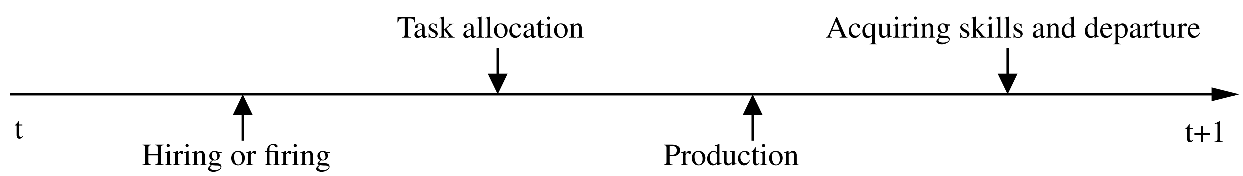


Figure 1: Stage Game Timeline

The Firm's Problem. Given wages $\{w_t^l, w_t^h\}_{t=0}^{\infty}$, firm i chooses the workforce size and task allocations to maximize its present discounted value:

$$\max_{\{n_{i,t}^l, n_{i,t}^h, z_{i,t}^l, z_{i,t}^h\}_{t=0}^{\infty}} \quad \sum_{t=0}^{\infty} \delta^t \left[(z_{i,t}^l + z_{i,t}^h)y - w_t^h n_{i,t}^h - w_t^l n_{i,t}^l\right],$$

subject to the capacity constraints (CC) for all t. We assume the revenue from a task is high enough ($\gamma^l y > (1-\delta)\underline{v}$) that the firm always finds it optimal to complete all z tasks.

Aggregation. The aggregate demand for juniors and seniors is $\int_i n_{i,t}^l \, di$ and $\int_i n_{i,t}^h \, di$. Let N_{t+1}^h denote the aggregate supply of senior workers in period $t+1$. The aggregate stock of seniors comes from two sources. The first is incumbent retention: the number of current seniors who stay, $(1-d_h)N_t^h$. The second is the talent pipeline: the number of current juniors who are allocated tasks, successfully acquire skills, and stay, $(1-d_l)\int_i \tilde{q}_{i,t} n_{i,t}^l \, di$.

Consequently, the aggregate supply of seniors follows:

$$N^h_{t+1} = (1 - d_h)N^h_t + (1 - d_l)\int_i \tilde{q}_{i,t} n^l_{i,t}\, di. \tag{FC}$$

Equilibrium. A dynamic general equilibrium consists of sequences of wages $\{w^l_t, w^h_t\}^\infty_{t=0}$, firm policies $\{n^l_{i,t}, n^h_{i,t}, z^l_{i,t}, z^h_{i,t}\}^\infty_{t=0}$, worker values $\{v^l_t, v^h_t\}^\infty_{t=0}$, and aggregate senior labor supply $\{N^h_t\}^\infty_{t=0}$ such that: given wages, the firm policies solve the firm's optimization problem; given wages and firm policies, the worker values satisfy the Bellman equations and participation conditions; the market for senior workers clears, meaning that aggregate demand equals aggregate supply; and the aggregate supply of senior workers evolves according to the flow equation (FC).

2.1 Remarks on the Model

Our model is deliberately simple to allow for a full characterization of the wage and employment dynamics of junior and senior workers. Below we discuss briefly how our results change when additional elements are incorporated into the model.

Homogeneous Tasks and the Role of Assignment. In our model, senior and junior workers perform the same type of task, differing only in their per-worker capacity. An alternative formulation, in the spirit of (Gibbons and Waldman, 1999*b*), would introduce a complex task and a routine task, and the firm assigns workers to tasks based on their human capital. In that setting, the firm would naturally assign senior workers to the complex task and junior workers to the routine one, and the task assignment serves as a form of promotion. Our results carry over to this richer environment, but the single-task formulation keeps the notation lighter and lets us focus on the forces that shape the internal labor market without the additional notation required by an assignment model.

Production Technology. The fixed-task production function, where the firm's output is limited by the number of available tasks, is a tractable special case. We later embed our framework in a more general production technology, including, for example, a Cobb-Douglas specification in which senior and junior labor are combined as inputs. Under such a specification, output responds smoothly to the workforce mix. Our main results on the long-run structure of the internal labor market and the transition dynamics following AI shocks are preserved under these more general technologies. The richer production function does, however, introduce an additional channel through which AI affects the wage structure: because senior and junior inputs are no longer perfect substitutes, a change in senior productivity alters the marginal product of juniors.

Firm-Specific Human Capital. The baseline model assumes that senior workers are equally productive regardless of whether they were promoted internally or hired from the external market. We later introduce firm-specific human capital, and we show that the main results on firm structure and the transition dynamics remain qualitatively unchanged. The extended model does, however, highlight an additional consequence of the hiring freeze that AI shocks can trigger: because fewer juniors enter the pipeline, the future supply of high-productivity internal seniors falls. This erosion of the internal talent stock reduces the firm's profitability in the future, a cost that is absent in the baseline specification.

Firing Costs and Incentives. Two additional forces that feature prominently in the internal labor market literature, firing costs and promotion incentives, are absent in our model. Adding either or both does not alter the characterization of the steady state for some parameter ranges, because the long-run structure depends only on flow rates and productivity parameters. The transition dynamics, however, become considerably harder to characterize. We view these as important extensions, but our central observation that today's junior hire affects future stock of seniors operates independently of them.

3 Preliminary Analysis

To begin the analysis, we first investigate the optimal task allocation and hiring policy for each firm. Because the firm can adjust the workforce composition costlessly, the firm's decisions across periods are de facto independent in the sense that the hiring choice in period t does not affect the value of the firm in period $t+1$. Therefore, given the wages in each period, the firm's optimization problem can be reduced to a sequence of static problems:

$$\max_{n^l_{i,t},n^h_{i,t},z^l_{i,t},z^h_{i,t}} \quad (z^l_{i,t}+z^h_{i,t})y - w^h_t n^h_{i,t} - w^l_t n^l_{i,t},$$

subject to the capacity constraints (CC).

This optimization problem results in the following task allocation and hiring policy.

Lemma 1. *For each firm i, the following holds in every period t: (i) $z^l_{i,t}+z^h_{i,t}=z$; (ii) $z^l_{i,t}=\gamma^l n^l_{i,t}$; (iii) $z^h_{i,t}=\gamma^h n^h_{i,t}$.*

Lemma 1 shows that all the capacity constraints (CC) bind in a competitive market. This follows because there is no gain for the firm in having slack in production (in the sense that workers have not worked to their full capacity). Reducing slack saves the wage costs for the firm.

Next, we describe the wage ratio between senior and junior workers in this economy.

Lemma 2. *If firms employ strictly positive quantities of both senior and junior workers ($n^h_{i,t} > 0$ and $n^l_{i,t} > 0$), the ratio of wages must equal the ratio of productivities:*

$$\frac{w^h_t}{w^l_t} = \frac{\gamma^h}{\gamma^l}.$$

Lemma 2 describes a no-arbitrage condition. The firm's production technology is linear, making junior and senior workers perfect substitutes. Because the firm can substitute γ^h/γ^l junior workers for one senior worker, the market wage premium for seniors must exactly match their productivity premium. Note that given the wage ratio, each firm is indifferent between hiring a junior and a senior worker.

4 Long-Run Outcome

We now characterize the long-run behavior of the economy. We begin by defining the steady state, and then establish its existence, uniqueness, and convergence properties. For our analysis, we assume that the rate of human capital accumulation is not too fast. That is, we assume the learning rate (q) satisfies $\frac{\gamma^h}{\gamma^l} q < \frac{2-d_h}{1-d_l}$. We discuss the case when this condition is not satisfied in Section 6.4.

Definition 1 (Steady State)**.** A steady state of the economy is a configuration $(N^{h*}, N^{l*}, w^{h*}, w^{l*})$ of aggregate senior employment, aggregate junior employment, and wages such that aggregate employment and wages are constant over time:

$$N^h_{t+1} = N^h_t = N^{h*}, \quad N^l_{t+1} = N^l_t = N^{l*}, w^h_{t+1} = w^h_t = w^{h*}, w^l_{t+1} = w^l_t = w^{l*}$$

and the equilibrium conditions are satisfied.

To describe the steady state, it is useful to separate the hierarchy's long-run *shape* from its long-run *scale*. We use $S^* := N^{l*}/N^{h*}$ to denote the *span*. In a representative firm, i.e., when all firms are identical, S^* determines the shape of the hierarchy. A firm's shape is a pyramid if $S^* > 1$ and is a diamond if $S^* < 1$. We use N^{h*} to denote the size of the senior layer; given S^*, the size of the junior layer is $N^{l*} = S^* N^{h*}$. This distinction between the shape and scale is useful because, in our model, shape is pinned down by worker flows, whereas scale is pinned down by productivity and the total task load. Also, let

$Z \equiv Mz$ denote the aggregate number of tasks in the economy. The following proposition characterizes the long-run outcome of the economy.

Proposition 1 (Long-Run Outcome)**.** *For any initial stock of seniors, the economy converges to a unique steady state in the long run. The long-run steady outcome is characterized as follows:*

(i) *The aggregate junior-to-senior ratio and aggregate employment levels are:*

$$S^* = \frac{d_h}{q(1-d_l)}, \quad N^{l*} = \frac{S^*}{\gamma^l S^* + \gamma^h} Z \quad \textit{and} \quad N^{h*} = \frac{1}{\gamma^l S^* + \gamma^h} Z.$$

(ii) *The equilibrium wages are:*

$$w^{l*} = \underline{v}(1-\delta)\left[\frac{1-\delta(1-d_h)+\delta q(1-d_l)}{1-\delta(1-d_h)+\frac{\gamma^h}{\gamma^l}\delta q(1-d_l)}\right] \quad \textit{and} \quad w^{h*} = \frac{\gamma^h}{\gamma^l} w^{l*}.$$

Proposition 1 shows that the economy converges to a unique steady state. In the long run, part (i) of Proposition 1 characterizes the shape and scale of a representative firm. A technical remark is that because juniors and seniors are perfect substitutes and the labor market is frictionless, individual firms are indifferent to their specific workforce composition. At the extreme, the aggregate steady state could theoretically be sustained by a mix of firms that employ only seniors and firms that employ only juniors. However, if internal seniors possess firm-specific human capital that external hires lack, this structural indifference disappears. As we show in Section 6.2, this friction forces the internal structure of every firm to mirror the economy-wide aggregate. We therefore assume that all firms are identical and focus on the representative firm.

In the steady state, the span S^* of a representative firm is determined by a simple flow-balance condition: the inflow into the senior layer, $q(1-d_l)N^{l*}$, must equal the outflow from the senior layer, $d_h N^{h*}$. This immediately implies $S^* = d_h/[q(1-d_l)]$. A key implication is that productivity does not enter this expression. A change in γ^h or γ^l changes how many workers are needed to complete current tasks, but it does not change how many juniors are needed per senior to keep the top layer constant. In that sense, productivity changes scale, not shape. This independence of S^* from productivity is a general long-run result. Because it comes from the stock-flow condition for the senior layer, it holds in more general production environments as well.

The expression for S^* shows when the hierarchy is more pyramid-like or more diamond-like. A higher senior departure rate d_h increases the need to replenish the top layer and therefore requires more juniors per senior. A higher junior departure rate d_l weakens the talent pipeline by reducing the fraction of juniors who remain long enough to be promoted. A lower learning rate q weakens the pipeline for the same reason. All three forces increase S^* and make the hierarchy more pyramidal. Conversely, when seniors stay longer, juniors stay longer, or juniors learn faster, the pipeline becomes more efficient and the hierarchy becomes more diamond-like.

We treat d_h and d_l as exogenous reduced-form turnover parameters. In a richer model, AI could affect them by changing outside opportunities and therefore worker exits. If so, AI would reshape the hierarchy through an additional turnover channel. But if departures mainly reflect retirement, family or home production, or other outside opportunities that are not themselves affected by productivity shock, holding d_h and d_l fixed is a reasonable first approximation that isolates the inflow-outflow mechanism.

Once the flow rates determine the span, the number of tasks (Z) and workers' productivity $(\gamma^l$ and $\gamma^h)$ determine the size of each layer. The firm adjusts the structure until the combined output of the two layers meets Z. Because the firm only needs to complete Z tasks, higher productivity in either the senior or junior worker reduces the number of workers in both levels.

Part (ii) of Proposition 1 describes the wage of the workers. Note that the wage of a worker is not equal to his marginal productivity. Instead, the wages are pinned down by the worker's outside option and reflect workers' present value. Juniors accept lower current wages because a junior job carries a promotion option. The wage profile is therefore backloaded: part of the compensation for juniors is paid not today, but in the form of a chance to access the senior wage tomorrow.

The junior wage w^l equals the flow outside option, $(1-\delta)\underline{v}$, multiplied by a factor strictly less than one. This is because sufficient junior labor supply drives the value of a junior position to $\underline{v}$, and workers accept lower current pay in exchange for the continuation value of acquiring skills and earning the senior wage w^h. The senior wage is then determined by the no-arbitrage condition in Lemma 2 $(w^h/w^l = \gamma^h/\gamma^l)$. This forward-looking logic is crucial for understanding the effect of AI.

Next, we turn to AI's effect on the economy. We distinguish three channels: AI may raise senior productivity γ^h, raise junior productivity γ^l, or raise the learning rate q. The first two are productivity shocks. The third is a learning shock. Proposition 2 shows that they have sharply different implications for organizational shape, employment, and wages.

Proposition 2 (The Long-Run Effect of AI). *In steady state, the economy responds to AI shocks as follows:*

(*i*) *If senior productivity γ^h increases, employment drops for both seniors and juniors (N^{h*} and*

N^{l*}), *but the aggregate junior-to-senior ratio* (S^*) *remains constant. The junior wage* (w^{l*}) *decreases, while the senior wage* (w^{h*}) *increases.*

(ii) *If junior productivity* γ^l *increases, employment drops for both seniors and juniors* (N^{h*} *and* N^{l*}), *but the aggregate junior-to-senior ratio* (S^*) *remains constant. The junior wage* (w^{l*}) *increases, while the senior wage* (w^{h*}) *decreases.*

(iii) *If the learning probability* q *increases, senior employment* (N^{h*}) *increases, junior employment* (N^{l*}) *decreases, and the aggregate junior-to-senior ratio* (S^*) *decreases. Both the junior wage* (w^l) *and the senior wage* (w^{h*}) *decrease, but the ratio* ($\frac{w^{h*}}{w^{l*}}$) *stays constant.*

Proposition 2 shows how different AI shocks affect the scale of employment, the shape of the hierarchy, and the wage ladder. Productivity shocks change scale and the wage ladder but leave shape unchanged. Learning shocks change all three.

Consider first an increase in senior productivity γ^h. Both senior and junior employment fall. The reason is not only that each senior can now perform more tasks. Once the span S^* is fixed by flows, a smaller senior layer mechanically requires a smaller junior layer as well. Thus a productivity improvement at the top contracts the whole hierarchy while preserving its shape. On the wage side, higher γ^h widens the productivity gap between seniors and juniors and raises the senior wage. Because the senior position becomes a more valuable career destination, firms can offer lower junior wages today. In equilibrium, juniors are paid their reservation lifetime value, not their current contribution alone, so firms backload compensation more aggressively when the future senior prize improves.

Next, consider an increase in junior productivity γ^l. Again, both layers shrink, because the firm needs fewer workers overall and the flow-determined span remains unchanged. But the wage effects reverse. A more productive junior worker becomes a closer substitute for a senior worker, so the scarcity value of the senior position falls. Equivalently, promotion becomes a less valuable prize in lifetime terms. That is why the senior wage falls when junior productivity rises. At the same time, juniors become more valuable in current production, and because the promotion prize is now smaller, firms must compensate them more upfront. As a result, the junior wage rises while the senior wage falls.

Finally, when AI raises the learning rate q, the key effect works through flows rather than current task capacity. Faster learning increases the expected inflow into the senior layer, so the firm needs fewer juniors to sustain a given senior workforce. The span therefore falls, making the hierarchy more diamond-like. This is the only AI shock in the baseline model that permanently reshapes organizational form in the long run. Wages also fall. A higher q raises the continuation value of a junior job, which lets firms lower junior wages today. Because the wage ratio is pinned down by the no-arbitrage condition $w^h/w^l = \gamma^h/\gamma^l$, senior wages fall in proportion.

It is useful to contrast our model with a benchmark in which the two layers operate independently. In such a benchmark, where two separate spot labor markets, a productivity shock to one layer would be entirely local: an increase in γ^h would reduce senior employment and raise the senior wage, with no effect whatsoever on juniors, and symmetrically for γ^l. Learning shocks cannot be easily analyzed because there is no pipeline connecting the layers.

Our model departs from this benchmark through two distinct linkages. The first is an *employment linkage* that operates through the flow equation. Because juniors today are the source of seniors tomorrow, the two layers' sizes are tied: a change in one layer's employment propagates to the other through the requirement that inflow match outflow in steady state. This is why a productivity improvement at the top, which directly reduces only the number of seniors needed, also reduces junior employment. Specifically, the span S^* is fixed by flows, so the junior layer must shrink in proportion in the long run.

The second is a *wage linkage* that arises because workers evaluate jobs dynamically. Because juniors are forward-looking, the junior wage today depends on the senior wage tomorrow. A change in γ^h does not affect the junior worker's productivity. But it changes the value of the senior position, which changes how much career surplus a junior job offers, which in turn changes the wage at which juniors are willing to accept.

Together, the two linkages generate a pattern that is common to all three shocks: junior employment falls. Whether AI makes seniors more productive, juniors more productive, or learning faster, the bottom of the hierarchy contracts. The mechanisms differ, but the implication for entry-level hiring is the same. This is consistent with the emerging empirical evidence that AI adoption reduces demand for junior workers across a range of settings.

Where the shocks differ is in their effect on organizational shape and on the wage ladder. On shape, the distinction is clean: productivity shocks leave the span unchanged, whereas the learning shock is the only one that permanently reshapes the hierarchy, shifting it from a pyramid toward a diamond.

On wages, all three shocks operate through the same channel of the option value of promotion, but they affect the option value differently. Productivity shocks change the size of the career prize. When γ^h rises, the gap between the senior and junior wage widens, making the prize larger; firms exploit this by offering lower junior wages today, so wages move in opposite directions across the two layers. When γ^l rises, the gap narrows, making the prize smaller; firms must compensate juniors more upfront, and again wages move in opposite directions but now favoring juniors.

The learning shock, by contrast, changes the speed at which the prize is reached rather than its size. A higher q means juniors access the senior wage sooner, raising the option value of a junior position even though the prize itself ($w^h/w^l = \gamma^h/\gamma^l$) is unchanged. Firms can therefore offer lower junior wages, and the increased flow of newly promoted seniors simultaneously depresses the senior wage. Both wages fall.

5 AI Shocks and Transition Dynamics

The previous section shows how AI shocks change the firm's structure and the wages of the workers in the long run. We now study the transition to the steady state.

Before turning to the formal analysis, we highlight a general feature of the transition in the baseline model. As established earlier, wages are determined by the no-arbitrage condition ($w^h/w^l = \gamma^h/\gamma^l$) and the free-entry condition for juniors ($v^l = \underline{v}/(1-\delta)$). Neither condition depends on the current stock of senior or junior workers. Consequently, once the AI shock hits and the productivity or learning parameters, wages jump immediately to their new steady-state levels and remain there throughout the transition. All of the transitional dynamics described below play out in the number of seniors and juniors employed rather than in wages.

This separation is, admittedly, a special feature of the baseline model: because the seniors and juniors are perfect substitutes, wages are pinned entirely by parameters and do not depend on the state of the workforce. An advantage of this structure is that it yields closed-form wage expressions and makes the source of the transitional dynamics transparent: they come entirely from the pipeline, not from wage adjustments. In the extension with more general production technology, both wages and the firm's structure change every period.

5.1 Productivity Shock

Suppose the representative firm begins in a steady state with N_0^{h*} seniors and N_0^{l*} juniors. At the beginning of period $t = 1$, AI increases the productivity of seniors from γ_0^h to γ_{AI}^h and juniors from γ_0^l to γ_{AI}^l. The following proposition characterizes the transition.

Proposition 3. *Assume the AI shock is not so extreme that the firm still hires juniors at $t = 1$ ($\gamma_{AI}^h < Z/N_0^{h*}$). Following an increase in productivity from (γ_0^h, γ_0^l) to $(\gamma_{AI}^h, \gamma_{AI}^l)$ at $t = 1$:*

(i) At $t = 1$, the senior number remains at N_0^{h*}, the junior number drops to $N_1^l < N_0^{l*}$, and the span decreases.

(*ii*) *For $t \geq 2$, the firm structure converges toward the new steady state according to the following cases:*

(*a*) *If $\frac{\gamma_{AI}^h}{\gamma_{AI}^l} > \frac{1-d_h}{q(1-d_l)}$, the transition oscillates: the senior number N_t^h alternates above and below its new steady state, and the junior number N_t^l moves in the opposite direction. The span fluctuates.*

(*b*) *If $\frac{\gamma_{AI}^h}{\gamma_{AI}^l} \leq \frac{1-d_h}{q(1-d_l)}$, the transition is monotonic: the senior number N_t^h decreases and the junior number N_t^l increases. The span increases steadily.*

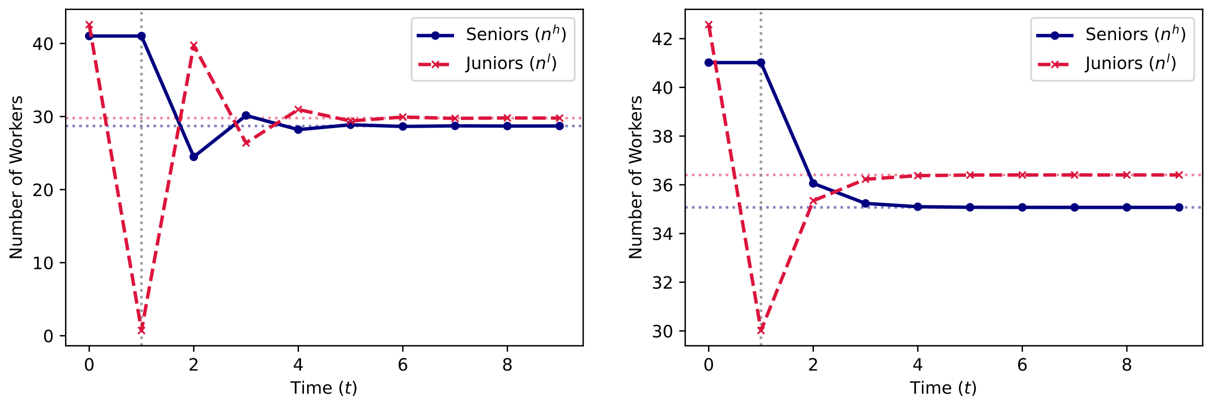
Proposition 3 shows that although productivity shocks eventually return the firm to its original span and the firm's shape remains a pyramid, the span adjusts in every period in the transition.

Specifically, the junior layer shrinks at $t = 1$ (part i). When the shock hits, the firm inherits a fixed stock of senior workers from the previous period. Because all workers are now more productive, the firm needs fewer workers in total to produce its target of Z tasks. Because the inherited seniors complete more tasks than before, there is less work for the junior layer. Furthermore, each junior can now handle more tasks. Facing this excess capacity, the firm hires fewer junior workers than before and may even fire some of the existing ones. The span therefore drops.

This shock to junior hiring affects the subsequent transition path. Because the firm relies on its junior layer to supply future seniors, cutting juniors today reduces the talent pipeline for tomorrow. How severely this pipeline is affected depends on the new productivity gap between seniors and juniors ($\gamma_{AI}^h/\gamma_{AI}^l$).

If the new productivity gap is large (Case (a) of Part ii), the initial cut to junior hiring is severe. This is because when the productivity gap is wide, one senior does the work of many juniors. As the excess seniors complete most of the tasks, the firm massively cuts its junior workforce, creating oscillations. By period $t = 2$, the number of seniors falls below the new steady-state level, creating a temporary shortage. To replace the missing seniors, the firm must hire a massive cohort of juniors. Because of this massive hiring, the span overshoots its pre-shock level, making the firm temporarily even more pyramidal than it was before the AI shock. This large cohort subsequently grows into an oversupply of seniors, causing another halt in junior hiring. This cycle of shortage and excess goes on and on, causing the oscillation in the firm's span. That is, the firm can swing between pyramids and diamonds before finally converging to the long-run steady state.

Conversely, if the new productivity gap is small (Case (b)), the initial excess of seniors causes only a mild reduction in junior hiring. Because this cut is small, the retained seniors and ongoing promotions are sufficient to prevent a severe pipeline break. The senior stock



(**a**) Oscillating: $\gamma_{AI}^h = 2.42$, $\gamma_{AI}^l = 1.02$.

(**b**) Smooth: $\gamma_{AI}^h = 1.45$, $\gamma_{AI}^l = 1.35$.

Figure 2: Simulation of the Transition where AI Increases γ^h and γ^l. Parameters: $d_l = 0.5$, $d_h = 0.41$, $q = 0.79$, $\gamma_0^h = 1.4$, $\gamma_0^l = 1$, and $Z = 100$.

then declines smoothly toward its new steady state, junior employment rises back from its initial drop, and the span monotonically returns to its long-run level.

5.2 Learning Shock

We now consider how the firm adjusts to a learning shock. Suppose the firm begins in a steady state with N_0^{h*} seniors and N_0^{l*} juniors. At the beginning of period $t = 1$, AI increases the learning rate of junior workers from q_0 to q_{AI}. The following proposition characterizes the transition dynamics.

Proposition 4. *Following an increase in the learning rate from q_0 to q_{AI} at $t = 1$:*

(*i*) *At $t = 1$, the senior and junior numbers remain at N_0^{h*} and N_0^{l*}.*

(*ii*) *For $t \geq 2$, the firm structure converges toward the new steady state according to the following cases:*

  (*a*) *If $q_{AI} > \frac{1-d_h}{1-d_l}\frac{\gamma^l}{\gamma^h}$, the transition oscillates. Following an initial increase in seniors and decrease in juniors (hence a decrease in span) at $t = 2$, the senior number alternates above and below the new steady-state level, and the junior number moves in the opposite direction. The span fluctuates.*

  (*b*) *If $q_{AI} \leq \frac{1-d_h}{1-d_l}\frac{\gamma^l}{\gamma^h}$, the transition is monotonic. The senior number increases, the junior number decreases, and the span steadily decreases.*

Proposition 4 shows that the learning shock results in a different transition dynamics from the productivity shock because it has a delayed impact on employment.

At $t = 1$, the firm structure remains entirely unchanged under the learning shock (part i). This stands in contrast to the productivity shock, which triggers an immediate contraction of the junior layer. The reason is that the learning shock operates through the pipeline rather than through current task capacity. At $t = 1$, juniors are acquiring the human capital at a faster rate, but the flow of new seniors has not yet materialized.

Wages, however, do change at $t = 1$. Even though no employment adjustment has occurred, the junior wage falls immediately because the option value of a junior position has increased: faster learning means quicker access to the senior wage. The senior wage falls in proportion, tied by the no-arbitrage condition. This is the wage linkage discussed earlier: wages respond to the change in career dynamics one period before the employment linkage begins to operate.

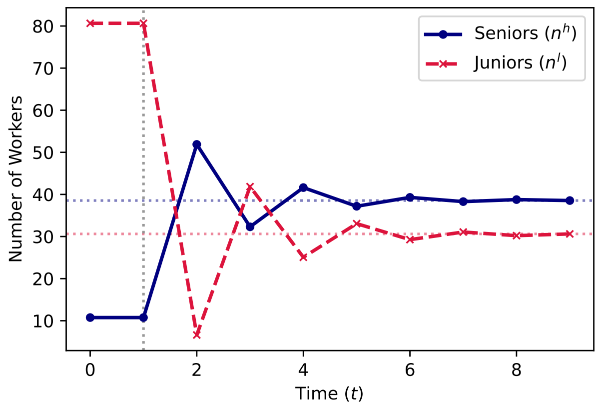
The effect on employment appears at $t = 2$ (part ii). Because juniors now learn faster, a larger fraction of them acquires skills and becomes the new seniors at the start of $t = 2$.

If the new learning rate is large (case a), this wave of new seniors at $t = 2$ is massive, pushing the senior number far above its new steady-state level. Facing this excess of seniors, the firm drastically cuts junior hiring. This massive cut starves the future talent pipeline. By period $t = 3$, the lack of juniors from the previous period creates a temporary shortage of seniors. To replace them, the firm must hire a large cohort of juniors. Because learning is fast, this large cohort matures into another excess of seniors at $t = 4$, and the firm cuts junior hiring again. The firm's span oscillates above and below the new steady-state level during transition. Consequently, the firm's structure can repeatedly oscillate between pyramids and diamonds before finally settling.

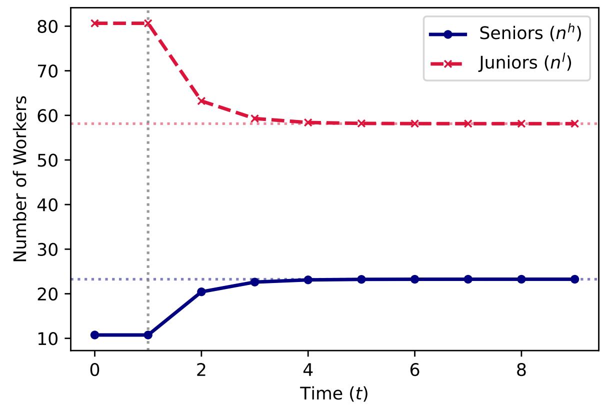
Conversely, if the new learning rate is relatively small (case b), the transition is smooth. Because the learning rate is modest, the wave of new seniors at $t = 2$ is smaller. The senior number grows, but it remains below its new steady-state level. Correspondingly, the firm reduces junior hiring mildly. Over time, the firm smoothly accumulates seniors and sheds juniors, monotonically decreasing its span until it reaches the new steady state.

6 Extensions

The baseline model isolates the talent pipeline mechanism in a simple environment: linear production, a fixed number of firms, and perfectly substitutable seniors. In this section, we relax these assumptions. First, we introduce general production technologies that allow seniors and juniors to be complementary. Second, we analyze the case where internally promoted seniors possess firm-specific human capital that external hires lack. Third, we allow for free entry of firms and analyze how industry scale adjusts to AI shocks.



(**a**) Oscillating: $q_{AI} = 0.95$.



(**b**) Smooth: $q_{AI} = 0.3$.

Figure 3: Transition Dynamics with Positive Learning Shock. Parameters: $d_l = 0.4$, $d_h = 0.45$, $q_0 = 0.1$, $\gamma^h = 1.8$, $\gamma^l = 1$, and $Z = 100$.

Then, we analyze rapid learning where the stability condition does not hold. Lastly, we map our baseline model to a job assignment framework. Together, these extensions preserve our primary findings while revealing new long-run and transitional implications for wages, employment, and firm value.

6.1 General Production Technology

The baseline model uses a linear production technology. This linearity makes senior and junior workers perfect substitutes. In many settings, however, these workers can be complements. Seniors might design a system's architecture, while juniors write the routine code. To capture this complementarity, we replace the linear production function with a more general specification, assuming long-term convergence.

Formally, in period t, firm i employs $n^h_{i,t}$ seniors and $n^l_{i,t}$ juniors to maximize $zy - w^h_t n^h_{i,t} - w^l_t n^l_{i,t}$, subject to the capacity constraint:

$$f(\gamma^h n^h_{i,t}, \gamma^l n^l_{i,t}) \geq z,$$

where f is increasing, concave, continuously differentiable, and satisfies constant returns to scale. We omit task allocation here because firms solve static problems in each period and therefore do not retain idle workers.

Under this general specification, the steady-state span remains $s^* = d_h/[q(1-d_l)]$, pinned down entirely by human capital flow rates. The effect of productivity shocks on wages, however, depends on the elasticity of substitution between the two inputs: when seniors and juniors are gross substitutes, a productivity shock to one layer devalues the other, as in the baseline; when they are gross complements, a productivity shock to one layer benefits the other. For clean closed-form expressions, we proceed with the Cobb-Douglas specification $f(\gamma^h n^h, \gamma^l n^l) = (\gamma^h n^h)^\alpha (\gamma^l n^l)^{1-\alpha}$, where $\alpha \in (0,1)$ is the output elasticity of senior labor.

Lemma 3. *In period t, the wage ratio is given by:*

$$\frac{w_t^h}{w_t^l} = \frac{\alpha}{1-\alpha} s_t.$$

Lemma 3 reveals two differences from the baseline. First, the wage ratio now depends on the span s_t rather than being fixed. When the span is wide—many juniors per senior—each senior is relatively scarce, and the wage premium rises. Conversely, a narrow span compresses the premium. Second, the productivity γ^h and γ^l do not appear in the wage ratio. Under Cobb-Douglas, a change in either productivity shifts the marginal products of both inputs in the same proportion, leaving their ratio unchanged. This contrasts with the linear case, where the wage ratio moves one-for-one with the productivity ratio.

The following lemma characterizes the steady-state equilibrium.

Lemma 4. *Under the Cobb-Douglas production technology, in steady state equilibrium, the span s^*, senior number n^{h*}, and junior number n^{l*} are:*

$$s^* = \frac{d_h}{q(1-d_l)}, \quad n^{h*} = \frac{z}{(\gamma^h)^\alpha (\gamma^l)^{1-\alpha} (s^*)^{1-\alpha}}, \quad n^{l*} = s^* n^{h*}.$$

The wages and wage ratio are:

$$w^{l*} = \underline{v}(1-\delta)\frac{1-\delta(1-d_h)+\delta q(1-d_l)}{1-\delta(1-d_h)+\frac{\alpha}{1-\alpha}\delta d_h}, \quad w^{h*} = \frac{\alpha}{1-\alpha}\left[\frac{d_h}{q(1-d_l)}\right] w^{l*}.$$

The steady-state span is identical to the baseline: $s^* = d_h/[q(1-d_l)]$. This shows that the long-run shape of the hierarchy is pinned down by human capital flow rates, independent of whether the production technology treats seniors and juniors as substitutes or complements. Production technology affects scale, how many workers of each type are needed, but not shape.

The following proposition shows how the steady state responds to AI shocks.

Proposition 5. *In the steady state equilibrium under Cobb-Douglas technology, the economy responds to AI shocks as follows:*

(i) *If productivity (γ^h or γ^l) increases, employment drops for both seniors and juniors (n^{h*} and n^{l*}), but the span (s^*) remains constant. The absolute wages (w^{h*} and w^{l*}) and the wage ratio (w^{h*}/w^{l*}) remain constant.*

(ii) *If the learning probability q increases, senior employment (n^{h*}) increases, junior employment (n^{l*}) decreases, and the span (s^*) decreases. The wage ratio (w^{h*}/w^{l*}) decreases. The junior wage (w^{l*}) increases, while the senior wage (w^{h*}) decreases.*

For a productivity shock, the firm shrinks in size but maintains its shape, just as in the linear model. But the wage implications can be richer for general production. In the linear model, raising γ^h widens the productivity gap and therefore widens the wage gap; raising γ^l narrows it. Under Cobb-Douglas, neither shock moves the wage ratio or the absolute wage levels at all. The reason is complementarity: when senior productivity rises, each senior produces more, but this also raises the marginal product of juniors who work alongside them. The gains are shared proportionally across both layers, leaving relative and absolute wages unchanged. Symmetrically, an increase in junior productivity raises the marginal product of seniors by the same proportion.[3]

For a learning shock, the structural response aligns with the baseline: faster learning produces more seniors, reduces the span, and shifts the hierarchy toward a diamond. The wage response, however, introduces a new channel that is absent in the linear model. In the baseline, the wage ratio $w^h/w^l = \gamma^h/\gamma^l$ is invariant to q. Under Cobb-Douglas, a smaller span means that seniors are less scarce relative to juniors, which compresses the wage ratio through Lemma 3. This compression reduces the future value of becoming a senior, so the firm must pay juniors a higher wage. As a result, the junior wage rises while the senior wage falls.

We now turn to the transition dynamics following an AI shock. We restrict attention to the stable case where the economy converges to the new steady state.

Proposition 6. *Following an increase in productivity or learning rate at $t = 1$, the firm's structure shows the following:*

(i) *At $t = 1$, the senior number remains at n_0^{h*}, and the junior number remains at n_0^{l*} if it is a purely learning shock, and strictly drops if it is a productivity shock;*

[3] The Cobb-Douglas specification sits at a boundary between two cases. When seniors and juniors are close substitutes, a productivity shock to one layer devalues the other, generating the opposing wage movements seen in the baseline. When they are strong complements, a productivity shock to one layer benefits the other even more than itself, reversing the wage effects. Cobb-Douglas, with its unit elasticity of substitution, is the boundary case where these forces exactly offset, leaving wages invariant to productivity changes.

(*ii*) *At $t \geq 2$, the firm structure converges toward the new steady state that can be either smooth or oscillating.*

During the transition, the wage ratio w_t^h/w_t^l comoves with the span s_t: it rises when the span widens and falls when the span narrows. In the long run, the wage ratio converges back to the pre-shock level following a productivity shock, and to a permanently lower level following a learning shock.

The transition paths for the firm's structure preserve the baseline's key features. A productivity shock creates excess senior capacity, squeezes out juniors, and can trigger the same oscillating pipeline dynamics described in Propositions 3 and 4. A learning shock leaves the firm unchanged at $t = 1$ but generates a wave of newly promoted seniors at $t = 2$, which can similarly set off oscillations when the output elasticity of senior labor α is high, where losing one senior requires hiring many more juniors to maintain production.

The novel feature under Cobb-Douglas is that wages are no longer constant during the transition. Because the wage ratio is tied to the span through Lemma 3, it moves with the firm's workforce composition in every period. When the firm temporarily carries excess seniors (a narrow span), the senior premium compresses. When a pipeline breakdown creates a senior shortage (a wide span), the premium rises.

This introduces a richer form of cohort inequality than in the baseline. In the linear model, all post-shock junior cohorts earn the same wage; inequality arises only through different access to jobs. Under Cobb-Douglas, cohorts face a trade-off between access today and pay tomorrow. A cohort entering during a senior oversupply confronts a hiring freeze. But the scarcity of juniors starves the pipeline, so the few who are hired will eventually mature into a period of senior shortage, where the wide span delivers an elevated wage premium. Conversely, the next cohort enters during that shortage and finds abundant junior positions, but will mature into a abundant senior layer where the compressed span reduces the premium they earn. The oscillating pipeline thus generates alternating cycles of hard entry with high future reward and easy entry with low future reward.

6.2 Firm-Specific Human Capital

Until now, we assumed senior workers are perfect substitutes across firms. In practice, internal experience generates firm-specific human capital—knowledge of workflows, protocols, or relationships—that external hires lack. We now extend the framework to incorporate this friction, assuming the stability condition $\frac{\gamma^h}{\gamma^l} q < \frac{2-d_h}{1-d_l}$.

We distinguish two types of seniors. Internal seniors (promoted from within) have productivity γ^h. External seniors (hired from outside) lack firm-specific capital and produce only $\gamma^h - \sigma$, where $\sigma \in (0, \gamma^h - \gamma^l)$. Each firm maximizes the discounted profits subject to the capacity constraint and a new talent constraint:

$$n_{i,t+1}^{h,in} \leq (1-d_h)(n_{i,t}^{h,in} + n_{i,t}^{h,ex}) + (1-d_l)\tilde{q}_t n_{i,t}^l$$

where $n_{i,t}^{h,in}$ is the number of internal seniors, $n_{i,t}^{h,ex}$ is the number of external seniors, and $n_{i,t}^l$ is the number of juniors. The talent constraint is the difference from the baseline model: hiring a junior today generates the future value of an internal senior tomorrow.

Lemma 5. *With firm-specific human capital, the steady-state equilibrium is unique and symmetric. In this equilibrium, every firm retains its entire stock of internal seniors and hires no external seniors ($n_i^{h,ex} = 0$). Furthermore, the firm hires exactly enough juniors to finish the remaining tasks.*

This equilibrium structure comes from the productivity advantage of internal seniors. Because internal seniors possess specific human capital that makes them more productive than external hires, the firm captures a rent on every internal senior it retains. To maximize these rents, the firm promotes exclusively from within and lets no internal talent leave. At the same time, the firm employs just enough to meet the production target, ensuring no wage costs are wasted on idle capacity.

Lemma 6. *In the steady-state equilibrium, the wage ratio is given by:*

$$\frac{w^h}{w^l} = \frac{(\gamma^h - \sigma) + \delta(1-d_h)\sigma}{\gamma^l + \delta q(1-d_l)\sigma} =: \Gamma.$$

Firm-specific human capital introduces future value into the wage ratio. For juniors, the denominator adds $\delta q(1-d_l)\sigma$, capturing the future value of specific skill acquisition. For seniors, the numerator adds $\delta(1-d_h)\sigma$, capturing the future value of retaining an external senior who becomes internal tomorrow. Additionally, because the senior wage is tied to the external market, the numerator uses $(\gamma^h - \sigma)$ instead of γ^h. Under this wage ratio Γ, the individual firm's structure and the market wage levels mirror Proposition 1.

Therefore, an AI shock affects the long-run firm structure $(n^{h,in*}, n^{l*}, s^*)$ and wage levels (w^{l*} and w^{h*}) just as before. But the learning shock's impact on the wage ratio changes. Without specific human capital, the wage ratio w^h/w^l is constant. Here, faster learning increases the value of the junior layer relative to the seniors, compressing the wage ratio.

Corollary 1. *In the steady state, if the learning rate q increases, the wage ratio $\frac{w^{h*}}{w^{l*}}$ decreases.*

To understand the transition dynamics after an AI shock, we need to examine the firm's dynamic hiring strategy.

Lemma 7. *In a symmetric equilibrium, each firm retains all internal seniors* ($n_t^{h,in} = (1-d_h)(n_{t-1}^{h,in} + n_{t-1}^{h,ex}) + (1-d_l)qn_{t-1}^l$) *and hires no external seniors* ($n_t^{h,ex} = 0$). *Moreover, the firm hires just enough juniors to finish the remaining tasks* ($n_t^l = \frac{z-\gamma^h n_t^{h,in}}{\gamma^l}$).

Along the transition path, firms prefer internal hiring because internal seniors possess specific capital σ. Noticeably, firms also refrain from hiring excess junior workers just to build future internal seniors. Although a junior worker fills the talent pipeline, this future value remains lower than the worker's current wage. Thus, the firm minimizes costs by keeping both the capacity and talent constraints binding in every period. Consequently, the firm's structural evolution after an AI shock directly mirrors the dynamics established in Propositions 3 and 4.

Because internal seniors possess specific capital σ that external hires lack, the firm captures a rent on every internal senior it retains. Consequently, the firm's present discounted value $V(n_t^h)$ is strictly increasing in its internal senior stock. The following proposition characterizes how the AI shocks affect this value during the transition.

Proposition 7. *Suppose that productivity* (γ^h *or* γ^l) *or the learning probability* (q) *permanently increases at* $t = 1$.

(i) *At* $t = 1$ *the firm's value strictly increases* ($V(n_1^{h,in}) > V(n_0^{h,in})$).

(ii) *During the transition* ($t > 1$), *the firm's value* $V(n_t^{h,in})$ *comoves with the senior stock* $n_t^{h,in}$.

Proposition 7 shows that the AI shock strictly increases the firm's value immediately. Because AI permanently reduces the labor required to complete tasks, this efficiency gain outweighs the short-run cost of a talent pipeline breakdown. However, the subsequent transition may entail temporary inefficiencies: whenever the senior stock dips below the steady state during oscillations, the firm captures fewer specific-capital rents.

6.3 Free Entry and Endogenous Industry Scale

In the baseline model, the mass of firms M is fixed, so AI acts purely as a labor-saving shock. In practice, lower production costs attract new entrants, expand industry output, and may offset the displacement of workers. To capture this channel, we endogenize the number of firms M_t. Throughout, we maintain the stability condition $\frac{\gamma^h}{\gamma^l} q < \frac{2-d_h}{1-d_l}$.

Let the inverse demand function for tasks be $y(Q_t)$, with $y'(Q) < 0$, where $Q_t = M_t z$ is aggregate output. There is a per-period fixed cost $F > 0$ to operate a firm. At the beginning of each period, potential entrants observe the current state of the economy and decide whether to enter. Consequently, free entry drives per-period profits to zero.

In steady state, the span S^*, the wages w^{l*} and w^{h*}, and the per-firm employment levels (in symmetric equilibrium) n^{l*} and n^{h*} are all determined exactly as in the baseline model: the span is pinned by flow rates, and wages are pinned by the no-arbitrage condition and the free-entry condition for workers. Therefore, AI shocks affect the wage structure in the same way as described in Proposition 2. What free entry adds is a new margin of adjustment: the number of firms, and hence the scale of the industry.

To see how this margin is determined, note that the no-arbitrage condition $w^{h*}/w^{l*} = \gamma^h/\gamma^l$ and the binding capacity constraint $\gamma^l n^{l*} + \gamma^h n^{h*} = z$ together imply that the total wage bill of a firm equals $\frac{w^{l*}}{\gamma^l}z$. That is, the cost per task is w^{l*}/γ^l, regardless of the firm's workforce mix. The free-entry condition — revenue equals labor costs plus the fixed cost — then pins down the equilibrium price:

$$y(Q^*) = c^* \equiv \frac{w^{l*}}{\gamma^l} + \frac{F}{z}. \tag{1}$$

Because $y(\cdot)$ is strictly decreasing, this equation uniquely determines aggregate output Q^* and hence the number of firms $M^* = Q^*/z$.

Lemma 8 (Long-Run Scale with Free Entry)**.** *Any AI shock that increases productivity (γ^h or γ^l) or the learning rate (q) strictly reduces the equilibrium per-task cost c^*, thereby strictly increasing aggregate output Q^* and the equilibrium number of firms M^*.*

Furthermore, the effect on aggregate employment depends on the price elasticity of demand $\eta \equiv -\frac{y(Q)}{y'(Q)Q}$:

(*i*) *For productivity shock, aggregate employment for both seniors and juniors rises if and only if demand is sufficiently elastic (η is high).*

(*ii*) *For learning shock, aggregate senior employment unambiguously rises. Aggregate junior employment rises if and only if demand is sufficiently elastic.*

The intuition is as follows. AI lowers the cost of completing tasks, which reduces the equilibrium price and draws in new firms until profits return to zero. Each firm employs fewer workers than before, but there are more firms. The net effect on aggregate employment therefore depends on the elasticity of demand: if demand is elastic, the expansion in the number of firms more than offsets the per-firm labor savings, and total employment rises; if demand is inelastic, the labor-saving effect dominates.

We now turn to the transition following a permanent AI shock at $t = 1$. Because wages do not depend on the workforce state, per-task cost drops to the new steady-state level immediately, and free entry implies that the mass of active firms becomes $M^*_{AI} > M_0$ from

$t = 1$ onward. The aggregate supply of senior workers, however, is predetermined by the past talent pipeline. This tension between immediate firm entry and a sluggish senior supply shapes the short-run dynamics.

Proposition 8 (Transition Dynamics with Free Entry)**.** *Following a permanent AI shock at $t = 1$, the aggregate transition dynamics are as follows:*

(*i*) *For $t \geq 1$, the number of firms is $M_t = M_{AI}^* > M_0$.*

(*ii*) *At $t = 1$, the aggregate senior stock remains at N_0^{h*}. Under a learning shock, aggregate junior hiring strictly increases. Under a productivity shock, aggregate junior hiring increases if demand is sufficiently elastic, but decreases if demand is inelastic.*

(*iii*) *For $t \geq 2$, aggregate employment converges to the new steady state. The transition path oscillates if $q_{AI} \frac{\gamma_{AI}^h}{\gamma_{AI}^l} > \frac{1-d_h}{1-d_l}$ and is monotonic otherwise. When demand is elastic ($N_0^{h*} < N_{AI}^{h*}$), the oscillation begins with a junior hiring boom; when demand is inelastic ($N_0^{h*} > N_{AI}^{h*}$), it begins with a junior hiring freeze.*

The key difference from the baseline is in Part (ii). With a fixed number of firms, AI always generates excess capacity and leads to an immediate contraction of the junior layer. With free entry, AI is also an expansionary shock: new firms enter, and the economy must staff them. Because the economy cannot produce experienced workers on the spot, the new capacity must be met by hiring juniors. If entry is large enough—that is, if demand is elastic—this expansion effect dominates the labor-saving effect, and aggregate junior hiring rises at $t = 1$.

This initial boom, however, creates its own imbalance. The large cohort of juniors eventually acquires skills and matures into seniors, producing an oversupply of senior workers in later periods. Firms then cut back on junior hiring, starving the pipeline, which leads to a future shortage of seniors, and so on. The pipeline still oscillates, but the starting phase is reversed: elastic demand turns the baseline "freeze-then-boom" cycle into a "boom-then-freeze" cycle.

6.4 Rapid Learning

In the baseline model, we assume $q \frac{\gamma^h}{\gamma^l} < \frac{2-d_h}{1-d_l}$, which ensures convergence to a steady state. We now consider what happens when this condition fails, for example, because AI raises q past the threshold. To simplify the analysis, we assume that unemployed seniors lose their skills in the subsequent period, and that there is a minimum wage $w_{\min} > 0$ for employed workers that is not too large.

When learning is sufficiently rapid, the steady state of Proposition 1 still exists but is unstable. A small increase in the number of seniors displaces juniors from tasks, which starves the pipeline and reduces future seniors. But because juniors learn so fast, the subsequent correction overshoots: the firm hires too many juniors, who quickly mature into too many seniors, triggering another round of cuts. Unlike the convergent case in Section 5, these oscillations grow rather than shrink. The economy eventually reaches a corner in which seniors can complete all tasks on their own, and no juniors are hired. At that point, the talent pipeline shuts down entirely.

Proposition 9 (Long-Run Outcome with Rapid Learning)**.** *Suppose $q\frac{\gamma^h}{\gamma^l} \geq \frac{2-d_h}{1-d_l}$ and unemployed seniors lose their skills after one period. Then the economy does not converge to a steady state. Instead, it eventually enters a permanent two-period cycle that alternates between two states:*

(i) ***Oversupply state.*** *The stock of seniors is large enough to complete all tasks. The firm employs Z/γ^h seniors at wage $w_O^h = w_{\min}$ and hires no juniors.*

(ii) ***Shortage state.*** *The senior stock has been depleted to $(1-d_h)Z/\gamma^h$. The firm hires $d_h Z/\gamma^l$ juniors to fill the remaining tasks. The wages are w_S^l and $w_S^h = (\gamma^h/\gamma^l)w_S^l$, both of which are above the baseline steady state level.*

The two-period cycle is the extreme case of the oscillations identified earlier. In Section 5, oscillations arise during the transition but eventually dampen. Here, learning is so fast that each cohort of juniors matures into a mass of seniors, and the pipeline breaks and rebuilds permanently.

Two implications follow. First, cohort inequality becomes permanent. In the convergent case, different cohorts face unequal employment prospects during the transition, but the inequality fades as the economy settles. Under rapid learning, cohorts entering in the senior shortage state work and accumulate human capital; those entering in the senior oversupply state find no positions.

Second, the wage structure differs from the baseline. In the oversupply state, the oversupply of seniors pushes their wage down to $w_{\min}$. This affects the shortage state as well, because a junior hired during a shortage who acquires the skill enters an oversupply state where the senior wage is only $w_{\min}$. The future value is therefore much smaller than in the baseline steady state. With a smaller future value, firms must pay juniors more upfront, so $w_S^l > w^{l*}$. The no-arbitrage condition then implies $w_S^h > w^{h*}$ as well. Rather than the constant wages of the baseline transition, workers face a recurring pattern: wages at the floor during oversupply, and above their steady-state levels during shortage.

6.5 Job Assignment and Hierarchical Structure

The baseline model treats the productivity difference between seniors and juniors as a primitive: seniors complete γ^h tasks per period, juniors complete γ^l, and both types work on the same tasks. One might wonder whether this reduced form is consistent with a richer environment in which the firm assigns workers to distinct jobs based on their human capital. We now show that it is.

Suppose the firm has a job ladder with two levels, $j \in \{1, 2\}$. Following Gibbons and Waldman (1999*b*), the output of worker i on job j is a linear function of ability:

$$y_{ijt} = d_j + c_j \eta_{it},$$

where $\eta_{it} \in \{0, 1\}$ equals zero for a junior and one for a senior. Job 1 is a routine position with parameters $d_1 = \gamma^l$ and $c_1 = 0$: it yields γ^l regardless of who fills it. Job 2 is an advanced position with parameters $d_2 = 0$ and $c_2 = \gamma^h$: it yields γ^h when filled by a senior but zero when filled by a junior.

Given this structure, the efficient assignment is immediate: juniors should be placed in job 1 and seniors in job 2. A junior in job 2 produces nothing, while a senior in job 1 wastes her additional ability. Under this assignment, a junior produces γ^l and a senior produces γ^h, which recovers exactly the production structure of our baseline model. The hierarchical shape of the firm — juniors at the bottom, seniors at the top — therefore emerges endogenously from optimal job assignment rather than being assumed directly.

7 Empirical Results

We use job postings data to ask whether the margins emphasized by the model are visible in observed hiring patterns. The data come from Burning Glass Institute U.S. postings and are restricted to company-source job boards. Junior postings are those requiring 0–2 years of experience, while senior postings are those requiring 6+ years. The quarterly sample runs from 2018-Q1 through 2025-Q4. The underlying samples are economically large: the tech/platform figures use 618,471 postings across 55 firms, the finance figures use 1,162,466 postings across 605 firms, and the sector comparison covers five NAICS-defined sectors whose average quarterly posting counts range from about 15,500 to 140,600. The purpose of the exercise below is to ask whether the data move along the quantity and wage margins highlighted by Propositions 2, 3, 4, and 5, rather than to identify the underlying AI shock.

7.1 Junior Hiring Freezes and Changes in Hiring Structure

Propositions 3 and 4 imply that AI can disrupt the junior pipeline during the transition. In the productivity case, junior hiring falls on impact because the inherited stock of seniors can complete a larger share of the firm's tasks. In the learning case, junior hiring can fall with a lag once a wave of promotions enlarges the senior layer. Figure 4 examines whether such disruptions are visible in the data. For tech/platform and finance, it plots the share of firms with no junior postings in a quarter and the share of firms whose junior postings fall to a small fraction of their own recent baseline. Both series are constructed at the firm-quarter level and then aggregated to the sector level.

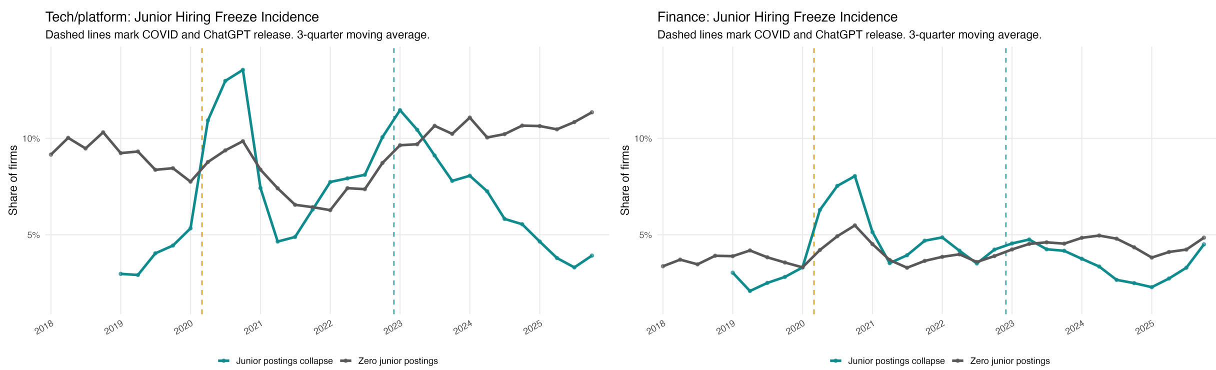


Figure 4: Incidence of junior-hiring freezes in tech/platform and finance.

Notes: The left panel shows tech/platform and the right panel shows finance. "Zero junior postings" is the share of classifiable firms with no junior postings in a quarter. "Junior postings collapse" is the share of firms whose junior postings fall to 20 percent or less of their own prior four-quarter junior-posting baseline, among firms with sufficient recent activity. Because this is defined relative to a firm's own lagged baseline, the collapse series can decline even when the zero-junior margin remains elevated. Both series are shown as 3-quarter moving averages, are not seasonally adjusted, and use company-source job-board postings only. The displayed quarterly calendar window runs from 2018-Q1 through 2025-Q4; no 2026 quarters are plotted.

In both tech/platform and finance, the post-2022 period is associated with a visibly higher incidence of firms with either no junior postings at all or a sharp collapse in junior postings relative to their own recent baseline. This is the closest empirical counterpart in the data to the temporary hiring freezes described by the transition results. The figure does not by itself distinguish sharply between the productivity and learning channels, since both can generate a thinner junior intake during adjustment. It does, however, show that the junior pipeline becomes materially less active in these sectors after late 2022. A decline in the "collapse" line should therefore not be read as a full normalization of junior hiring, since firms can move from a sharp-drop state into a persistently low-junior or zero-junior state.

If the inflow into the junior layer weakens persistently, the vacancy-side hierarchy should also flatten. This margin is especially important because Proposition 2 shows that, in the baseline model, only the learning shock changes the long-run junior-to-senior ratio; productivity shocks change the scale of employment but leave the span unchanged. Figure 5 therefore turns to the junior/senior postings ratio itself and plots it for tech/platform and finance using a common 3-quarter moving average.

In both sectors the ratio declines after the AI period begins, with the change especially pronounced in tech/platform and also visible in finance. This is the clearest descriptive counterpart to a movement *from a pyramid toward a diamond*. Viewed through the lens of Proposition 2, the figure points more naturally toward the learning channel than toward a pure productivity channel, since a persistent decline in the junior/senior ratio is the margin that the learning shock changes in the long run. At the same time, the figure remains descriptive. A sector-level decline in the vacancy-side span could also reflect changes in firm composition, differences in the timing of AI adoption, or other forces that are outside the model.

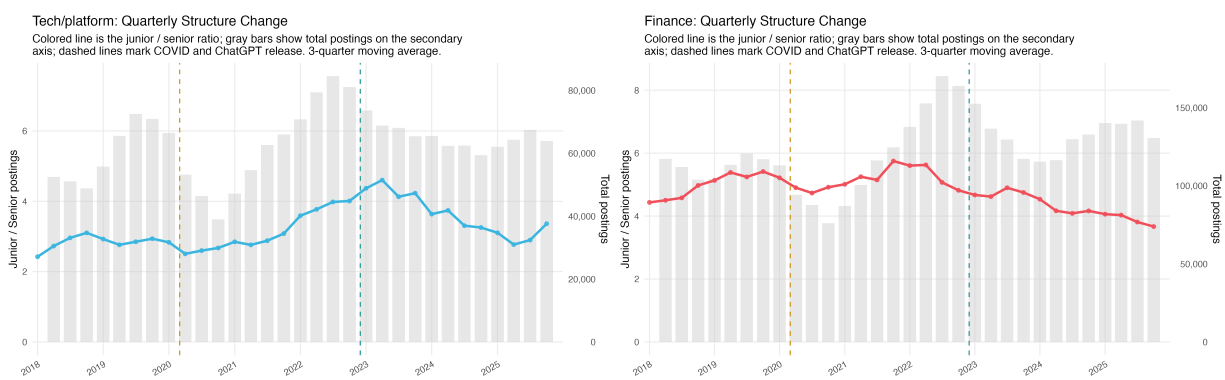


Figure 5: Junior-to-senior postings ratio in tech/platform and finance.

Notes: The left panel plots the quarterly mean of the firm-level junior/senior postings ratio for tech/platform and the right panel plots the same object for finance, using all classifiable firms in each sector. Both the colored line and the light gray bars are shown as 3-quarter trailing moving averages; the bars represent total postings on the secondary axis. The figures are not seasonally adjusted. The displayed quarterly calendar window runs from 2018-Q1 through 2025-Q4; no 2026 quarters are plotted.

The two figures above establish the quantity-side evidence in a narrow but transparent way. The remaining question is whether the sectors that exhibit flatter hiring pyramids also exhibit changes in relative pay that line up with the theoretical predictions. That comparison matters because the paper distinguishes learning and productivity channels most clearly when quantity and wage movements are read together.

7.2 Sector-Level Changes in Hiring Structure and Wage Premia

Figure 6 brings together the two empirical margins that speak most directly to the theory: the junior-to-senior hiring ratio and the senior-to-junior salary ratio. For each sector, we compute the average quarterly junior/senior postings ratio over 2021–2022 and over 2024–2025, and we do the same for the quarterly senior/junior posted-salary ratio. The figure then plots the percent change between the later and earlier window. This deliberately compresses the time series into a single descriptive comparison. The exercise is therefore not meant to identify a structural shock. Rather, it is meant to show whether the joint movement in quantities and relative pay looks more consistent with one theoretical channel than another. For the current run, the average quarterly postings across sectors range from about 15,500 to 140,600, and the average number of firms per quarter ranges from about 1,700 to 6,100.

The usefulness of the figure is that the model's comparative statics are inherently two-dimensional. In Proposition 2, a senior-productivity shock should move sectors mainly upward with limited horizontal movement: the wage premium rises while the long-run span remains roughly unchanged. A junior-productivity shock should move sectors mainly downward, again with limited horizontal movement. A learning shock should move sectors mainly leftward, because the junior/senior ratio falls while the wage ratio changes little.

The extensions imply some nuances beyond this benchmark. Under Cobb-Douglas type production and under firm-specific human capital, a flatter hierarchy can coincide with a compressed wage premium, so a left-and-down pattern can still be consistent with learning-based mechanisms rather than only with junior-productivity shocks.

Finance is the case that is most in line with the baseline learning-shock prediction: the junior/senior ratio falls markedly while the senior/junior salary ratio changes relatively little. Accounting lies closer to a senior-productivity pattern, with limited movement in the hiring ratio and a higher wage premium. Tech/platform and IT services also move leftward, indicating flatter hiring pyramids, but they do so alongside a narrowing wage premium. That pattern is harder to reconcile with the baseline linear model alone, but it is consistent with the richer versions of the model in which the wage premium compresses as the pyramid flattens. Strategy lies in between and is best read as mixed.

Taken together, the figure is consistent with the paper's broader message that AI is associated with a contraction of the junior layer, while the wage response is heterogeneous across sectors. In that sense, the quantity side of the evidence is the most uniform, whereas the wage side helps distinguish which theoretical channel looks more plausible in different parts of the economy. Much more empirical work remains to be done before treating this as a test of the model, but as a first descriptive pass the figure is useful for organizing which channels appear more or less plausible in the current data.

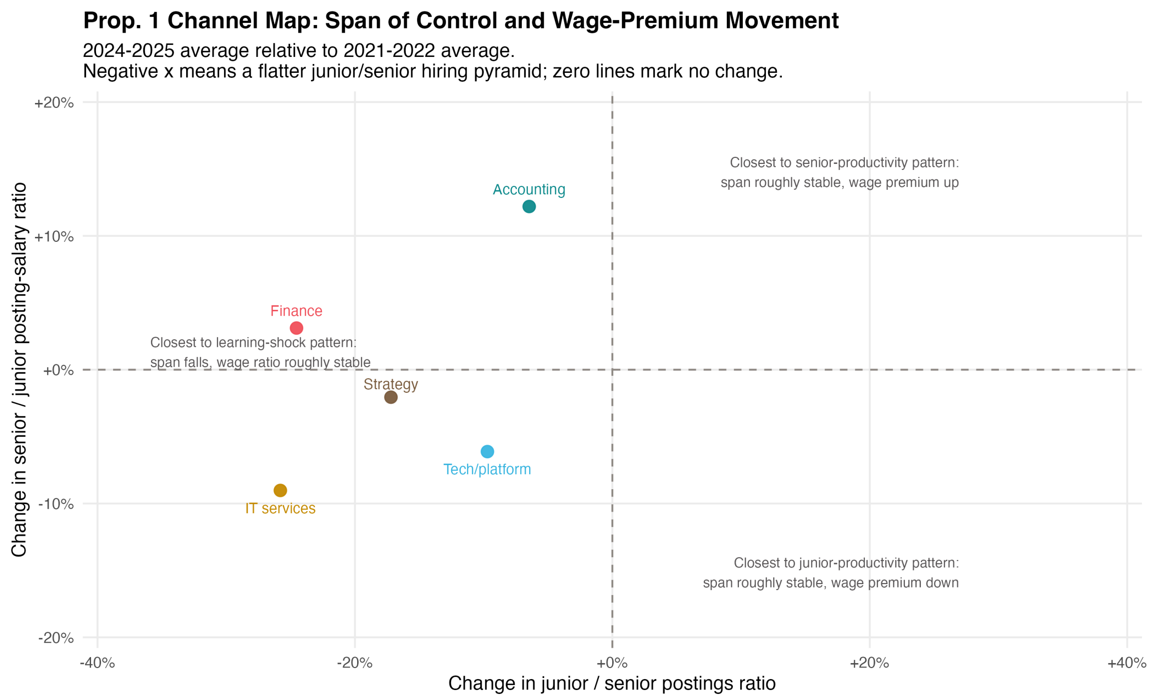


Figure 6: Sector-level changes in hiring structure and wage premia.

Notes: Each point is a NAICS-defined sector built from company-source job-board postings only: tech/platform (51), finance (52), IT services (5415), accounting (5412), and strategy consulting (54161). The x-axis is the percent change in the quarterly junior/senior postings ratio between the 2021–2022 average and the 2024–2025 average, so points to the left indicate a flatter vacancy-side pyramid. The y-axis is the percent change in the quarterly senior/junior posted-salary ratio over the same windows. Posted salaries are constructed from advertised salary ranges using the midpoint when both endpoints are observed, with one-sided and single-salary fallbacks when necessary. Each window contains eight quarters. Because the empirical objects are vacancy flows and advertised salaries rather than internal staffing stocks and realized wages, the figure should be read as descriptive evidence on which theoretical channel appears more consistent with the data, not as structural shock identification.

8 Concluding Remarks

There is little doubt that the massive technology shock of AI will transform firms substantially. Even more so, it will transform the career logic that firms have followed over many decades: Hire top talent, develop it through the ranks of the hierarchy, and thus maintain strategic advantages in product markets.

Despite its simplicity, our model captures some typical features of ILM, in particular, the need of firms to employ juniors to fill the talent pipeline. The model also allows for thinking about the different ways in which AI affects workplaces, individual productivity gains on different layers and the speed of learning (here, modeled through the short cut of how likely a junior person acquires the skills needed to become senior).

The model offers implications about firm size (generally falling), firm profit and number of firms under free entry. It also provides an analytical framework to think about worker welfare. We show that because of oscillations between the steady state, some cohorts of juniors might be hit by negative, and others by positive employment shocks. The concerns about future labor markets may be well grounded, but probably more because of oscillations than new steady states.

We deviate from a typical knowledge hierarchy as in Garicano (2000) and Garicano and Rossi-Hansberg (2015). Rather we consider a top-down hierarchy in which a CEO implements strategy by allocating tasks to agents in the different hierarchical layers, or, equivalently, delegates work packages that seniors can delegate further to the junior workers. This makes it possible to focus on the effect of juniors' and seniors' productivity effects, and learning but does not allow for considering how changes in problem solving capacity may affect the hierarchy. We have also abstracted from incentive issues that are certainly an important element of talent management[4] and may need complementary analysis. Lastly, we assume that all bargaining power is held by the firms. See, for example, Huang, Li and Luo (2026) for a more balanced treatment of the bargaining power in the labor market when AI improves learning.

We hope though this model is a good start to think about these important issues and would see it as an advantage that it is embedded in a labor market and offers reflections on the future of organizations and junior wages. We are also confident that we can allow for changes in the business model in the industry and are working on it.

References

Arrow, Kenneth J. 1962. "The economic implications of learning by doing." *The review of economic studies*, 29(3): 155–173. 1

Auriol, Emmanuelle, Guido Friebel, and Frauke Von Bieberstein. 2016. "The firm as the locus of social comparisons: Standard promotion practices versus up-or-out." *Journal of Economic Behavior & Organization*, 121: 41–59. 3

[4]There is a large literature on this, for instance, Waldman (1984); Milgrom and Oster (1987); Carmichael (1988); Fairburn and Malcomson (1994); Prendergast and Topel (1996); Bar-Isaac and Leaver (2021). Recently, many researchers investigate more on the role of manager's incentives (Hoffman and Tadelis, 2021; Friebel, Heinz and Zubanov, 2022; Minni, 2023; Haegele, 2024; Friebel and Raith, 2026).

Baker, George, Michael Gibbs, and Bengt Holmstrom. 1994*a*. "The internal economics of the firm: Evidence from personnel data." *The Quarterly Journal of Economics*, 109(4): 881–919. 1, 3

Baker, George, Michael Gibbs, and Bengt Holmstrom. 1994*b*. "The wage policy of a firm." *The Quarterly Journal of Economics*, 109(4): 921–955. 3

Bar-Isaac, Heski, and Clare Leaver. 2021. "Investment in Human Capital Under Endogenous Asymmetric Information." 32

Bastani, Hamsa, Osbert Bastani, Alp Sungu, Haosen Ge, Özge Kabakcı, and Rei Mariman. 2025. "Generative AI without guardrails can harm learning: Evidence from high school mathematics." *Proceedings of the National Academy of Sciences*, 122(26): e2422633122. 5

Beckmann, Martin. 1978. *Rank in organizations.* Berlin:Springer. 4

Bianchi, Nicola, Giulia Bovini, Jin Li, Matteo Paradisi, and Michael Powell. 2023. "Career spillovers in internal labour markets." *The Review of Economic Studies*, 90(4): 1800–1831. 3, 4

Brynjolfsson, Erik, Bharat Chandar, and Ruyu Chen. 2025. "Canaries in the Coal Mine? Six Facts about the Recent Employment Effects of Artificial Intelligence." Stanford Digital Economy Lab Working Paper. 1

Brynjolfsson, Erik, Danielle Li, and Lindsey Raymond. 2025. "Generative AI at work." *The Quarterly Journal of Economics*, 140(2): 889–942. 5

Carmichael, H Lorne. 1988. "Incentives in academics: why is there tenure?" *Journal of political Economy*, 96(3): 453–472. 32

Chen, Yvonne Jie, Jie Gong, Jin Li, and Zibo Zhao. 2025. "Better Technology, Worse Motivation: GenAI's Mediocrity Trap." *Available at SSRN 5208163*. 5

Doeringer, Peter B., and Michael J. Piore. 1971. *Internal Labor Markets and Manpower Analysis.* Lexington, Mass:Heath. 1, 3

Fairburn, James A, and James M Malcomson. 1994. "Rewarding performance by promotion to a different job." *European Economic Review*, 38(3-4): 683–690. 32

Friebel, Guido, and Elena Panova. 2008. "Insider Privatization and Careers: A Clinical Study of a Russian Firm in Transition." In *The Analysis of Firms and Employees: Quantitative and Qualitative Approaches*. , ed. Stefan Bender, Julia Lane, Kathryn L. Shaw, Fredrik Andersson and Till von Wachter, 253–266. Chicago, IL:University of Chicago Press. 4

Friebel, Guido, and Michael Raith. 2026. "Talent Management: The Role of Bosses." *Journal of Labor Economics*. Forthcoming. 3, 32

Friebel, Guido, Matthias Heinz, and Nikolay Zubanov. 2022. "Middle managers, personnel turnover, and performance: A long-term field experiment in a retail chain." *Management Science*, 68(1): 211–229. 32

Fudenberg, Drew, and Luis Rayo. 2019. "Training and effort dynamics in apprenticeship." *American Economic Review*, 109(11): 3780–3812. 4

Garicano, Luis. 2000. "Hierarchies and the Organization of Knowledge in Production." *Journal of political economy*, 108(5): 874–904. 32

Garicano, Luis, and Esteban Rossi-Hansberg. 2015. "Knowledge-based hierarchies: Using organizations to understand the economy." *Annual Review of Economics*, 7(1): 1–30. 32

Garicano, Luis, and Luis Rayo. 2017. "Relational knowledge transfers." *American Economic Review*, 107(9): 2695–2730. 4

Garicano, Luis, and Luis Rayo. 2025. "Training in the Age of AI: A Theory of Apprenticeship Viability." Centre for Economic Policy Research Discussion Paper DP20634. 4

Gibbons, Robert. 1997. "Incentives and Careers in Organizations." In *Advances in Econometrics: Theory and Applications*. Vol. 2, , ed. David M. Kreps and Kenneth J. Wallis, Chapter 1, 1–37. Cambridge:Cambridge University Press. 3

Gibbons, Robert, and Michael Waldman. 1999*a*. "Careers in organizations: Theory and evidence." In *Handbook of Labor Economics*. Vol. 3, 2373–2437. Elsevier. 1, 3

Gibbons, Robert, and Michael Waldman. 1999*b*. "A theory of wage and promotion dynamics inside firms." *The Quarterly Journal of Economics*, 114(4): 1321–1358. 3, 7, 27

Gibbons, Robert, and Michael Waldman. 2004. "Task-specific human capital." *American Economic Review*, 94(2): 203–207. 1

Haegele, Ingrid. 2024. "Talent hoarding in organizations." *arXiv preprint arXiv:2206.15098*. 32

Hoffman, Mitchell, and Steven Tadelis. 2021. "People management skills, employee attrition, and manager rewards: An empirical analysis." *Journal of political economy*, 129(1): 243–285. 32

Huang, Yao, Jin Li, and Yuchen Luo. 2026. "Artificial Intelligence, On-the-Job Learning, and Labor Market Effects." Working paper. 32

Huitfeldt, Ingrid, Andreas R Kostøl, Jan Nimczik, and Andrea Weber. 2023. "Internal labor markets: A worker flow approach." *Journal of Econometrics*, 233(2): 661–688. 3

Hui, Xiang, Oren Reshef, and Luofeng Zhou. 2024. "The short-term effects of generative artificial intelligence on employment: Evidence from an online labor market." *Organization Science*, 35(6): 1977–1989. 1

Humlum, Anders, and Emilie Vestergaard. 2025. "The unequal adoption of ChatGPT exacerbates existing inequalities among workers." *Proceedings of the National Academy of Sciences*, 122(1): e2414972121. 5

Humlum, Anders, and Emilie Vestergaard. 2026. "Still Waters, Rapid Currents: Early Labor Market Transformation under Generative AI." Working paper. 1

Ide, Enrique. 2026. "Automation, AI, and the intergenerational transmission of knowledge." *arXiv preprint arXiv:2507.16078*. 4

Ke, Rongzhu, Jin Li, and Michael Powell. 2018. "Managing careers in organizations." *Journal of Labor Economics*, 36(1): 197–252. 3, 4

Kräkel, Matthias, and Anja Schöttner. 2012. "Internal labor markets and worker rents." *Journal of Economic Behavior & Organization*, 84(2): 491–509. 3

Lazear, Edward P., and Paul Oyer. 2013. "Personnel Economics." In *Handbook of Organizational Economics*. , ed. Robert Gibbons and John Roberts, Chapter 12, 479–519. Princeton, NJ:Princeton University Press. 3

Lazear, Edward P, and Sherwin Rosen. 1981. "Rank-order tournaments as optimum labor contracts." *Journal of Political Economy*, 89(5): 841–864. 3

Lichtinger, Guy, and Seyed Mahdi Hosseini Maasoum. 2025. "Generative AI as seniority-biased technological change: Evidence from US resume and job posting data." *Available at SSRN*. 1

MacLeod, W Bentley, and James M Malcomson. 1988. "Reputation and hierarchy in dynamic models of employment." *Journal of Political Economy*, 96(4): 832–854. 3

Malcomson, James M. 1984. "Work incentives, hierarchy, and internal labor markets." *Journal of political economy*, 92(3): 486–507. 3

Milgrom, Paul, and Sharon Oster. 1987. "Job discrimination, market forces, and the invisibility hypothesis." *The Quarterly Journal of Economics*, 102(3): 453–476. 32

Minni, Virginia. 2023. "Making the invisible hand visible: Managers and the allocation of workers to jobs." 32

Noy, Shakked, and Whitney Zhang. 2023. "Experimental evidence on the productivity effects of generative artificial intelligence." *Science*, 381(6654): 187–192. 5

Osterman, Paul. 2024. "Reconsidering occupational internal labor markets: Incidence and consequences." *ILR Review*, 77(3): 366–395. 3

Pastorino, Elena. 2024. "Careers in firms: The role of learning about ability and human capital acquisition." *Journal of Political Economy*, 132(6): 1994–2073. 3

Prendergast, Canice. 1993. "The role of promotion in inducing specific human capital acquisition." *The Quarterly Journal of Economics*, 108(2): 523–534. 3

Prendergast, Canice, and Robert H Topel. 1996. "Favoritism in organizations." *Journal of Political Economy*, 104(5): 958–978. 32

Priest, Dan. 2025. "In the age of AI: Speed matters more, scale matters less, innovation matters most." *PwC Tech Effect*, Accessed: January 14, 2026. 1

Rosenbaum, James. 1984. *Career mobility in a corporate hierarchy.* New York:Academic Press. 4

Rosen, Sherwin. 1986. "Prizes and incentives in elimination tournaments." *American Economic Review*, 76(4): 701–715. 3

Sharwood, Simon. 2025. "AWS CEO says using AI to replace junior staff is 'Dumbest thing I've ever heard'." 1

Simon, Herbert A. 1951. "The analysis of promotional opportunities." *Personnel*, 27(4): 282–285. 4

Stewman, Shelby, and Suresh Konda. 1983. "Careers and organizational labor markets: Demographic models of organizational behavior." *American Journal of Sociology*, 88(4): 637–685. 4

Waldman, Michael. 1984. "Job assignments, signalling, and efficiency." *The Rand journal of economics*, 15(2): 255–267. 32

Waldman, Michael. 2003. "Ex ante versus ex post optimal promotion rules: The case of internal promotion." *Economic Inquiry*, 41(1): 27–41. 3

Waldman, Michael. 2013. "Theory and evidence in internal labor markets." In *The handbook of organizational economics*. , ed. Robert Gibbons and John Roberts, Chapter 13, 520–571. Princeton University Press. 1, 3

White, Harrison C. 1970. *Chains of opportunity: System models of mobility in organizations.* Harvard University Press. 4

Zabojnik, Jan, and Dan Bernhardt. 2001. "Corporate tournaments, human capital acquisition, and the firm Size—Wage relation." *The Review of Economic Studies*, 68(3): 693–716. 3

A Proofs

Proof of Lemma 1. We solve the simplified optimization problem using the Lagrangian function:

$$\begin{aligned}\mathcal{L} =&(z^l_{i,t} + z^h_{i,t})y - w^h_t n^h_{i,t} - w^l_t n^l_{i,t}\\ &+ \lambda_{1,t}(\gamma^l n^l_{i,t} - z^l_{i,t})\\ &+ \lambda_{2,t}(\gamma^h n^h_{i,t} - z^h_{i,t})\\ &+ \lambda_{3,t}(z - z^l_{i,t} - z^h_{i,t}),\end{aligned}$$

where $\lambda_{1,t}$, $\lambda_{2,t}$ and $\lambda_{3,t}$ are Lagrangian multipliers for the capacity constraints (CC). The first order conditions on the four control variables are

$$\begin{aligned} n^l_{i,t} :&\quad -w^l_t + \lambda_{1,t}\gamma^l = 0;\\ n^h_{i,t} :&\quad -w^h_t + \lambda_{2,t}\gamma^h = 0;\\ z^l_{i,t} :&\quad y - \lambda_{1,t} - \lambda_{3,t} = 0;\\ z^h_{i,t} :&\quad y - \lambda_{2,t} - \lambda_{3,t} = 0.\end{aligned}$$

This shows that $\lambda_{1,t} = w^l_t/\gamma^l > 0$ and $\lambda_{2,t} = w^h_t/\gamma^h > 0$, meaning that $\gamma^l n^l_{i,t} = z^l_{i,t}$ and $\gamma^h n^h_{i,t} = z^h_{i,t}$. Moreover, $\lambda_{3,t} = y - \lambda_{1,t} = y - w^l_t/\gamma^l$. From ($VF^l$), we have

$$w^l_t = v^l_t - \delta(1-d_l)[qv^h_{t+1} + (1-q)v^l_{t+1}] - \delta d_l \underline{v}.$$

Free entry condition makes sure that $v^l_t = v^l_{t+1} = \underline{v}$. Since $v^h_{t+1} \geq \underline{v}$, we must have $w^l_t \leq (1-\delta)\underline{v}$. Together with assumption of a sufficiently large y, we have $\lambda_{3,t} > 0$, meaning that $z^l_{i,t} + z^h_{i,t} = z$. ■

Proof of Lemma 2. From the proof of Lemma 1, we have $\lambda_{1,t} = y - \lambda_{3,t} = \lambda_{2,t}$. This means $w^l_t/\gamma^l = w^h_t/\gamma^h$. ■

Proof of Proposition 1. Part (i): In a symmetric equilibrium, the flow condition for a representative firm is given by

$$N^h_{t+1} = (1-d_h)N^h_t + q(1-d_l)N^l_t.$$

Together with the steady state definition, we have

$$S^* = \frac{N^{l*}}{N^{h*}} = \frac{d_h}{q(1-d_l)}.$$

Lemma 1 shows that all the capacity constraints (CC) bind. This means $N_t^l = (Z - \gamma^h N_t^h)/\gamma^l$. Hence, we have

$$N^{h*} = (1-d_h)N^{h*} + \frac{q}{\gamma^l}(1-d_l)(Z - \gamma^h N^{h*}),$$

or $N^{h*} = \frac{q(1-d_l)}{\gamma^l d_h + \gamma^h q(1-d_l)} Z = \frac{1}{\gamma^l S^* + \gamma^h} Z$. Hence, $N^{l*} = (Z - \gamma^h N^{h*})/\gamma^l = \frac{d_h}{\gamma^l d_h + \gamma^h q(1-d_l)} Z = \frac{S^*}{\gamma^l S^* + \gamma^h} Z$.

Part (ii): From Lemma 2, we have $w_t^l/\gamma^l = w_t^h/\gamma^h$. Therefore, in a steady state, we must have

$$w^{h*} = \frac{\gamma^h}{\gamma^l} w^{l*}.$$

With the steady state and free entry, we have $v_t^l = v_{t+1}^l = \underline{v}$ and $v_t^h = v_{t+1}^h$. Together with the value functions (VF^l) and (VF^h), we have

$$\underline{v} = w^{l*} + \delta(1-d_l)[q v_t^h + (1-q)\underline{v}] + \delta d_l \underline{v},$$
$$v_t^h = \frac{\gamma^h}{\gamma^l} w^{l*} + \delta(1-d_h) v_t^h + \delta d_h \underline{v}.$$

Hence, we can get $w^{l*} = \underline{v}(1-\delta)\left[\frac{1-\delta(1-d_h)+\delta q(1-d_l)}{1-\delta(1-d_h)+\frac{\gamma^h}{\gamma^l}\delta q(1-d_l)}\right]$. ■

Proof of Proposition 2. These results are directly from Proposition 1.

Part (i):

$$\frac{\partial N^{h*}}{\partial \gamma^h} < 0, \quad \frac{\partial N^{l*}}{\partial \gamma^h} < 0, \quad \frac{\partial S^*}{\partial \gamma^h} = 0, \quad \frac{\partial w^{l*}}{\partial \gamma^h} < 0 \quad \text{and} \quad \frac{\partial w^{h*}}{\partial \gamma^h} > 0.$$

Part (ii):

$$\frac{\partial N^{h*}}{\partial \gamma^l} < 0, \quad \frac{\partial N^{l*}}{\partial \gamma^l} < 0, \quad \frac{\partial S^*}{\partial \gamma^l} = 0, \quad \frac{\partial w^{l*}}{\partial \gamma^l} > 0 \quad \text{and} \quad \frac{\partial w^{h*}}{\partial \gamma^l} < 0.$$

Part (iii):

$$\frac{\partial N^{h*}}{\partial q} > 0, \quad \frac{\partial N^{l*}}{\partial q} < 0, \quad \frac{\partial S^*}{\partial q} < 0, \quad \frac{\partial w^{l*}}{\partial q} < 0 \quad \text{and} \quad \frac{\partial w^{h*}}{\partial q} < 0.$$

From Lemma 2, $w^{h*}/w^{l*} = \gamma^h/\gamma^l$, meaning that a change in q has no effect on the wage ratio. ■

Proof of Proposition 3. In equilibrium, the flow condition for a representative firm is:

$$N_{t+1}^h = (1-d_h)N_t^h + q(1-d_l)N_t^l. \tag{2}$$

We analyze the transition dynamics using (2) and the binding capacity constraints from Lemma 1.

Part (i): Impact at $t = 1$. At the beginning of $t = 1$, the number of seniors is inherited from the pre-shock steady state: $N_1^h = N_0^{h*}$. With the new productivities $(\gamma_{AI}^h, \gamma_{AI}^l)$, the binding capacity constraint requires:

$$N_1^l = \frac{Z - \gamma_{AI}^h N_0^{h*}}{\gamma_{AI}^l}.$$

The assumption $\gamma_{AI}^h < Z/N_0^{h*}$ ensures $N_1^l > 0$. Since $\gamma_{AI}^h > \gamma_0^h$ and/or $\gamma_{AI}^l > \gamma_0^l$, we have

$$N_1^l = \frac{Z - \gamma_{AI}^h N_0^{h*}}{\gamma_{AI}^l} < \frac{Z - \gamma_0^h N_0^{h*}}{\gamma_0^l} = N_0^{l*}.$$

The span at $t = 1$ is $S_1 = N_1^l / N_0^{h*} < N_0^{l*}/N_0^{h*} = S_0^*$.

Part (ii): Dynamics for $t \geq 2$. Substituting $N_t^l = (Z - \gamma_{AI}^h N_t^h)/\gamma_{AI}^l$ into the flow equation yields the linear first-order difference equation:

$$N_{t+1}^h = \underbrace{\left[(1 - d_h) - \frac{q(1-d_l)\gamma_{AI}^h}{\gamma_{AI}^l}\right]}_{\equiv \phi} N_t^h + \frac{q(1-d_l)Z}{\gamma_{AI}^l}.$$

The general solution is $N_t^h - N_{AI}^{h*} = (\phi)^{t-1}(N_1^h - N_{AI}^{h*})$. Under the stability condition, $|\phi| < 1$.

We verify that $N_t^l > 0$ for all t. From Proposition 2, $N_{AI}^{h*} < N_0^{h*} = N_1^h$. By assumption, $N_1^h < Z/\gamma_{AI}^h$. Because $|\phi| < 1$, the distance from steady state shrinks: $|N_t^h - N_{AI}^{h*}| \leq |N_1^h - N_{AI}^{h*}|$, so $N_t^h \leq \max(N_1^h, N_{AI}^{h*}) = N_1^h < Z/\gamma_{AI}^h$, ensuring $N_t^l > 0$ for all t.

The transition path depends on the sign of ϕ:

Case (a): If $\frac{\gamma_{AI}^h}{\gamma_{AI}^l} > \frac{1-d_h}{q(1-d_l)}$, then $\phi < 0$ (with $\phi \in (-1, 0)$ under stability). The deviation $N_t^h - N_{AI}^{h*}$ alternates in sign, so N_t^h oscillates around the new steady state. Consequently, N_t^l and the span S_t fluctuate.

Case (b): If $\frac{\gamma_{AI}^h}{\gamma_{AI}^l} \leq \frac{1-d_h}{q(1-d_l)}$, then $\phi \in [0, 1)$. Since $N_1^h > N_{AI}^{h*}$, the sequence $\{N_t^h\}$ decreases monotonically toward N_{AI}^{h*}, and N_t^l increases monotonically. The span S_t increases monotonically toward S^*. ■

Proof of Proposition 4. The proof follows the same structure as that of Proposition 3.

Part (i): Impact at $t = 1$. At $t = 1$, the learning rate changes but productivities are unchanged. Neither the stock of seniors nor the capacity constraint changes at $t = 1$: $N_1^h = N_0^{h*}$ and $N_1^l = (Z - \gamma^h N_0^{h*})/\gamma^l = N_0^{l*}$. The firm structure is unchanged.

Part (ii): Dynamics for $t \geq 2$. Substituting $N_t^l = (Z - \gamma^h N_t^h)/\gamma^l$ into the flow equation with the new learning rate q_{AI} yields:

$$N^h_{t+1} = \left[(1-d_h) - \frac{q_{AI}(1-d_l)\gamma^h}{\gamma^l}\right] N^h_t + \frac{q_{AI}(1-d_l)Z}{\gamma^l}.$$

Let $\phi = (1-d_h) - q_{AI}(1-d_l)\gamma^h/\gamma^l$. The general solution is $N^h_t - N^{h*}_{AI} = (\phi)^{t-2}(N^h_2 - N^{h*}_{AI})$. Under the stability condition $|\phi| < 1$.

At $t = 2$, the senior stock is $N^h_2 = (1-d_h)N^{h*}_0 + q_{AI}(1-d_l)N^{l*}_0$. Since $q_{AI} > q_0$, we have $N^h_2 > (1-d_h)N^{h*}_0 + q_0(1-d_l)N^{l*}_0 = N^{h*}_0$, so the senior stock increases at $t = 2$.

We verify $N^l_t > 0$ for all t. Since neither γ^h nor γ^l changes, and the new steady state has $N^{h*}_{AI} > N^{h*}_0$ (Proposition 2), we need $N^h_t < Z/\gamma^h$ for all t. This follows because $N^{h*}_{AI} = Z/(\gamma^l S^*_{AI} + \gamma^h) < Z/\gamma^h$, and the oscillations are bounded by the stability condition.

The transition path depends on the sign of ϕ:

Case (a): If $q_{AI} > \frac{1-d_h}{1-d_l}\frac{\gamma^l}{\gamma^h}$, then $\phi < 0$, and N^h_t oscillates around N^{h*}_{AI}.

Case (b): If $q_{AI} \leq \frac{1-d_h}{1-d_l}\frac{\gamma^l}{\gamma^h}$, then $\phi \geq 0$, and N^h_t increases monotonically toward N^{h*}_{AI}. ■

Proof of Lemma 3. The capacity constraint binds: $(\gamma^h n^h)^\alpha(\gamma^l n^l)^{1-\alpha} = z$. The Lagrangian for the cost-minimization problem is $\mathcal{L} = w^h n^h + w^l n^l - \lambda[(\gamma^h n^h)^\alpha(\gamma^l n^l)^{1-\alpha} - z]$. The first-order conditions are:

$$w^h = \lambda\alpha(\gamma^h)^\alpha(n^h)^{\alpha-1}(\gamma^l n^l)^{1-\alpha},$$
$$w^l = \lambda(1-\alpha)(\gamma^h n^h)^\alpha(\gamma^l)^{1-\alpha}(n^l)^{-\alpha}.$$

Dividing yields $w^h/w^l = [\alpha/(1-\alpha)](n^l/n^h) = [\alpha/(1-\alpha)]s_t$. ■

Proof of Lemma 4. The span $s^* = d_h/[q(1-d_l)]$ follows from the flow condition as in the baseline. Given s^*, the capacity constraint $(\gamma^h n^{h*})^\alpha(\gamma^l s^* n^{h*})^{1-\alpha} = z$ yields:

$$n^{h*} = \frac{z}{(\gamma^h)^\alpha(\gamma^l)^{1-\alpha}(s^*)^{1-\alpha}}.$$

For wages, the free-entry condition requires $v^l = \underline{v}$ for all t. In steady state, the senior Bellman equation gives $v^{h*} = (w^{h*} + \delta d_h \underline{v})/[1-\delta(1-d_h)]$. Substituting into the junior Bellman equation yields:

$$\underline{v} = w^{l*} + \delta(1-d_l)q\, v^{h*} + \delta(1-d_l)(1-q)\underline{v} + \delta d_l \underline{v}.$$

Isolating w^{l*} and substituting $w^{h*} = [\alpha/(1-\alpha)]s^* w^{l*}$ with $s^* = d_h/[q(1-d_l)]$:

$$w^{l*}\left[1-\delta(1-d_h)+\frac{\alpha}{1-\alpha}\delta d_h\right] = \underline{v}(1-\delta)\left[1-\delta(1-d_h)+\delta q(1-d_l)\right],$$

which gives the stated expression. The senior wage follows from $w^{h*} = [\alpha/(1-\alpha)]s^* w^{l*}$. ■

Proof of Proposition 5. *Part* (i). The span $s^* = d_h/[q(1-d_l)]$ is invariant to γ^h and γ^l. From $n^{h*} = z/[(\gamma^h)^\alpha(\gamma^l)^{1-\alpha}(s^*)^{1-\alpha}]$, an increase in either productivity reduces n^{h*} and hence $n^{l*} = s^* n^{h*}$. The wage expressions depend only on δ, d_h, d_l, q, α, and $\underline{v}$—not on γ^h or γ^l—so both wages and the wage ratio are unchanged.

Part (ii). An increase in q lowers s^*. For $n^{h*} = z/[(\gamma^h)^\alpha(\gamma^l)^{1-\alpha}(s^*)^{1-\alpha}]$, since s^* decreases and $1-\alpha > 0$, n^{h*} increases. For n^{l*}, write $n^{l*} = z(s^*)^\alpha/[(\gamma^h)^\alpha(\gamma^l)^{1-\alpha}]$; since $\alpha \in (0,1)$ and s^* decreases in q, n^{l*} decreases. The wage ratio $[\alpha/(1-\alpha)]s^*$ decreases because s^* does.

For the junior wage, the numerator of w^{l*} is increasing in q while the denominator is invariant to q, so w^{l*} increases. For the senior wage, $w^{h*} = [\alpha/(1-\alpha)][d_h/(q(1-d_l))]w^{l*}$. Taking the derivative with respect to q:

$$\frac{\partial}{\partial q}\left[\frac{1-\delta(1-d_h)+\delta q(1-d_l)}{q}\right] = \frac{-(1-\delta(1-d_h))}{q^2} < 0,$$

so w^{h*} decreases. ■

Proof of Proposition 6. *Part* (i). At $t = 1$, seniors are predetermined at n_0^{h*}. For a productivity shock, the capacity constraint requires $(\gamma_{AI}^h n_0^{h*})^\alpha(\gamma_{AI}^l n_1^l)^{1-\alpha} = z$. Equating with the pre-shock condition $(\gamma_0^h n_0^{h*})^\alpha(\gamma_0^l n_0^{l*})^{1-\alpha} = z$ gives $n_1^l = \left(\frac{\gamma_0^h}{\gamma_{AI}^h}\right)^{\alpha/(1-\alpha)}\left(\frac{\gamma_0^l}{\gamma_{AI}^l}\right) n_0^{l*} < n_0^{l*}$. For a learning shock, neither f nor its arguments change at $t = 1$ (i.e., $\gamma_{AI} = \gamma_0$), so $n_1^l = n_0^{l*}$.

Part (ii). For $t \geq 2$, the binding capacity constraint gives $n_t^l = (1/\gamma_{AI}^l)z^{1/(1-\alpha)}(\gamma_{AI}^h)^{-\alpha/(1-\alpha)}(n_t^h)^{-\alpha/(1-\alpha)}$. Substituting into the flow equation yields:

$$n_{t+1}^h = (1-d_h)n_t^h + \Omega(n_t^h)^{-\alpha/(1-\alpha)},$$

where $\Omega := q(1-d_l)(1/\gamma_{AI}^l)z^{1/(1-\alpha)}(\gamma_{AI}^h)^{-\alpha/(1-\alpha)} > 0$.

Define $\phi(n^h) := (1-d_h)n^h + \Omega(n^h)^{-\alpha/(1-\alpha)}$. The steady state satisfies $d_h n^{h*} = \Omega(n^{h*})^{-\alpha/(1-\alpha)}$, yielding $\Omega = d_h(n^{h*})^{1/(1-\alpha)}$. The derivative at steady state is:

$$\phi'(n^{h*}) = (1-d_h) - \frac{\alpha}{1-\alpha}\frac{\Omega}{(n^{h*})^{1/(1-\alpha)}} = (1-d_h) - \frac{\alpha}{1-\alpha}d_h = 1 - \frac{d_h}{1-\alpha}.$$

This determines the local dynamics: smooth monotonic convergence if $|\phi'(n^{h*})| < 1$ and

$\phi' \geq 0$ (i.e., $d_h \leq 1-\alpha$); damped oscillation if $\phi' \in (-1, 0)$ (i.e., $1-\alpha < d_h < 2(1-\alpha)$); unstable dynamics if $\phi' \leq -1$ (i.e., $d_h \geq 2(1-\alpha)$).

Wage dynamics. The wage ratio $w_t^h/w_t^l = [\alpha/(1-\alpha)]s_t$ follows from Lemma 3. Since s^* is invariant to productivity, the wage ratio returns to its pre-shock level following a productivity shock. Since s^* decreases in q, the wage ratio is permanently lower after a learning shock. ■

Proof of Lemma 5. Consider a firm maximizes the present discounted value of profits subject to the capacity constraint and the talent flow constraint. Let λ_t and μ_t denote the Lagrange multipliers for the capacity and talent constraints, respectively. The Lagrangian is:

$$\mathcal{L} = \sum_{t=0}^{\infty} \delta^t \left\{ \begin{aligned} & zy - w_t^h(n_{i,t}^{h,in} + n_{i,t}^{h,ex}) - w_t^l n_{i,t}^l \\ & + \lambda_{i,t}[\gamma^h n_{i,t}^{h,in} + (\gamma^h - \sigma) n_{i,t}^{h,ex} + \gamma^l n_{i,t}^l - z] \\ & + \mu_{i,t}[(1-d_h)(n_{i,t}^{h,in} + n_{i,t}^{h,ex}) + q(1-d_l)n_{i,t}^l - n_{i,t+1}^{h,in}] \end{aligned} \right\}.$$

In a stationary equilibrium, the wages and multipliers are constant ($w_t^h = w^h$, $w_t^l = w^l$, $\lambda_{i,t} = \lambda_i$, $\mu_{i,t} = \mu_i$).

i. We first prove that $n_{it}^{h,ex} = 0$ so that $n_{it}^l > 0$ and $n_{it}^{h,in} > 0$. Suppose, by contradiction, that there exists a stationary equilibrium where the aggregate mass of external seniors hired is strictly positive ($\int n_i^{h,ex} di > 0$).

(i) There exists a firm i that hires external seniors ($n_i^{h,ex} > 0$). For this firm, the talent constraint must be binding. Otherwise, that means the firm lets some internal seniors go away while hiring externally. Then firm i can increase the internal hiring and reduces external hiring in a way to keep the capacity constant. This would strictly lower the wage bill.

Now we know that, for this firm i, $n_i^{h,ex} > 0$ and $n_i^{h,in} > 0$. Consequently, the first-order conditions with respect to $n_i^{h,in}$ and $n_i^{h,ex}$ (which are strictly positive) are:

$$\mu_i = \delta[-w^h + \lambda_i \gamma^h + \mu_i(1-d_h)]. \tag{3}$$

$$w^h = \lambda_i(\gamma^h - \sigma) + \mu_i(1-d_h), \tag{4}$$

Substituting (4) into (3) yields $\mu_i = \delta\lambda_i\sigma$, which captures the present value of the specific human capital premium σ. We substitute $\mu_i = \delta\lambda_i\sigma$ back into the first-order condition for $n_{i,t}^{h,ex}$ and the condition for $n_{i,t}^l$ (which is nonnegative):

$$w^h = \lambda_i[\gamma^h - \sigma + \delta\sigma(1-d_h)],$$

$$w^l \geq \lambda_i[\gamma^l + \delta\sigma q(1-d_l)].$$

Taking the ratio of wages leads to the following upper bound (we show later wages are strictly positive):

$$\frac{w^h}{w^l} \leq \frac{\gamma^h - \sigma[1 - \delta(1 - d_h)]}{\gamma^l + \delta\sigma q(1 - d_l)}. \tag{5}$$

Since $\sigma > 0$, the numerator is strictly less than γ^h, and the denominator is strictly greater than γ^l. Thus, for any firm i hiring external seniors, $\frac{w^h}{w^l} < \frac{\gamma^h}{\gamma^l}$.

(ii) On the other hand, for external seniors to be hired, there must be a firm j releasing their internal talent. That is, firm j's talent constraint is slack. This implies that $\mu_j = 0$. First notice that firm j does not hire externally ($n_j^{h,ex} = 0$). If firm j hires externally, then it can always reduce the external hiring by ε and increase internal hiring by $\frac{\gamma^h - \sigma}{\gamma^h}\varepsilon$, as the talent constraint is slack. This keeps the capacity constant while strictly decreasing the wage bill.

Second, notice that firm j employs juniors ($n_j^l > 0$). If $n_j^l = 0$ and $n_j^{h,ex} = 0$, the talent constraint implies $n_j^{h,in} < (1 - d_h)n_j^{h,in}$, which is impossible. Moreover, the capacity constraint binds. If not, the firm could reduce costs by firing juniors and strictly raising the profit. This implies that $\lambda_j > 0$.

Because for firm j, $n_j^l > 0$ and $n_j^{h,ex} = 0$, the first-order conditions for firm j with respect to n_j^l and $n_j^{h,ex}$ are

$$w^l = \lambda_j \gamma^l,$$
$$w^h \geq \lambda_j \gamma^h.$$

Therefore, wages are strictly positive. Taking the ratio yields:

$$\frac{w^h}{w^l} \geq \frac{\gamma^h}{\gamma^l}. \tag{6}$$

Comparing (5) and (6), we have $\frac{w^h}{w^l} < \frac{\gamma^h}{\gamma^l} \leq \frac{w^h}{w^l}$, a contradiction. Thus, no firms hire external seniors in stationary equilibrium. Every firm hires juniors and hires seniors only from the ILM ($n_{it}^l > 0$ and $n_{it}^{h,in} > 0$).

ii. Next we prove that the capacity and talent constraints are binding. If the capacity constraint is slack, then each firm can reduce $n_{it}^{h,in}$ and n_{it}^l in proportion to satisfy the flow constraint while strictly reduces the wage bill. Second, suppose the talent constraint is slack for a positive mass of firms ($n_i^{h,in} < (1 - d_h)n_i^{h,in} + q(1 - d_l)n_i^l$). These firms release talent into the external market, implying the aggregate supply of external seniors is strictly positive. However, we have shown that the demand for external seniors is zero for all firms. This violates the market clearing condition. Thus, the talent constraint must bind:

$d_h n_i^{h,in} = q(1-d_l)n_i^l$. Because both capacity and talent constraints are binding, they uniquely pin down $n_t^{h,in}$ and n_t^l in the stationary equilibrium and every firm is symmetric. ■

Proof of Lemma 6. From Lemma 5, we know $n_t^l > 0$, $n_t^{h,in} > 0$, and $n_t^{h,ex} = 0$. Therefore, the first-order condition for junior workers n_t^l is:

$$\frac{\partial \mathcal{L}}{\partial n_t^l} = -w^l + \lambda_t \gamma^l + \mu_t q(1-d_l) = 0.$$

For internal seniors $n_{t+1}^{h,in}$, it is:

$$\frac{\partial \mathcal{L}}{\partial n_{t+1}^{h,in}} = -\mu_t + \delta\left[-w^h + \lambda_{t+1}\gamma^h + \mu_{t+1}(1-d_h)\right] = 0.$$

For external seniors $n_t^{h,ex}$, it is:

$$\frac{\partial \mathcal{L}}{\partial n_t^{h,ex}} = -w^h + \lambda_t(\gamma^h - \sigma) + \mu_t(1-d_h) = 0.$$

The condition holds with equality because of competitive equilibrium.

In the stationary symmetric equilibrium, multipliers are constant ($\lambda_t = \lambda, \mu_t = \mu$). Rearranging these equations yields:

$$\begin{aligned} w^h &= \lambda(\gamma^h - \sigma) + (\delta\lambda\sigma)(1-d_h) = \lambda[\gamma^h - \sigma + \delta\sigma(1-d_h)], \\ w^l &= \lambda\gamma^l + (\delta\lambda\sigma)q(1-d_l) = \lambda[\gamma^l + \delta\sigma q(1-d_l)]. \end{aligned}$$

The multiplier λ is positive because the capacity constraint binds. Dividing w^h by w^l eliminates λ and yields the result Γ. ■

Proof of the Corollary 1. This follows from Lemma 6. ■

Proof of Lemma 7. We first prove that, if there is a symmetric equilibrium involving external hiring, we can always construct a symmetric equilibrium that achieves the same profit for the firm without external hiring but strictly increases the firm's capacity.

Suppose that firms hire external seniors during the transition ($n_t^{h,ex} > 0$). In a symmetric equilibrium, the supply of external seniors corresponds to the seniors that firms fail to retain. Thus, a firm hiring $n_t^{h,ex}$ external seniors could explicitly choose to retain these workers instead.

Consider this alternative strategy: the firm sets $\tilde{n}_t^{h,in} = n_t^{h,in} + n_t^{h,ex}$ and $\tilde{n}_t^{h,ex} = 0$. This reallocation maintains the total senior workforce and the wage bill $w_t^h(n_t^{h,in} + n_t^{h,ex})$.

Furthermore, because the total number of seniors remains unchanged, the firm carries the exact same talent pool into the next period, satisfying all dynamic constraints. However, because internal seniors possess firm-specific capital while external seniors do not, total capacity increases. Therefore, it is without loss of generality to focus on the symmetric equilibrium with internal hiring only. We next prove that the slack capacity constraint implies that the original plan is suboptimal if the firm's optimal strategy requires the capacity constraint to bind.

To show that the capacity and talent constraints bind, we analyze the firm's dynamic optimization problem. The firm maximizes discounted profits subject to both constraints. Although firms hire only internal seniors in equilibrium, the availability of external hires determines the market wage w_t^h. Let λ_t be the Lagrange multiplier for the capacity constraint and μ_t be the multiplier for the talent constraint. The Lagrangian is:

$$\begin{aligned}\mathcal{L} = \sum_{t=0}^{\infty} \delta^t \Big\{ & zy - w_t^h n_t^{h,ex} - w_t^h n_t^{h,in} - w_t^l n_t^l \\ & + \lambda_t \left[(\gamma^h - \sigma) n_t^{h,ex} + \gamma^h n_t^{h,in} + \gamma^l n_t^l - z \right] \\ & + \mu_t \left[(1-d_h)(n_t^{h,in} + n_t^{h,ex}) + q(1-d_l) n_t^l - n_{t+1}^{h,in} \right] \Big\}.\end{aligned}$$

The first-order conditions for $n_t^{h,ex}$, n_t^l, and $n_{t+1}^{h,in}$ characterize the firm's optimal hiring:

$$\begin{aligned}-w_t^h + \lambda_t(\gamma^h - \sigma) + \mu_t(1-d_h) &= 0, \\ -w_t^l + \lambda_t \gamma^l + \mu_t q(1-d_l) &= 0, \\ \delta \left[-w_{t+1}^h + \lambda_{t+1}\gamma^h + \mu_{t+1}(1-d_h) \right] - \mu_t &= 0.\end{aligned}$$

These three conditions imply $\mu_t = \delta \lambda_{t+1} \sigma$ and

$$\begin{aligned}w_t^h &= \lambda_t(\gamma^h - \sigma) + \delta\sigma(1-d_h)\lambda_{t+1}, \\ w_t^l &= \lambda_t \gamma^l + \delta\sigma q(1-d_l)\lambda_{t+1}.\end{aligned}$$

That is, in a dynamic symmetric equilibrium, the no-arbitrage condition becomes $\frac{w_t^h}{w_t^l} = \frac{\lambda_t(\gamma^h-\sigma)+\delta\sigma(1-d_h)\lambda_{t+1}}{\lambda_t\gamma^l+\delta\sigma q(1-d_l)\lambda_{t+1}}$. In the linear economy with competitive labor markets, shadow prices adjust instantly to the new steady-state levels following a permanent shock: $\lambda_t = \lambda$. As a result, $\frac{w^h}{w^l} = \frac{(\gamma^h-\sigma)+\delta(1-d_h)\sigma}{\gamma^l+\delta q(1-d_l)\sigma}$.

Finally, we need to prove λ (and hence μ) is positive. This is done by solving the wages out from the no-arbitrage condition, workers' value equations, and the free-entry condition as before. The expressions of the wages can be easily calculated from the no-arbitrage condition and the value functions. Because the derived market wages are strictly positive, the shadow price λ (and consequently μ) must be strictly positive. By the Kuhn-Tucker conditions, $\lambda > 0$ and $\mu > 0$ imply that both constraints bind at the optimum. This concludes the proof. ■

We first prove a lemma characterizing the firm's value function.

Lemma A.1. *With firm-specific human capital, the firm's value function is linear in its internal senior stock:*

$$V(n^{h,in}) = \lambda\sigma\, n^{h,in} + \frac{z(y-\lambda)}{1-\delta},$$

where $\lambda = \frac{w^l}{\gamma^l+\delta\sigma q(1-d_l)}$ *is the shadow price of capacity.*

Proof of Lemma A.1. We conjecture a linear value function $V(n^{h,in}) = A + Bn^{h,in}$. Substituting this into the Bellman equation $V(n_t^{h,in}) = zy - w^h n_t^{h,in} - w^l n_t^l + \delta V(n_{t+1}^{h,in})$ and using the wage expressions from Lemma 6, we match coefficients.

First, the binding capacity constraint implies $n_t^l = (z - \gamma^h n_t^{h,in})/\gamma^l$. The transition law is $n_{t+1}^{h,in} = \phi n_t^{h,in} + C$, where ϕ and C are constants defined by the binding flow constraint. Matching the slope coefficients yields $B = \lambda\sigma$. Matching the constant terms yields $A = \frac{z(y-\lambda)}{1-\delta}$.

The expression for λ follows from the junior wage equation derived in the proof of Lemma 6, $w^l = \lambda[\gamma^l + \delta\sigma q(1-d_l)]$. ■

Proof of Proposition 7. For part (i), differentiating λ with respect to the parameters shows that $\frac{\partial\lambda}{\partial\gamma^h} < 0$, $\frac{\partial\lambda}{\partial\gamma^l} < 0$, and $\frac{\partial\lambda}{\partial q} < 0$. This implies the shadow cost of capacity falls with AI improvements.

Using the value function $V(n_t^{h,in}) = \lambda\sigma n_t^{h,in} + \frac{z(y-\lambda)}{1-\delta}$, we differentiate with respect to λ:

$$\frac{\partial V}{\partial\lambda} = \sigma n_t^{h,in} - \frac{z}{1-\delta}.$$

Since the firm's total capacity is z and internal seniors produce $\gamma^h > \sigma$, the total specific capital rent cannot exceed the total capacity value, so $\sigma n_t^{h,in} < z < \frac{z}{1-\delta}$. Thus, $\frac{\partial V}{\partial\lambda} < 0$. Because AI lowers λ, and lower λ increases V, the firm value increases instantly. Simultaneously, the marginal value of an incumbent is $\lambda\sigma$, which strictly falls as λ falls.

For part (ii), from Lemma A.1, $V(n_t^{h,in})$ is linear and increasing in $n_t^{h,in}$ (since $\lambda\sigma > 0$). From the transition analysis, we know $n_t^{h,in}$ can oscillate during the transition. Therefore, $V(n_t)$ can also oscillate. Specifically, if $n_t^{h,in} < n^{h,in*}$, then $V(n_t^{h,in}) < V(n^{h,in*})$. ■

Proof of Lemma 8. We first show that c^* strictly decreases following any positive AI shock. From Equation (1), $c^* = \frac{w^{l*}}{\gamma^l} + \frac{F}{z}$. Substituting the steady-state junior wage from Proposition 1 and letting $A \equiv 1 - \delta(1 - d_h) > 0$ and $B \equiv \delta q(1 - d_l) > 0$, we can write:

$$c^* = \underline{v}(1 - \delta)\frac{A + B}{\gamma^l A + \gamma^h B} + \frac{F}{z}.$$

The comparative statics follow by inspection. The denominator $\gamma^l A + \gamma^h B$ is strictly increasing in γ^h and in γ^l, so $\frac{\partial c^*}{\partial \gamma^h} < 0$ and $\frac{\partial c^*}{\partial \gamma^l} < 0$. For a learning shock, B is strictly increasing in q. Let $f(B) = \frac{A+B}{\gamma^l A + \gamma^h B}$. Differentiating with respect to B gives a numerator of $(\gamma^l A + \gamma^h B) - (A + B)\gamma^h = A(\gamma^l - \gamma^h) < 0$, so $\frac{\partial c^*}{\partial q} < 0$.

Because c^* falls in all three cases, the zero-profit condition requires the equilibrium price $y(Q^*)$ to fall. Since demand is downward sloping, Q^* strictly increases, and hence $M^* = Q^*/z$ strictly increases.

Next, we analyze the effect on aggregate employment. Let $\eta \equiv -\frac{y(Q^*)}{y'(Q^*)Q^*} > 0$ be the price elasticity of demand. Since $y(Q^*) = c^*$, differentiating yields $\frac{d \ln Q^*}{dx} = -\eta \frac{d \ln c^*}{dx} = \eta \frac{|dc^*/dx|}{c^*}$ for any parameter $x \in \{\gamma^h, \gamma^l, q\}$.

Aggregate senior employment is $N^{h*} = M^* n^{h*} = \frac{Q^*}{\gamma^l S^* + \gamma^h}$. Taking the natural logarithm gives:

$$\ln N^{h*} = \ln Q^* - \ln\left(\gamma^l S^* + \gamma^h\right).$$

Case 1: Productivity Shock ($\gamma \in \{\gamma^h, \gamma^l\}$).
The span $S^* = \frac{d_h}{q(1-d_l)}$ is independent of γ. Therefore, $\frac{d \ln\left(\gamma^l S^* + \gamma^h\right)}{d\gamma} > 0$. Differentiating aggregate senior employment with respect to γ:

$$\frac{d \ln N^{h*}}{d\gamma} = \eta \frac{|dc^*/d\gamma|}{c^*} - \frac{d \ln\left(\gamma^l S^* + \gamma^h\right)}{d\gamma}.$$

This expression is strictly positive if and only if $\eta > \frac{c^*}{|dc^*/d\gamma|} \frac{d \ln\left(\gamma^l S^* + \gamma^h\right)}{d\gamma} \equiv \bar{\eta}_\gamma$. Thus, aggregate senior employment rises if and only if demand is sufficiently elastic ($\eta > \bar{\eta}_\gamma$). Because $N^{l*} = S^* N^{h*}$ and S^* is constant, aggregate junior employment moves in the exact same direction, subject to the identical threshold.

Case 2: Learning Shock (q).
Differentiating aggregate senior employment with respect to q gives:

$$\frac{d \ln N^{h*}}{dq} = \eta \frac{|dc^*/dq|}{c^*} - \frac{d \ln\left(\gamma^l S^* + \gamma^h\right)}{dq}.$$

Because S^* is strictly decreasing in q, the term $\gamma^l S^* + \gamma^h$ strictly decreases, meaning

$\frac{d\ln(\gamma^l S^* + \gamma^h)}{dq} < 0$. Since both $\eta\frac{|dc^*/dq|}{c^*} > 0$ and $-\frac{d\ln(\gamma^l S^* + \gamma^h)}{dq} > 0$, we have $\frac{d\ln N^{h*}}{dq} > 0$ unconditionally. Therefore, aggregate senior employment unambiguously rises, regardless of the elasticity of demand.

For aggregate junior employment, $N^{l*} = S^* N^{h*}$. Taking the derivative:

$$\frac{d\ln N^{l*}}{dq} = \frac{d\ln S^*}{dq} + \frac{d\ln N^{h*}}{dq} = \frac{d\ln S^*}{dq} + \eta\frac{|dc^*/dq|}{c^*} - \frac{d\ln\left(\gamma^l S^* + \gamma^h\right)}{dq}.$$

Because $\frac{d\ln S^*}{dq} < 0$, the net effect depends on η. Aggregate junior employment rises if and only if η is sufficiently large to offset the drop in the span S^*. ■

Proof of Proposition 8. Part (i): At $t = 1$, the permanent AI shock changes the parameters. As shown in the baseline model, wages depend only on parameters and adjust immediately. The per-task cost therefore drops to c^*_{AI} at once, and the free-entry condition $y(M_t z) = c^*_{AI}$ implies $M_t = M^*_{AI}$ for all $t \geq 1$. By Lemma 8, $M^*_{AI} > M_0$.

Part (ii): At $t = 1$, the aggregate supply of seniors is fixed by the previous period's flow: $N^h_1 = M_0 n^{h*}_0 = N^{h*}_0$. With M^*_{AI} active firms, the binding aggregate capacity constraint gives

$$N^l_1 = \frac{M^*_{AI} z - \gamma^h_{AI} N^{h*}_0}{\gamma^l_{AI}}.$$

For a *learning shock*, productivities are unchanged ($\gamma^h_{AI} = \gamma^h_0$, $\gamma^l_{AI} = \gamma^l_0$), so $N^l_1 - N^{l*}_0 = (M^*_{AI} - M_0)z/\gamma^l_0 > 0$: junior hiring strictly increases.

For a *productivity shock*, M^*_{AI} is increasing in the elasticity of demand η. When η is large, M^*_{AI} is large, and the expansion term $M^*_{AI} z$ dominates, so $N^l_1 > N^{l*}_0$. When η is small, M^*_{AI} is close to M_0, the higher productivities γ^h_{AI} and γ^l_{AI} reduce the need for workers, and $N^l_1 < N^{l*}_0$.

Part (iii): For $t \geq 2$, the aggregate flow condition is $N^h_{t+1} = (1 - d_h)N^h_t + q_{AI}(1 - d_l)N^l_t$. Substituting $N^l_t = (M^*_{AI} z - \gamma^h_{AI} N^h_t)/\gamma^l_{AI}$ gives the linear difference equation

$$N^h_{t+1} = \underbrace{\left[1 - d_h - q_{AI}(1 - d_l)\frac{\gamma^h_{AI}}{\gamma^l_{AI}}\right]}_{\equiv \phi} N^h_t + \frac{q_{AI}(1 - d_l)M^*_{AI} z}{\gamma^l_{AI}}.$$

The dynamics oscillate if and only if $\phi < 0$, which holds when $q_{AI}\frac{\gamma^h_{AI}}{\gamma^l_{AI}} > \frac{1-d_h}{1-d_l}$.

The starting phase of the transition depends on the initial deviation $N^{h*}_0 - N^{h*}_{AI}$. By Lemma 8, under a learning shock, $N^{h*}_0 < N^{h*}_{AI}$ always holds. Under a productivity shock, $N^{h*}_0 < N^{h*}_{AI}$ holds if and only if demand is sufficiently elastic. Whenever $N^{h*}_0 < N^{h*}_{AI}$, the economy enters the transition with a relative shortage of seniors. The initial junior hiring boom at $t = 1$ produces a wave of new seniors at $t = 2$ that overshoots the steady state ($N_2^h > N_{AI}^{h*}$), meaning the oscillation begins from above. Conversely, when $N_0^{h*} > N_{AI}^{h*}$ (which only occurs under a productivity shock with inelastic demand), the economy starts with an oversupply of seniors, junior hiring contracts at $t = 1$, and the oscillation begins from below.

■

Proof of Proposition 9. We first show that the steady state is unstable, then characterize the two-period cycle and its wages.

Instability. From the task constraint (Lemma 1), $N_t^l = (Z - \gamma^h N_t^h)/\gamma^l$. Substituting into the flow equation (FC):

$$N_{t+1}^h = \underbrace{\left(1 - d_h - q(1 - d_l)\frac{\gamma^h}{\gamma^l}\right)}_{\equiv \phi} N_t^h + \frac{q(1 - d_l)}{\gamma^l} Z.$$

Writing $N_t^h = N^{h*} + \varepsilon_t$ gives $\varepsilon_{t+1} = \phi\, \varepsilon_t$. Since $\phi < 1$ always holds, stability requires $\phi > -1$, i.e., $q\frac{\gamma^h}{\gamma^l} < \frac{2-d_h}{1-d_l}$. When this condition fails, $\phi \leq -1$, and perturbations grow weakly in magnitude while alternating in sign.

Reaching the corner. Since $|\phi| \geq 1$ and $\phi < 0$, the sequence $\{N_t^h\}$ oscillates with non-decreasing amplitude around N^{h*}. There exists a period T such that $\gamma^h N_T^h \geq Z$, at which point $N_T^l = 0$.

The two-period cycle. *Oversupply state.* The firm needs only Z/γ^h seniors to complete all tasks. However, the total mass of seniors at the beginning of the oversupply state is $N_O^h = (1 - d_h)^2 \frac{Z}{\gamma^h} + q(1 - d_l)\frac{d_h Z}{\gamma^l}$. Since $N_O^h > Z/\gamma^h$, supply exceeds demand. Competition drives the senior wage to $w_{\min}$. No juniors are hired. Employed seniors stay in the firm. Unemployed seniors lose their skills and receive their outside option. Thus, a senior entering the oversupply state is employed with probability $P = \frac{Z/\gamma^h}{N_O^h} < 1$.

Shortage state. Of the Z/γ^h employed seniors from the oversupply state, a fraction d_h exits, leaving $(1 - d_h)Z/\gamma^h$ seniors who can complete $(1 - d_h)Z$ tasks. The remaining $d_h Z$ tasks require $d_h Z/\gamma^l$ juniors. Both types are employed, so by Lemma 2, $w_S^h/w_S^l = \gamma^h/\gamma^l$.

Persistence of the cycle. From the shortage state, the next-period senior stock is

$$N^h = (1 - d_h)\frac{(1 - d_h)Z}{\gamma^h} + q(1 - d_l)\frac{d_h Z}{\gamma^l},$$

where the first term counts surviving seniors and the second counts newly promoted juniors. (No juniors were hired in the preceding oversupply state, so that cohort contributes zero.)

The task capacity of these seniors is

$$\gamma^h N^h = (1-d_h)^2 Z + q(1-d_l)\frac{\gamma^h}{\gamma^l}\, d_h\, Z.$$

This exceeds Z if and only if $q\frac{\gamma^h}{\gamma^l} \geq \frac{2-d_h}{1-d_l}$, which holds by assumption. The economy returns to the oversupply state, and the cycle repeats. Note that N^h matches exactly N^h_O.

Wages in the shortage state. Let $\bar{v} = \underline{v}/(1-\delta)$ denote the present value of the outside option. Let v^h_O and v^h_S denote the expected continuation values of a senior entering the oversupply and shortage states, respectively. Incorporating the employment probability P, these values satisfy:

$$v^h_O = P\left[w_{\min} + \delta(1-d_h)\, v^h_S + \delta\, d_h\, \bar{v}\right] + (1-P)\bar{v}, \tag{7}$$

$$v^h_S = w^h_S + \delta(1-d_h)\, v^h_O + \delta\, d_h\, \bar{v}. \tag{8}$$

Solving (7)–(8) for v^h_O:

$$v^h_O = \frac{P w_{\min} + P\delta(1-d_h)\, w^h_S + \{P\delta\, d_h\,[1+\delta(1-d_h)] + 1 - P\}\,\bar{v}}{1 - P\delta^2(1-d_h)^2}. \tag{9}$$

For a junior in the shortage state, a junior who is not promoted enters the oversupply state, where no juniors are hired, and returns to the junior pool in the following shortage state. The continuation value of an unpromoted junior is therefore $\underline{v} + \delta\, v^l_S$. By the free-entry condition, $v^l_S = \bar{v}$, so this continuation value equals $\underline{v} + \delta\,\bar{v} = \bar{v}$. The junior value function in the shortage state is

$$\bar{v} = w^l_S + \delta(1-d_l)\left[q\, v^h_O + (1-q)\,\bar{v}\right] + \delta\, d_l\, \bar{v}.$$

Rearranging:

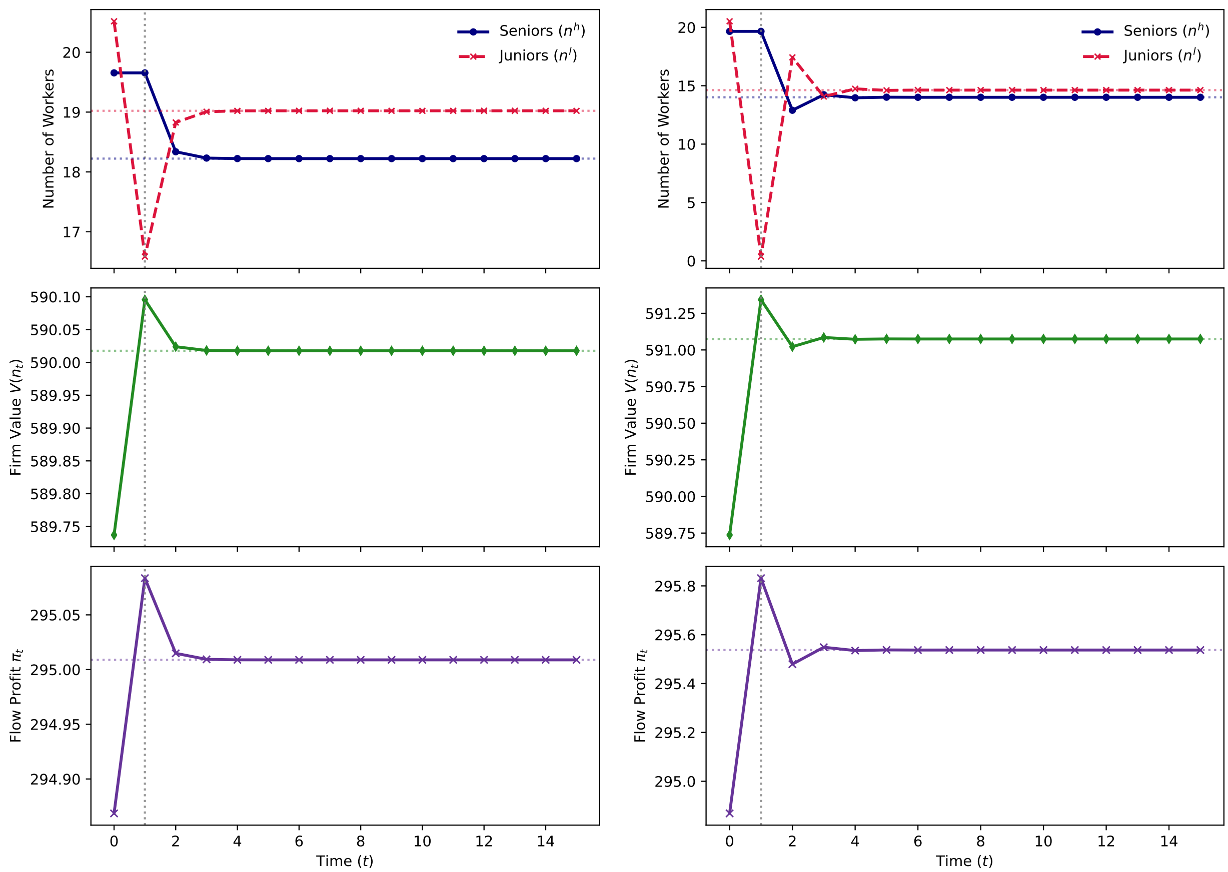
$$\bar{v}\,[1-\delta+\delta(1-d_l)\,q] = w^l_S + \delta(1-d_l)\, q\, v^h_O.$$

Substituting (9) with $w^h_S = (\gamma^h/\gamma^l)\, w^l_S$, using $\bar{v} = \underline{v}/(1-\delta)$, and solving for w^l_S yields:

$$w^l_S = \frac{\underline{v}\,\{1 - P\delta^2(1-d_h)^2 + P\delta q(1-d_l)[1+\delta(1-d_h)]\} - P\delta\, q(1-d_l)\, w_{\min}}{1 - P\delta^2(1-d_h)^2 + P\delta^2\, q(1-d_l)(1-d_h)\,\frac{\gamma^h}{\gamma^l}}.$$

■

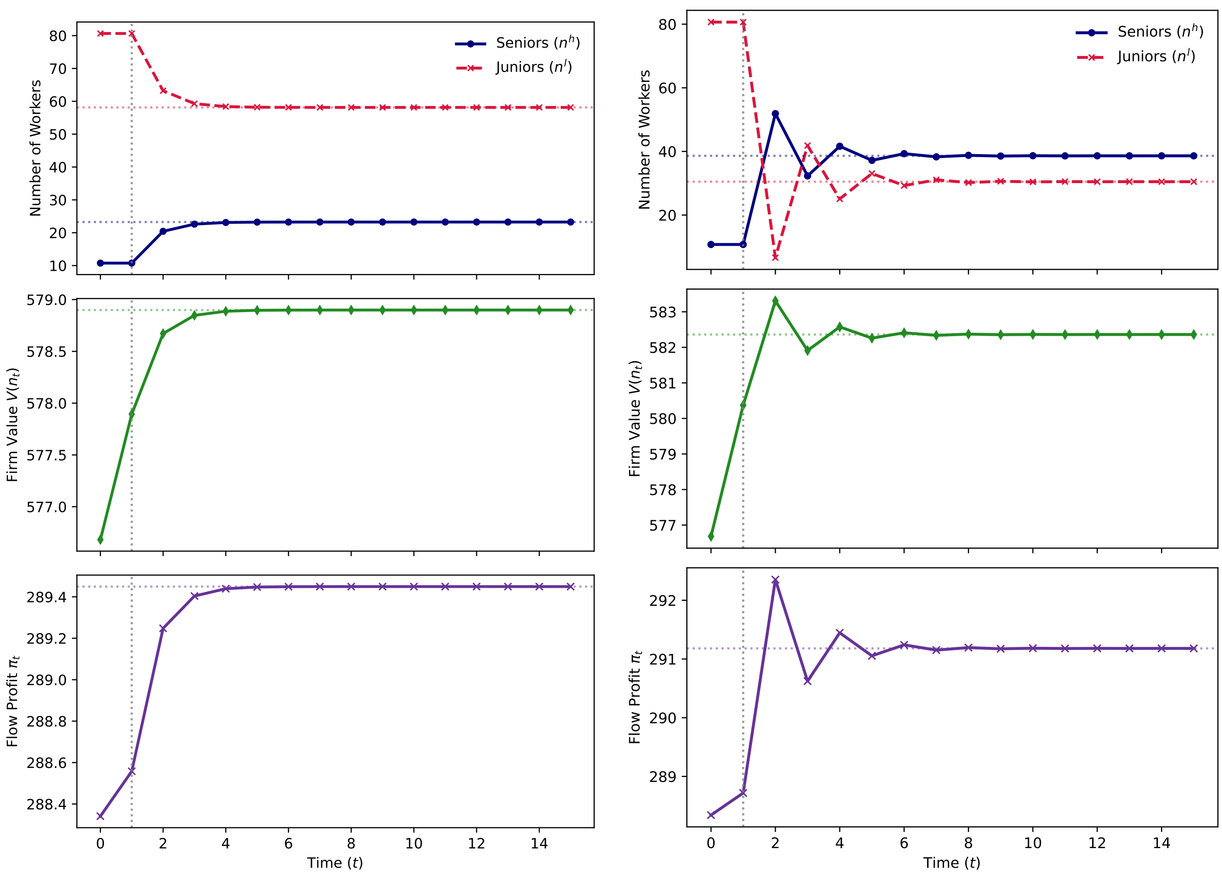
B Figures



(**a**) Smooth Transition.

(**b**) Oscillation.

Figure 7: Simulation of the Transition of Discounted Firm Value and Current Profit. Parameters: $y = 3$, $\delta = 0.5$, $\sigma = 0.99$, $\underline{v} = 0.25$, and the remaining parameters are same as Figure **??**.



(**a**) Smooth Transition.

(**b**) Oscillation.

Figure 8: Simulation of the Transition of Discounted Firm Value and Current Profit. Parameters: $y = 3$, $\delta = 0.5$, $\sigma = 0.7$, $\underline{v} = 0.25$, and the remaining parameters are same as Figure **??**.