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

Guido Friebel
Goethe University Frankfurt & CEPR

Yao Huang
The University of Hong Kong

Jin Li
The University of Hong Kong

Soumitra Shukla
The Burning Glass Institute

Andrew Zhang
Goethe University Frankfurt

August 19, 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, junior hiring contracts sharply and flattens the hierarchy; when the pre-shock span is near one, the pyramid temporarily turns into a diamond. In the transition, the span overshoots its new level and oscillates 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

*We are grateful to Alex Imas, Anders Humlum, and Luis Garicano for helpful comments and to seminar participants at the NBER Organizational Economics Spring 2026 meeting, IOEA, and CEPR. The authors can be reached at: Guido Friebel (gfriebel@wiwi.uni-frankfurt.de), Yao Huang (yaohuang@connect.hku.hk), Jin Li (jli1@hku.hk), Soumitra Shukla (sshukla@burningglassinstitute.org), and Andrew Zhang (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 hence 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?

The Model and what it delivers. In each of many firms in the economy, a CEO allocates a fixed set of tasks to senior and junior workers. 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. Because a senior can only be produced by a junior hired a period earlier, the size of today's junior cohort determines tomorrow's supply of seniors. Firms hire period by period; the link between the two layers therefore runs through the market rather than through any firm's forward planning. 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 in the long run. Perhaps surprisingly, this holds even though the short-run disruption to junior hiring can be severe. In a nutshell, pyramids remain pyramids. The wages of seniors and juniors do shift, however. Seniors earn a premium that reflects how much more productive they are, so the senior wage rises when seniors become more productive and falls when juniors become more productive. The junior wage moves in the opposite direction, because juniors are forward-looking and anticipate the future senior wage. A higher senior wage lets the firm pay juniors less, while a lower senior wage forces it to pay them more.

Second, if the main effect of AI is on how fast juniors learn and can become seniors, a permanent structural change occurs. Pyramids can become diamonds. To see why, consider the long-run steady state. The number of seniors who leave each period equals the number of juniors who learn the skills and become new seniors. It is this learning rate, the flow rate from juniors to seniors, that governs the ratio of juniors to seniors, and thus the firm's shape. When learning is slow, many juniors are needed for each new senior, so the firm is more

pyramidal. When AI speeds up learning, fewer juniors are needed, so the firm is closer to a diamond. This also explains why productivity shocks leave the shape unchanged, as they do not change how fast juniors become seniors. The learning shock is the only AI shock that reshapes the hierarchy in the long run. For the wages, both fall. Faster learning lets juniors become seniors sooner, so the firm pays juniors less today, and the senior wage falls by the same proportion, since relative productivity between juniors and seniors stays unchanged.

Third, we characterize the transition dynamics following both types of shock. A productivity shock cuts junior hiring sharply on impact, narrowing the junior layer and flattening the hierarchy. The transition need not be smooth. The span oscillates around the new steady state when the senior–junior productivity gap is sufficiently wide, or junior learning sufficiently fast. The structure is that of a cobweb Freeman (1975), with two differences: the lag sits inside the firm's own pipeline rather than in workers' schooling decisions, and the cycle survives fully forward-looking wages rather than resting on expectational error. Oscillation is not the result of a forecasting error but represents the equilibrium quantity path, delayed by exactly the length of the pipeline. Depending on the size of the shock and the pre-shock span, these oscillations may mean that the hierarchy crosses the pyramid–diamond boundary or keep the pyramid structure. Section 6 develops all of this.

Why should an optimizing firm overshoot rather than glide to its new steady state? The answer lies in the interaction of three features. First, there is a lag: a junior hired today cannot fill a senior slot until next period, so production and the filling of the talent pipeline are decoupled in time. Second, when seniors are far more productive than juniors, replacing a single missing senior requires hiring many juniors. Third, when these juniors learn fast, a large share of that cohort matures into seniors together the following period. Putting these forces together, the initial hiring cutback leaves the firm short of seniors next period, so it must hire a large junior cohort to make up the deficit—larger, the wider the productivity gap. This large cohort, when learning is fast, then matures into an equally large wave of seniors, which now exceeds what the firm needs, triggering another recruiting cutback. The swings are therefore most violent precisely when AI widens the senior–junior productivity gap or accelerates junior learning.

These swings have costs. 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.

A caveat is worth noting. The oscillation depends on juniors being available to hire when

firms' pipelines are rebuilt. If hiring cutbacks persist long enough that potential juniors move elsewhere or never enter the career track, the rebuilding cohort never materializes, and the cycle weakens rather than repeats. The oscillation, in other words, requires firms to be able to refill their junior intake fast enough to keep the loop alive.

We extend the model in a number of ways, not only to investigate the robustness, but also to provide structure for the empirical observations presented in Section 8. A theoretically particularly relevant extension concerns 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 is under stress, 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). In our model firms hire period by period, and contracting the junior layer is individually optimal at every date. Yet when human capital is firm-specific, every period in which the pipeline runs short forfeits rents that an external hire cannot restore. The model therefore identifies a cost of pipeline volatility without giving the firm an instrument against it. Whether a CEO would in practice smooth junior intake depends on frictions we do not model, such as adjustment costs or limits on how fast a junior cohort can be recruited; we return to this in Section 9.

Related literature. 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

[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).

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 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. What has no counterpart in those frameworks is the stock-flow relation between the two layers: today's junior cohort is tomorrow's senior stock, and it is this relation, not any individual firm's planning problem, that pins down the long-run shape of the hierarchy and generates its transition path.

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. This basic production function is not without charm because it is grounded in the emerging empirical literature showing that AI allows more tasks to be completed in a given time and leads to faster learning (Noy and Zhang, 2023; Brynjolfsson, Li and Raymond, 2025; Chen et al., 2025; Humlum and Vestergaard, 2025). Beyond labor demand, reliance on AI introduces risks to human capital formation and creative diversity.

We consequently investigate whether 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.

Empirical observations. We begin the paper by providing some evidence on how junior and senior hiring and postings are currently changing, using data from the Burning Glass Institute (BGI). In Section 8, we then dig deeper into these data to mirror the propositions of the model with the data.

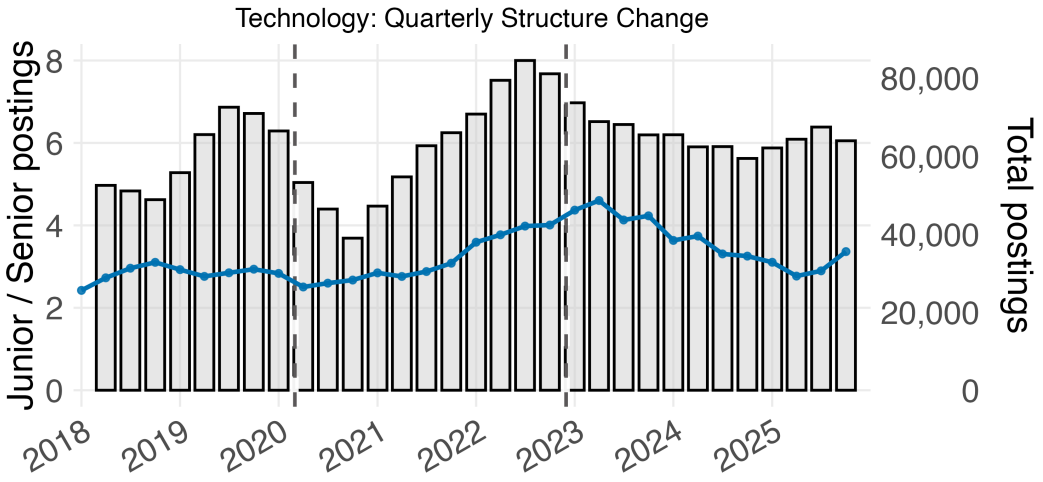
The exercise is descriptive: none of the figures identifies a causal effect of AI on hierarchies. Three facts stand out. First, the senior layer is increasingly held together by incumbents rather than replenished through the pipeline. The retained share rises in every focus sector after 2022, while internal promotions and both external-arrival margins fall. Second, in all sectors that we looked at, the junior/senior span narrows and the senior wage premium compresses at the same time. The compression comes from the bottom, as median junior real wages rise by four to nine percent while senior wages move little. Third, the contraction in junior postings begins immediately after 2022Q4 rather than with a lag.

No single shock in the baseline model produces this combination. Richer production settings, however, deliver what the baseline cannot. Under Cobb-Douglas production, a learning shock narrows the span, compresses the wage ratio, raises the junior wage, and lowers the senior wage: all four observed margins, with the correct signs. With firm-specific human capital, a learning shock again narrows the span and compresses the premium, and it additionally accounts for the shift toward retention, since internal seniors carry rents that external hires cannot replace. Either of these models delivers the joint pattern on its own. Hence, beyond robustness these versions of the model are the ones in line with the data. Adding the immediate timing of the junior contraction, the natural reading is a joint shock: a productivity component that cuts junior hiring on impact, and a learning component operating through complementarity or firm-specific capital.

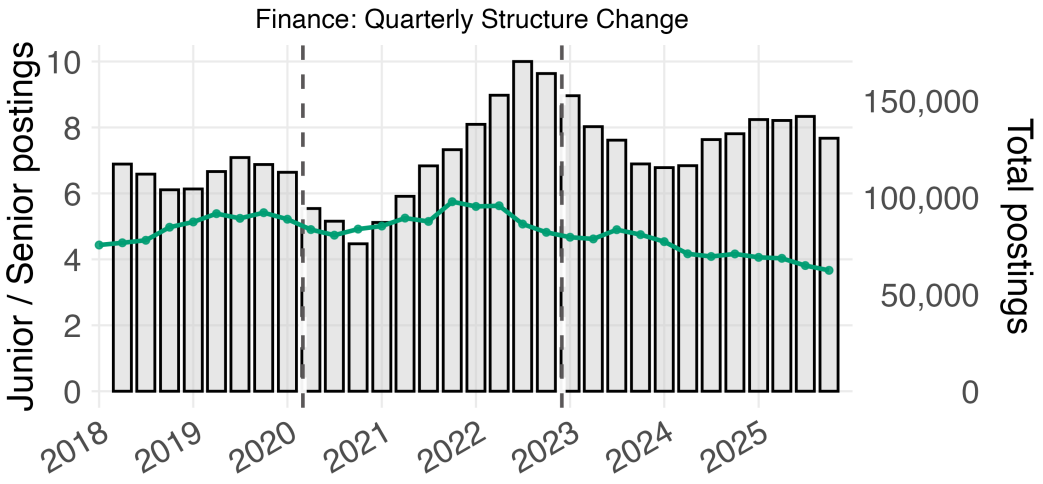
2 Junior and Senior Hiring and Positions

We start with advertised hiring data from the Burning Glass Institute (BGI). These data show the decline in junior relative to senior postings after the release of ChatGPT. Figure 1 plots the ratio of junior to senior vacancies (line) against total vacancies (bars). In the technology industry, junior vacancies rise relative to senior vacancies through the

run-up to ChatGPT and then fall (except for a slight increase in 2025Q4). In finance, the junior-to-senior ratio is flatter before 2022 but also drifts down afterward.



(**a**) Technology



(**b**) Finance

Figure 1: *BGI job postings.* Ratio of junior to senior job vacancies (line, left axis) and total vacancies (bars, right axis) in the technology and finance sectors, 2018–2025, shown as a three-quarter moving average. The junior/senior line is the mean firm-level ratio among firms with both junior and senior postings in a quarter, not the pooled aggregate ratio of all junior postings to all senior postings. Technology is NAICS sector 51 (Information) and finance is NAICS sector 52 (Finance and Insurance). Vacancies are from Burning Glass Institute company-source U.S. job postings, with junior and senior vacancies defined by their stated experience requirements. Dashed vertical lines mark the onset of COVID-19 (2020Q1) and the public release of ChatGPT (2022Q4). Construction is described in Appendix C.1.

The vacancy data measure advertised hiring, not the stock of workers inside firms. As such, they do not show whether the realized hierarchy inside firms is changing. Figure 2

turns to BGI worker profiles, which are constructed from LinkedIn profile histories processed by BGI. The figure shows the share of firms whose senior-titled workers outnumber junior-titled workers. This share rises after 2022, most visibly from 59% to 65% in information/technology and from 52% to 59% in finance. While this figure does not provide a causal link between the release of ChatGPT and the change in the firm hierarchy, the figure helps motivate a key channel the paper wants to model, namely that firms become more top-heavy. Appendix Figure D.1 shows that the same movement toward a more senior-heavy hierarchy is visible for the typical large firm, not only in the size-weighted aggregate.

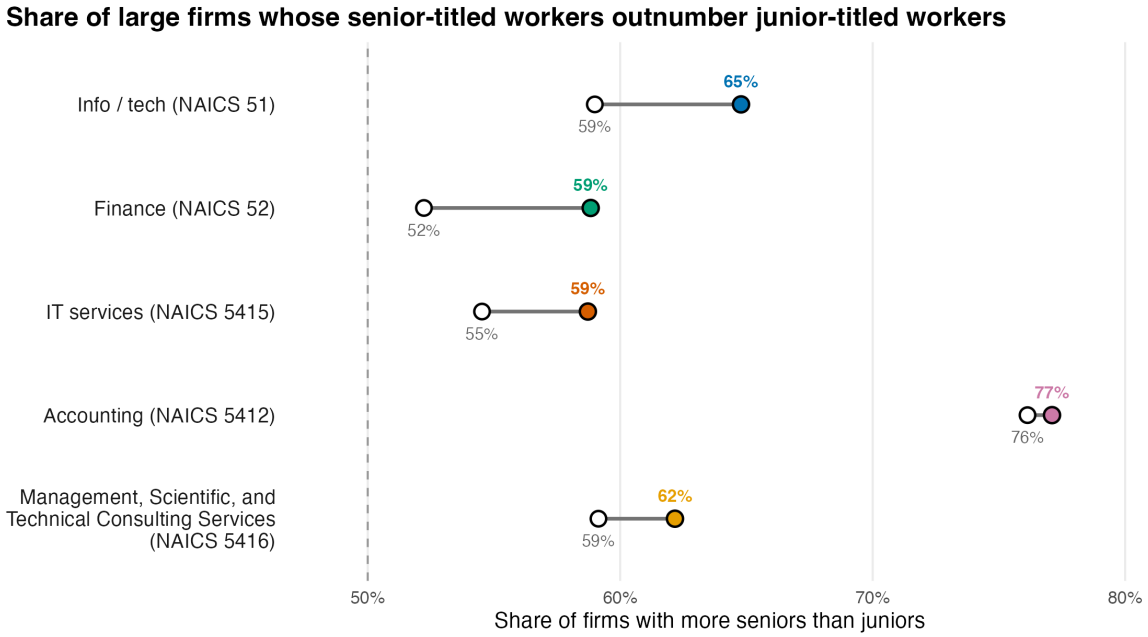


Figure 2: *BGI worker profiles.* Share of large firms whose senior-titled workers outnumber junior-titled workers. Open circles show the pre-ChatGPT average (2021–2022) and filled circles show the post-ChatGPT average (2023–2025), by sector. A firm is included in a sector-year if it has at least 100 worker-profile person-years. Workers sit on a title-rank ladder: R0 (interns and trainees) and R1 (entry-level, unleveled individual contributors) are the junior ranks, while R2 (senior individual contributors—senior, lead, principal, or staff titles) and above, up through managers (R3), directors (R4), vice presidents (R5), and executives (R6), are senior. Ranks come from a rule-based classifier that scans each worker's standardized job title for seniority keywords (*intern*, *senior*, *principal*, *director*, *chief*, and so on) and assigns the most senior match. Construction is described in Appendix C.2.

Together, Figures 1 and 2 help motivate our main analysis that models whether a thinner junior layer is a pause or a pipeline shift, how firms rebuild seniors, or which cohorts bear the adjustment. Sections 3–7 develop these ideas theoretically. Section 8 then returns to the BGI worker profiles data to measure the corresponding objects and check if they are consistent with the model's predictions.

3 Model Setup

In this section, we develop a competitive labor-market 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 $n^l_{i,t} \geq 0$ and $n^h_{i,t} \geq 0$ 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 $\{w^l_t, w^h_t\}_{t=0}^{\infty}$ denote the sequence of market wages. Given these wages, we can write the present discounted values of being a junior and a senior, v^l_t and v^h_t, at the beginning of period t as:

$$v^l_t = w^l_t + \delta(1 - d_l)\left[\tilde{q}_{i,t} v^h_{t+1} + (1 - \tilde{q}_{i,t}) v^l_{t+1}\right] + \delta d_l \underline{v}, \tag{VF^l}$$

$$v^h_t = w^h_t + \delta(1 - d_h) v^h_{t+1} + \delta d_h \underline{v}. \tag{VF^h}$$

Timing. Figure 3 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.

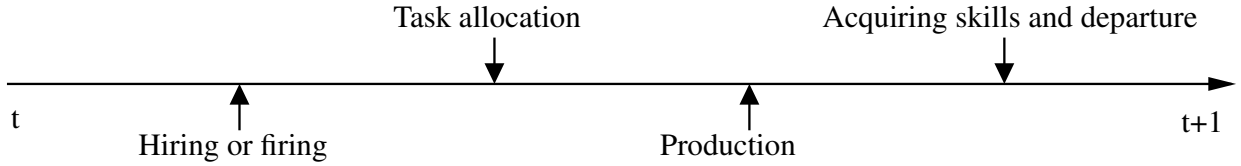


Figure 3: Stage Game Timeline

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. When several workforce choices are equally profitable for a given task allocation, the firm chooses the one that uses the fewest juniors and seniors.

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:

$$N_{t+1}^h = (1-d_h)N_t^h + (1-d_l)\int_i \tilde{q}_{i,t} n_{i,t}^l \, di. \tag{FC}$$

Equilibrium. A competitive labor-market equilibrium consists of sequences of wages $\{w_t^l, w_t^h\}_{t=0}^{\infty}$, firm policies $\{n_{i,t}^l, n_{i,t}^h, z_{i,t}^l, z_{i,t}^h\}_{t=0}^{\infty}$, worker values $\{v_t^l, v_t^h\}_{t=0}^{\infty}$, and aggregate senior labor supply $\{N_t^h\}_{t=0}^{\infty}$ 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).

Assumptions The setup imposes four restrictions. First, production is linear and the task load is fixed. Section 7.1 embeds the framework in a general constant-returns technology, including Cobb-Douglas, in which the two types may be complements and firm scale is endogenous. Second, internally promoted and externally hired seniors are equally productive. Section 7.2 introduces firm-specific human capital and shows that the pipeline then has no external substitute. Third, seniors and juniors perform the same task and differ only in per-worker capacity. An assignment formulation in the spirit of Gibbons and Waldman (1999b), with a routine and a complex task, is equivalent to this one, as we show in Section 7.6. Fourth, we abstract from firing costs and promotion incentives, and we do not analyze them anywhere in the paper. We discuss this in Section 9. Section 7 also relaxes the fixed mass of firms, the constant learning rate, and the stability condition. We keep the baseline deliberately simple because it permits a full characterization of the employment and wage dynamics of both layers.

4 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 unused labor capacity weakly lowers the firm's wage bill. When it leaves profits unchanged, the tie-breaking convention selects the no-slack employment choice.

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* $n^l_{i,t} > 0$)*, then their wage costs per unit of task capacity are equal:*

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

Lemma 2 describes a no-arbitrage condition. We will refer to the ratio of senior v. junior wages as the senior wage premium, and use the term in ratio form throughout rather than as a wage difference. Lemma 2 shows that the senior wage premium is fixed by relative productivity alone, and moves neither with the firm's structure nor along the transition. This is a consequence of perfect substitutability. Under the general production technology of Section 7.1 the premium instead depends on the span, and with firm-specific human capital (Section 7.2) it depends in addition on the capital that juniors accumulate while they learn. Section 8 will then measure the premium as the ratio of median senior to median junior real wages from the BGI data.

5 Long-Run Outcome

We now characterize the long-run behavior of the economy. We begin by defining a steady state and then establish its existence and uniqueness.

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:

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

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)**.** *The economy has a unique aggregate 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 characterizes the unique aggregate 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 7.2, 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 a 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 (γ^l and γ^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 illustrates the long-run effects of AI on employment, hierarchical structure and wage gap. Both productivity shocks change the employment and wage gap, while the hierarchical structure remains the same. A learning shock has impacts on all three aspects.

First, suppose there is an increase in senior productivity γ^h. Hence, each senior can produce more output. Given the fixed number of total tasks, the production target can be finished with fewer workers. The span S^* is fixed in the long run, as it is only affected by flow rates (d_h, d_l and q). Hence, the senior and junior employment will decrease proportionally. Lemma 2 has shown that the wage ratio equals the productivity ratio. This implies that the increase in senior productivity enlarges the wage gap. However, the senior productivity shock does not simply push up the senior wage. Anticipating a wage rise in the senior position, junior workers are willing to accept a lower wage at the beginning. In the long-run equilibrium, the expected value of junior positions is fixed by the outside option. Therefore, the increase in senior wage makes it possible for firms to backload the compensation more aggressively.

Next, suppose there is an increase in junior productivity γ^l. Hence, each junior can produce more output. For the same reason, senior and junior employment will drop proportionally in the long run. Different from the senior productivity shock, now the productivity gap narrows, leading to a narrower wage gap. Expecting a drop in the senior wage in the future, junior workers would require a higher current wage, making it more difficult for firms to backload compensation.

Finally, suppose there is an increase in learning rate q. Hence, junior workers get promoted faster after acquiring the necessary skills. This accelerates the talent supply to the senior layer. Firms require fewer junior workers to maintain the talent pipeline in the long run, leading to a drop in the span. This is the only AI shock in our baseline model that permanently reshapes the hierarchical structure, making it more diamond-like. Anticipating that they can get promoted faster with a higher q, junior workers would be willing to accept a lower current junior wage. From the no-arbitrage condition in Lemma 2, the senior wage also falls in proportion.

It is useful to contrast our model with a benchmark in which the two layers operate

independently. In such a benchmark, where there are 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 that 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.

6 AI Shocks and Transition Dynamics

Section 5 showed which shocks change the long-run shape of the hierarchy and how all three change wages. We now turn to the path between steady states. Two features of the model simplify our analysis on the transition dynamics. First, hiring in any period is a static optimum. Since the workforce can be adjusted at no cost, firms' optimization problem is separable across periods, as shown in Section 4. Firms therefore do not hire juniors today in anticipation of tomorrow's senior needs. Second, wages play no role during the transition. The no-arbitrage and free-entry conditions pin down wages without reference to current employment, so wages jump to their new steady-state levels when the shock hits and stay there.[3] To understand the transition process, we look into AI's initial impact on hierarchical structure and the way firms converge to steady state.

AI's initial impact on employment falls on junior hiring. Consider a productivity shock. The firm produces in period $t = 1$ with the senior stock inherited from the previous period. If seniors become more productive, they leave less work for the junior layer; if juniors become more productive, fewer of them are needed for the work that remains. Either way, all firms cut back on junior hiring or lay off juniors: junior employment and the span fall immediately, and the pyramid moves toward a diamond. A learning shock, by contrast, affects no one's current productivity, so employment does not change at $t = 1$. Its effect arrives one period later, when the first cohort of faster-learning juniors reaches the senior layer.

After the initial impact, both productivity and learning shocks trigger the same adjustment process. Substituting the binding capacity constraints (Lemma 1) into the flow condition (FC) shows that the senior stock follows

$$N^h_{t+1} - N^{h*}_{AI} = \zeta \left(N^h_t - N^{h*}_{AI} \right), \qquad t \geq 1, \tag{1}$$

where N^{h*}_{AI} is the new steady-state senior stock and the *transition coefficient*

$$\zeta \equiv (1 - d_h) - q(1 - d_l)\frac{\gamma^h}{\gamma^l}$$

[3] This is a consequence of perfect substitutability. Under the general production technology of Section 7.1, wages move with the span along the transition; see Appendix B.1.

is evaluated at the post-shock parameters.

The transition coefficient measures the net effect of an extra senior today on the senior stock tomorrow: retention minus pipeline crowding-out. The retention part is $1 - d_h$: an extra senior remains a senior with this probability. The crowding-out part is $q(1 - d_l)\gamma^h/\gamma^l$: an extra senior completes the work of γ^h/γ^l juniors, so the firm hires that many fewer juniors, and a fraction $q(1 - d_l)$ of these missing juniors would have become seniors. Because some seniors depart ($d_h > 0$), ζ is always below one. We assume throughout that $\zeta > -1$, which is the stability condition $q\gamma^h/\gamma^l < (2 - d_h)/(1 - d_l)$. Under this condition, any deviation from the steady state will eventually vanish. Section 7.5 studies rapid learning, where $\zeta < -1$ and the oscillation becomes a permanent cycle.

We restrict attention to shocks for which junior employment stays positive along the transition.[4] The following proposition characterizes the transition.

Proposition 3 (Transition Dynamics)**.** *Suppose that, at the beginning of period $t = 1$, AI permanently raises productivity (γ^h, γ^l, or both) or the learning rate q, and evaluate ζ at the post-shock parameters. Then:*

(i) Impact. *Under a productivity shock, the senior stock is unchanged at $t = 1$ and junior employment falls. Under a learning shock, employment is unchanged at $t = 1$, and the senior stock rises at $t = 2$.*

(ii) Convergence. *For $t \geq 1$, the deviation of the senior stock from its new steady-state level is multiplied by ζ each period, and junior employment and the span deviate in the opposite direction. If $\zeta < 0$, the transition oscillates: the senior stock alternates above and below its new steady-state level, while junior employment and the span alternate on the opposite side. If $\zeta \geq 0$, the transition is monotonic. In both cases, the deviations shrink by the factor $|\zeta|$ each period.*

Proposition 3 shows that the type of shock matters only for the initial impact while the transition path is characterized by the coefficient ζ. A productivity shock lowers the senior stock the firm needs in the long run, so the path starts from a senior surplus, and it starts immediately, with the hiring freeze at $t = 1$. A learning shock raises the senior stock the new steady state supports, so the path starts from a senior shortage, and it starts one period late, when the first faster-trained cohort arrives. The sign of ζ decides the shape of the path. Oscillation happens when $\zeta < 0$ while a smooth transition happens when $\zeta \geq 0$. The transition speed is decided by the absolute value $|\zeta|$.

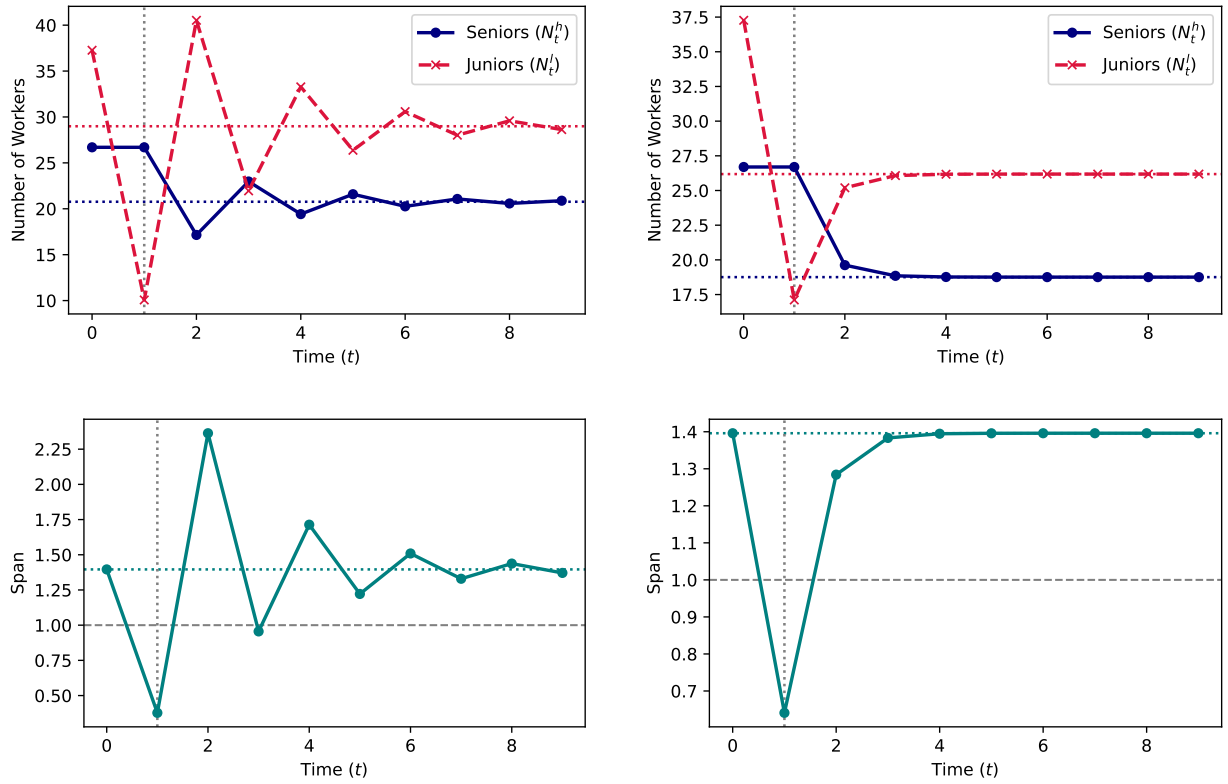
[4] It suffices that $\gamma^h_{AI} N^{h*}_0 < Z$ for a productivity shock and $(q_{AI} - q_0)(1 - d_l)\gamma^h/\gamma^l < 1$ for a learning shock. Appendix A verifies that junior employment then remains positive in every period.

To see why the path oscillates when $\zeta < 0$, suppose the firm hires one fewer senior than its steady-state stock at the end of a certain period. To complete all tasks, it hires γ^h/γ^l additional juniors. Next period, a fraction $q(1-d_l)$ of them become seniors, so the pipeline delivers $q(1-d_l)\gamma^h/\gamma^l$ new seniors. The missing senior would have remained at the firm only with probability $1-d_h$. With $\zeta < 0$, the replacement exceeds the loss, and the shortage in seniors turns into a surplus. The surplus triggers a hiring freeze, which starves the pipeline, and a new shortage appears one period later. The firm is not making mistakes in forecasting. In every period it hires exactly the workers it currently needs. The oscillation happens because a junior takes one period to become a senior. The firm hires to fix today's productivity gap, which leads to too many seniors tomorrow.

Both shocks can push ζ below zero, but the mechanisms are different. A productivity shock that widens the gap γ^h/γ^l makes the crowding-out effect stronger: one missing senior requires many replacement juniors. A learning shock that raises q makes human capital acquisition faster: a larger share of the replacement cohort matures into seniors together. The oscillations are therefore strongest exactly where AI is most likely to matter, on the productivity gap and on the speed of learning. We now trace the two paths in turn.

Productivity shock. Suppose AI raises productivities from (γ_0^h, γ_0^l) to $(\gamma_{AI}^h, \gamma_{AI}^l)$ at $t = 1$. In the long run the span returns to its pre-shock level, because productivity does not enter the flow condition (Proposition 2). However, junior employment and the span of control fall immediately after the shock. If the post-shock productivity gap is wide, so that $\zeta < 0$ (equivalently, $\gamma_{AI}^h/\gamma_{AI}^l > \frac{1-d_h}{q(1-d_l)}$), the span subsequently crosses its steady-state level in every period. Because the pre- and post-shock steady state spans are the same, the first crossing at $t = 2$ is an overshoot. The firm temporarily becomes more pyramidal than before the shock, after which the oscillations decay. If the gap is narrow, so that $\zeta \geq 0$, the senior stock declines smoothly to its new level while junior employment recovers monotonically. Figure 4 shows both cases: in panel 4a, AI widens the productivity ratio and the span oscillates; in panel 4b, AI compresses the ratio and the span returns smoothly.

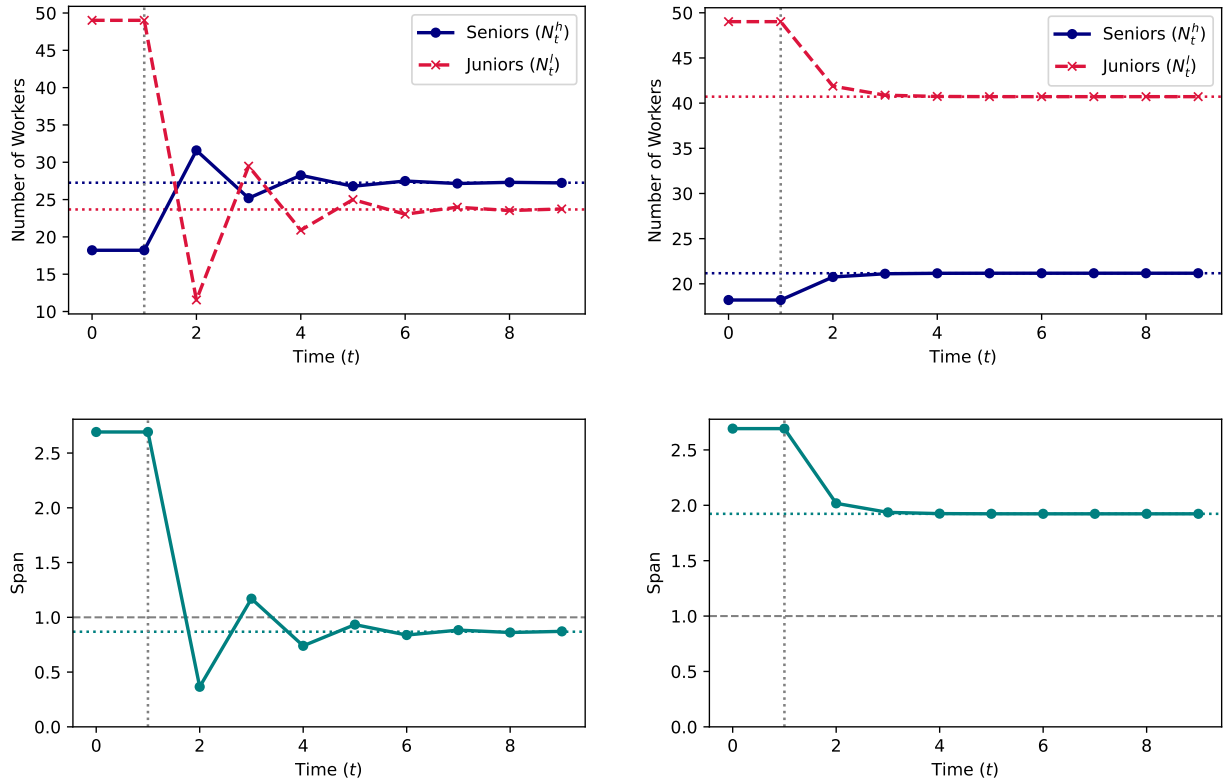
Learning shock. Suppose AI raises the learning rate from q_0 to q_{AI} at $t = 1$. The long-run span now falls (Proposition 2), so the firm converges to a more diamond-like shape rather than back to its original one. At $t = 1$, employment is unchanged while wages fall at once because faster promotion raises the option value of a junior job. Employment begins to adjust at $t = 2$, when the enlarged cohort of newly trained seniors arrives. If the new learning rate is high, so that $\zeta < 0$ (equivalently, $q_{AI} > \frac{1-d_h}{1-d_l}\frac{\gamma^l}{\gamma^h}$), the senior stock, junior employment, and the span alternate around their new steady-state levels with shrinking



(**a**) Oscillating (**b**) Monotonic

Figure 4: Simulation of the transition where AI increases γ^h and γ^l. Parameters are $d_h = 0.49$, $d_l = 0.35$, $q = 0.54$, and $Z = 100$. Both panels start from $(\gamma_0^h, \gamma_0^l) = (2.35, 1)$. In panel (a), AI changes productivities to $(\gamma_{AI}^h, \gamma_{AI}^l) = (3.35, 1.05)$, widening the senior–junior productivity ratio from 2.35 to 3.19 and generating an oscillating transition. In panel (b), AI changes productivities to $(2.40, 2.10)$, compressing the ratio to 1.14 and generating a monotonic transition. For this illustration, one model period can be interpreted as multiple years, so d_l, d_h, and q are cumulative probabilities over the period rather than annual rates.

amplitude. If the new learning rate is modest, so that $\zeta \geq 0$, the senior stock rises smoothly to its new level while junior employment and the span fall monotonically. Figure 5 shows both cases.



(**a**) Oscillating (**b**) Monotonic

Figure 5: Simulation of the transition where AI increases the learning rate q. Parameters are $d_h = d_l = 0.35$, $(\gamma^h, \gamma^l) = (2.8, 1)$, and $Z = 100$. Both panels start from $q_0 = 0.20$. In panel (a), AI raises the learning rate to $q_{AI} = 0.62$, generating an oscillating transition and lowering the steady-state span from 2.69 to 0.87. In panel (b), AI raises the learning rate to $q_{AI} = 0.28$, generating a monotonic transition and lowering the steady-state span from 2.69 to 1.92. For this illustration, one model period can be interpreted as multiple years, so d_l, d_h, q_0, and q_{AI} are cumulative probabilities over the period rather than annual rates.

7 Richer models

The baseline isolates the talent pipeline mechanism in a simple environment: linear production, a fixed number of tasks, a fixed mass of firms, perfectly substitutable seniors, and a learning rate that does not depend on firm structure. We now relax these assumptions in turn. The long-run span survives all of them, because it is pinned down by the stock-flow condition for the senior layer rather than by the production technology. What the extensions change is the wage structure, the scale of the firm, and the amplitude of the transition.

Section 7.1 introduces a general constant-returns technology in which seniors and juniors may be complements and output responds to the shock. The span still depends

only on flow rates, and the pipeline dynamics persist. But with imperfect substitution the span enters both marginal products, so the senior premium now moves with the span, across steady states and along the transition alike. This enriches the cohort inequality result: because the premium covaries with the span, cohorts face a trade-off between ease of entry and future reward. It also dampens the swings, since a cohort entering a hiring boom anticipates a compressed premium and requires a higher wage today.

Section 7.2 lets internally promoted seniors carry firm-specific human capital κ. Firms then hire seniors exclusively from within, and the structure evolves as in Propositions 3 and 4 with $\gamma^h + \kappa$ in place of γ^h. Specific capital does not smooth the adjustment but widens the region in which it oscillates, because each internal senior now replaces more juniors (Corollary B.1). The oscillation is also costly here in a way it is not in the baseline: firm value moves one-for-one with the stock of internal seniors (Proposition 5), so every period in which the pipeline runs short forfeits rents that an external hire cannot restore.

Section 7.3 allows free entry and shows that the entry margin can reverse the baseline freeze-then-boom cycle into a boom-then-freeze cycle. Section 7.4 lets the learning rate depend on the span, which deepens the long-run fall in the span but dampens the transition. Section 7.5 drops the stability condition and shows that under rapid learning the oscillation becomes a permanent two-period cycle. Section 7.6 maps the baseline into a job-assignment framework and shows that optimal assignment reproduces the baseline production structure exactly, so none of the analysis depends on the single-task formulation.

7.1 General Production Technology

The baseline assumes that seniors and juniors are perfect substitutes and that each firm faces a fixed number of tasks. This section relaxes both and reports the long-run results. Details and transition dynamics are in Appendix B.1.

As in Lucas (1978), production takes two steps. Seniors and juniors produce intermediate output through a technology with constant returns to scale, and the CEO turns that intermediate output into final output through a strictly concave managerial technology. Two parameters matter. The *elasticity of substitution* σ measures how easily senior labor replaces junior labor. The two types are substitutes if $\sigma > 1$ and complements if $\sigma < 1$, with Cobb-Douglas at $\sigma = 1$. The *scale elasticity* ϵ_g measures how far the firm expands its output when the marginal cost of production falls.

With imperfect substitution, the span affects marginal products of juniors and seniors and hence their wages. When output can adjust, an AI shock may expand production and employment rather than reduce workers. Table 1 reports the long-run response to each of

the three AI shocks, in the baseline, under Cobb–Douglas, and under a general technology.

Table 1: Long-run responses to AI shocks

Outcome	Baseline	Cobb–Douglas	General production
Productivity shock: $\gamma^h \uparrow$ *or* $\gamma^l \uparrow$			
Span	Unchanged.	Unchanged.	Unchanged.
Wages	Higher senior productivity raises the senior premium and the senior wage and lowers the junior wage; higher junior productivity reverses these effects.	The senior premium and both wages are unchanged.	As in the baseline if $\sigma > 1$. All three signs reverse if $\sigma < 1$.
Output and employment	Output is fixed by the task constraint; senior and junior employment both fall.	Output rises. Senior and junior employment both rise if $\epsilon_g > 1$, fall if $\epsilon_g < 1$, and are unchanged if $\epsilon_g = 1$.	Output rises unless the two types are strong complements. Senior and junior employment move together and rise only if output expands enough to offset the labor saved; they fall otherwise.
Learning shock: $q \uparrow$			
Span	Narrows.	Narrows.	Narrows.
Wages	The senior premium is unchanged; both wages fall.	The senior premium and senior wage fall; the junior wage rises.	The senior premium and senior wage fall; the junior wage rises unless the two types are strong substitutes.
Output and employment	Output is fixed by the task constraint; senior employment rises and junior employment falls.	Output rises if production is senior-intensive. Senior employment rises unless output contracts enough; junior employment falls unless output expands enough.	Output rises if the two types are sufficiently substitutable. Senior employment rises unless output contracts enough; junior employment falls unless output expands enough.

Notes: The table reports steady-state responses to a small rise in each parameter. The learning-shock results in the general-production column assume that seniors earn a wage premium.

The wage responses to a productivity shock depend on substitutability. Consider a senior productivity shock. When seniors and juniors are substitutes, the marginal product of seniors rises relative to that of juniors, so the wage ratio rises. When they are complements, the marginal product of juniors rises by more, so the wage ratio falls. The junior productivity shock is symmetric. Cobb-Douglas is the boundary, where a gain to either type raises both marginal products in the same proportion and leaves the wage ratio unchanged.

Faster learning narrows the span as in the baseline, but here the narrower span also lowers the senior premium, because seniors become relatively more abundant. This changes

the junior wage response. Juniors still become seniors sooner, which lets firms pay less today, as in the baseline. But the senior premium they will earn is lower, and this forces firms to pay more today. The second effect is the stronger one under Cobb-Douglas, and more generally unless the two types are strong substitutes.

7.2 Firm-Specific Human Capital

Previously, we assumed senior workers are identical across firms. In practice, internal experience generates firm-specific human capital that external hires lack. We now extend the framework to incorporate this friction.

We distinguish seniors by their experience at the firm. External seniors are new to the firm and have productivity γ^h. Internal seniors have prior experience at the firm and productivity $\gamma^h + \kappa$, where $\kappa > 0$ represents firm-specific human capital. Assume the stability condition $\frac{\gamma^h+\kappa}{\gamma^l} q < \frac{2-d_h}{1-d_l}$. 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.

Lemma 3. *With firm-specific human capital, the steady-state equilibrium is unique and symmetric:*

(*i*) *Each firm retains all internal seniors, hires no external seniors, and hires exactly enough juniors to complete the remaining tasks. The span and employment levels are therefore*

$$s^* = \frac{d_h}{q(1-d_l)}, \qquad n^{l*} = \frac{s^* z}{\gamma^l s^* + \gamma^h + \kappa}, \qquad n^{h,in*} = \frac{z}{\gamma^l s^* + \gamma^h + \kappa}, \qquad n^{h,ex*} = 0.$$

(*ii*) *The wage ratio and wage levels are*

$$\frac{w^{h*}}{w^{l*}} = \frac{\gamma^h + \delta(1-d_h)\kappa}{\gamma^l + \delta q(1-d_l)\kappa} =: \Gamma, \quad w^{l*} = (1-\delta)\underline{v}\frac{1-\delta(1-d_h)+\delta q(1-d_l)}{1-\delta(1-d_h)+\Gamma\delta q(1-d_l)}, \quad w^{h*} = \Gamma w^{l*}.$$

Part (i) shows that the firm relies entirely on its internal labor market. Because internal seniors are more productive than external ones, the firm retains all internal seniors and hires seniors only from within. It hires just enough juniors to meet its production target.

Part (ii) shows how firm-specific human capital enters wages. The wage ratio reflects both current productivity and the future value of the specific capital created when juniors

acquire skills and when external seniors become internal. The worker value equations then determine wage levels.

The effects of a productivity shock on firm structure and wage levels remain as in the baseline. But because the learning rate now affects how quickly juniors acquire firm-specific human capital, the wage effects of a learning shock can change.

Corollary 1. *In the steady state, suppose the learning rate q increases. Then:*

(*i*) *the wage ratio w^{h*}/w^{l*} strictly decreases;*

(*ii*) *the senior wage w^{h*} strictly decreases;*

(*iii*) *there exists a threshold $\kappa^\dagger > 0$ such that the junior wage w^{l*} strictly increases if $\kappa > \kappa^\dagger$ and strictly decreases if $\kappa < \kappa^\dagger$.*

Faster learning compresses the senior wage premium. This is because a junior now does two things: she produces output today and builds firm-specific capital for tomorrow. Faster learning speeds up the second, so the junior layer becomes more valuable relative to the senior layer. The wage ratio then falls, and the senior wage falls with it.

This compression can also push the junior wage up. Recall that a junior accepts a low wage today for the chance to become a senior and earn the senior wage tomorrow. Faster learning means juniors become seniors sooner, and that alone lowers the junior wage in the baseline. But with firm-specific capital, faster learning also lowers the senior wage premium. When firm-specific capital is important, this fall in the senior wage premium outweighs the faster promotion. Firms must then raise the junior wage to keep junior workers from taking the outside option.

We now turn to the transition. Appendix B.2 shows that in an interior symmetric equilibrium firms retain all internal seniors, hire no external seniors, and do not hire extra juniors merely to build the pipeline. The firm's structure evolves as in Proposition 3, with $\gamma^h + \kappa$ in place of γ^h. The transition oscillates when $\frac{\gamma^h_{AI}+\kappa}{\gamma^l_{AI}} > \frac{1-d_h}{q(1-d_l)}$. This shows that although firm-specific capital makes the pipeline more valuable, it does not smooth the transition. A larger κ means that each internal senior replaces more juniors, which amplifies the oscillation in firm structure.

Oscillations now can be costly, because internal seniors carry rents that external hires do not.

Proposition 4. *The firm's value at an internal senior stock $n^{h,in}$ is*

$$V(n^{h,in}) = \lambda\kappa\, n^{h,in} + \frac{z(y-\lambda)}{1-\delta}, \tag{2}$$

where $\lambda = w^l/[\gamma^l + \delta\kappa q(1-d_l)]$ is the cost of completing a task. A permanent increase in γ^h, γ^l, or q at $t=1$ lowers λ. Two implications follow.

(*i*) *At $t=1$ the firm's value strictly increases.*

(*ii*) *Along the transition the firm's value moves in step with $n_t^{h,in}$.*

An AI shock raises firm value on impact because it reduces the marginal cost of completing a task. But whenever the senior stock drops below its steady-state level during the transition, the firm captures fewer specific-capital rents and its value falls with it.

7.3 Free Entry

The baseline model fixes the mass of firms at M. We now allow firms to enter and show that entry may increase aggregate employment and change the transition path.

Let $y(Q_t)$ be the inverse demand for tasks, where $y'(Q) < 0$ and $Q_t = M_t z$. Each firm incurs a fixed operating cost $F > 0$ per period. Potential firms decide whether to enter at the beginning of each period. Free entry therefore implies zero profits. We maintain the stability condition $q\gamma^h/\gamma^l < (2-d_h)/(1-d_l)$ and focus on shocks for which junior employment remains positive throughout the transition.

In steady state, the span, wages, and per-firm employment are determined as in the baseline model. The zero-profit condition then determines aggregate output and the mass of firms. The long-run effects of AI are as follows.

Proposition 5 (Long-Run Effects with Free Entry)**.** *In steady state, any marginal increase in γ^h, γ^l, or q increases aggregate output and the mass of firms.*

(*i*) *For a productivity shock, senior and junior employment rise if and only if demand is sufficiently elastic.*

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

Free entry therefore adds an expansion effect to the labor-saving effect of AI. AI lowers the firm's per-task cost, so new firms enter until profits return to zero. Although each firm employs fewer workers, there are now more firms. Following a productivity shock, senior and junior employment rise when demand is sufficiently elastic for this expansion to dominate the labor-saving effect. A learning shock also shifts the workforce toward seniors, so senior employment always rises, while junior employment rises only when demand is sufficiently elastic.

The increase in the number of firms also changes the transition. After an AI shock, wages and per-task cost adjust immediately. The mass of firms therefore jumps to its new steady-state level. The senior stock, however, is determined by the previous period.

Proposition 6 (Transition with Free Entry)**.** *Following a permanent AI shock at $t = 1$, firm entry is immediate, while employment adjusts gradually:*

(*i*) *The mass of firms jumps to its new steady-state level at $t = 1$ and remains there: $M_t = M_{AI}^* > M_0$ for all $t \geq 1$.*

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

(*iii*) *For $t \geq 2$, employment oscillates as it converges if $\zeta < 0$, and converges monotonically otherwise.*

The key difference from the baseline is the initial effect on junior hiring. With a fixed mass of firms, junior hiring falls in the early stages after the AI shock. With free entry, lower costs also attract new firms. Because senior employment cannot adjust at $t = 1$, the additional output must be produced by hiring juniors. Junior hiring therefore rises after a learning shock and, after a productivity shock, when demand is sufficiently elastic. If employment oscillates, these additional juniors become seniors at $t = 2$, leading firms to reduce junior hiring. Free entry can therefore reverse the baseline "freeze-then-boom" cycle into a "boom-then-freeze" cycle.

7.4 Learning and the Span of Control

The baseline holds the learning rate constant. We now allow learning to depend on the span of control S_t. Each junior learns with probability $q(S_t)$, where $q'(S) < 0$, because, for example, a wider span gives each junior less supervision and slows learning.

Under the regularity conditions in Appendix B.4, there is a unique steady-state span S^* satisfying

$$(1 - d_l)q(S^*)S^* = d_h. \tag{3}$$

It shows that the steady-state span still balances inflows into the senior layer with outflows from it. The difference is that learning now responds to the span, so S^* is determined implicitly. Given S^*, employment and wages are determined as in the baseline, with q replaced by $q(S^*)$. Productivity shocks therefore leave the span unchanged and have the same wage effects as before.

A learning shock brings new effects. Specifically, we model a learning shock as a proportional improvement, $q_\chi(S) = \chi q(S)$ where $\chi > 1$ and $\chi q(S) \leq 1$.

Proposition 7. *Let S^* be the pre-shock steady-state span. A learning shock $\chi > 1$ lowers the span to S^*/χ in the constant-q baseline and lowers the span strictly below S^*/χ when learning depends on the span. Both wages fall by more than in the constant-q baseline.*

The proposition shows that span-dependent learning amplifies the long-run effect of a learning shock on the firm's shape and workers' wages. To see why, recall that a higher learning rate means more juniors become seniors, so the span falls. When learning depends on the span, this narrower span gives each junior more supervision and raises learning further. The span therefore falls further and the economy settles at a lower span than the baseline predicts (Figure 6). This feedback also amplifies the wage effects. By raising the effective learning rate beyond the direct increase, it further lowers the junior wage. The senior wage falls with it because the wage ratio remains fixed.

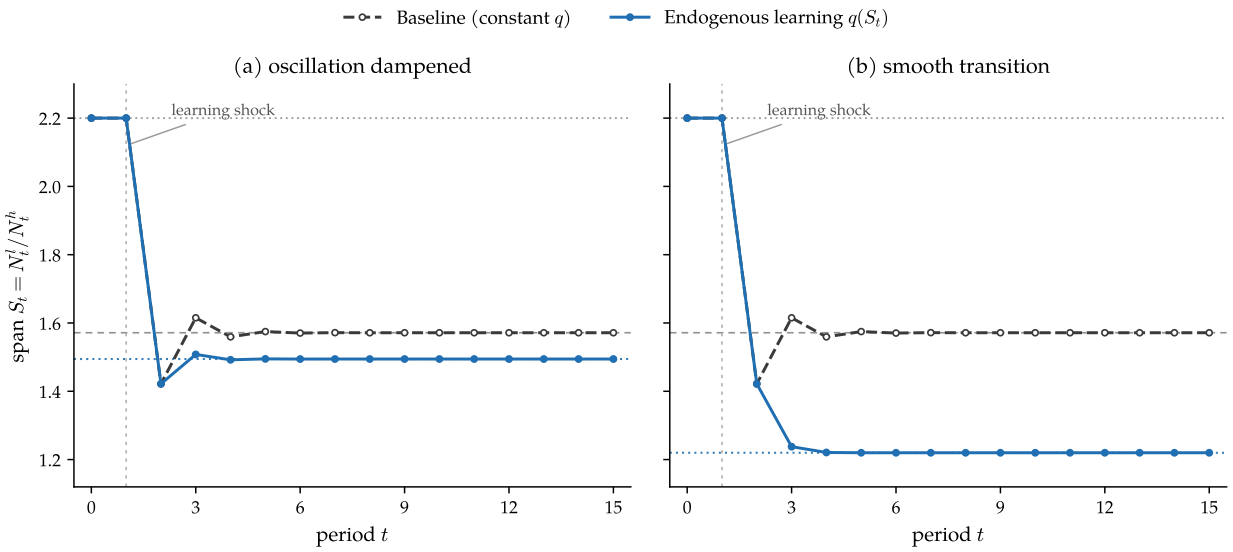


Figure 6: Transition of the span after a learning shock. Both panels start from the same pre-shock steady state S^* and apply the same shock χ at $t = 1$. The baseline (dashed) holds the learning rate fixed at $q(S^*)$ and is identical across panels; span-dependent learning (solid) lets the rate respond to the span through $q(s) = \bar{q}/(1 + a\, s^\rho)$, with $\bar{q} = q_0(1 + u)$ and $a = u/(S^*)^\rho$ (so that $q(S^*) = q_0$ and $\eta_q(S^*) = \rho u/(1 + u)$). Parameters: $d_l = 0.35$, $d_h = 0.44$, $\gamma^h = 3$, $\gamma^l = 1$, $Z = 100$, $\chi = 1.4$, $S^* = 2.2$ (so $q_0 = q(S^*) = 0.308$), $\rho = 0.9$, and $u = 0.20$ in (a), $u = 1.189$ in (b).

Although this feedback deepens the long-run fall in the span, Figure 6 shows that it also tends to dampen the oscillations in transition compared to the baseline. To see why, recall that in the constant-q baseline, the initial fall in the span leaves few juniors to become seniors in the next period. The senior stock then falls and the firm hires more juniors,

and the span rebounds. Span-dependent learning weakens this oscillation. When the span narrows, the remaining juniors learn faster, so the inflow of new seniors falls by less. The span therefore oscillates less in panel (a) and even converges monotonically in panel (b). Industries in which learning depends more on close supervision, such as surgery or research training, may therefore experience a larger long-run change in hierarchy but a smoother adjustment path.

7.5 Rapid Learning

The baseline assumes $q\gamma^h/\gamma^l < (2-d_h)/(1-d_l)$, which ensures convergence to a steady state. We now consider what happens when the learning rate exceeds this threshold, for example, because AI raises q. For simplicity, we assume that unemployed seniors lose their skills after one period and that there is a sufficiently low minimum wage $w_{\min} > 0$ for employed workers.

Proposition 8. *Suppose $\zeta < -1$. Starting from any senior stock other than the steady-state level, the economy eventually enters a permanent two-period cycle:*

(*i*) *Oversupply state. The senior stock exceeds the number needed to complete all tasks. Firms employ Z/γ^h seniors at wage $w_O^h = w_{\min}$ and hire no juniors.*

(*ii*) *Shortage state. The senior stock falls to $(1-d_h)Z/\gamma^h$. Firms hire $d_h Z/\gamma^l$ juniors to complete the remaining tasks. Wages are w_S^l and $w_S^h = (\gamma^h/\gamma^l)w_S^l$, with $w_S^l > w^{l*}$ and $w_S^h > w^{h*}$.*

Proposition 8 shows that rapid learning turns the temporary oscillations identified earlier into a permanent cycle. To see why, suppose the senior stock falls slightly below its steady-state level. Firms hire more juniors to complete the remaining tasks. Because learning is rapid, too many of these juniors become seniors in the next period, causing firms to sharply cut junior hiring. Too few juniors then become seniors in the following period. Each swing is larger than the preceding one until the senior stock is large enough to complete all tasks. Firms then hire no juniors and employ only the seniors they need. Senior departures create a shortage in the next period, and the juniors hired to fill the remaining tasks produce enough seniors to put the economy back into oversupply.

The cycle also makes wages alternate: the senior wage falls to the minimum during oversupply, while both wages rise above their steady-state levels during shortage. This is because in the oversupply state, seniors compete for too few positions, so employed seniors earn only $w_{\min}$. A junior who acquires skills during the shortage state enters the oversupply state in the next period, where the wage is low and employment is uncertain.

So firms must pay juniors a higher current wage, $w_S^l > w^{l*}$. The no-arbitrage condition then gives $w_S^h > w^{h*}$.

As the hiring and wages alternate, career outcomes depend permanently on when workers enter the labor market. Juniors entering during a senior shortage find jobs and initially earn high wages. In the next period, those who become seniors face unemployment risk and earn only the minimum wage if employed. Juniors entering during a senior oversupply find no jobs at all.

7.6 Job Assignment and Hierarchical Structure

In the baseline model, senior and junior workers perform the same type of task and differ only in their per-worker capacity. We now consider an alternative formulation, in the spirit of Gibbons and Waldman (1999*b*), in which the firm has a complex task and a routine task and assigns workers to the two tasks based on their human capital. We show that this formulation delivers the same production structure as the baseline model.

Suppose the firm has two tasks. Task 1 is routine and task 2 is complex. Let $\eta_{it} \in \{0, 1\}$ denote worker i's human capital in period t, with $\eta_{it} = 0$ for a junior and $\eta_{it} = 1$ for a senior. The worker's output on task j is

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

The routine task has $d_1 = \gamma^l$ and $c_1 = 0$, so either type of worker produces γ^l. The complex task has $d_2 = 0$ and $c_2 = \gamma^h$, so a senior produces γ^h, while a junior produces nothing.

Optimal task assignment reproduces the baseline production structure and gives learning a natural interpretation as promotion. Because a junior produces only on the routine task, while a senior produces more on the complex task than on the routine task, the firm assigns juniors to the routine task and seniors to the complex task. Under this assignment, juniors produce γ^l and seniors produce γ^h, exactly as in the baseline model. When a junior acquires human capital and becomes a senior, she moves from the routine task to the complex task and thus up the job ladder.

8 Empirical Patterns

We now return to the BGI worker-profile data and ask which of the model's AI-related channels the data can support. The exercise is descriptive: none of the figures identifies a causal effect of AI on firm hierarchies. But the model makes predictions about steady-state effects on wages and spans and the transition dynamics. It does so across three shocks. The data are rich enough to assess which of the channels would be consistent with the moments that are available in our data, and which would fail to do so.

We proceed in three steps: the composition of the senior layer, the joint movement of span and wages, and the timing of the adjustment.

Composition of the senior layer: retain, not rebuild. The tension at the heart of the model is that junior workers are both involved in current production and determine the future supply of seniors. In the flow condition (FC), the senior layer is replenished from two sources: retained incumbents and the talent pipeline. Figure 7 decomposes the observed senior layer into four parts: (i) retained same-firm seniors, (ii) internal junior-to-senior promotions, (iii) external promoted arrivals (the latter two together constitute the pipeline), and (iv) external senior arrivals. After 2022, the retained share rises in every focus sector while all three flow shares fall. What we observe is that the senior layer is increasingly held together (by incumbents) rather than replenished (through the pipeline). Appendix Figure D.2 shows the same decomposition in levels, so the pattern is not a share artifact. This margin matters in the model: a senior layer sustained by retention rather than replenishment gives up part of the specific-capital rents that make internal training valuable (Section 7.2). Note that a fall in the quit rate of seniors for reasons unrelated to AI could result in a similar picture. We do not see any reason why this should be the case though.

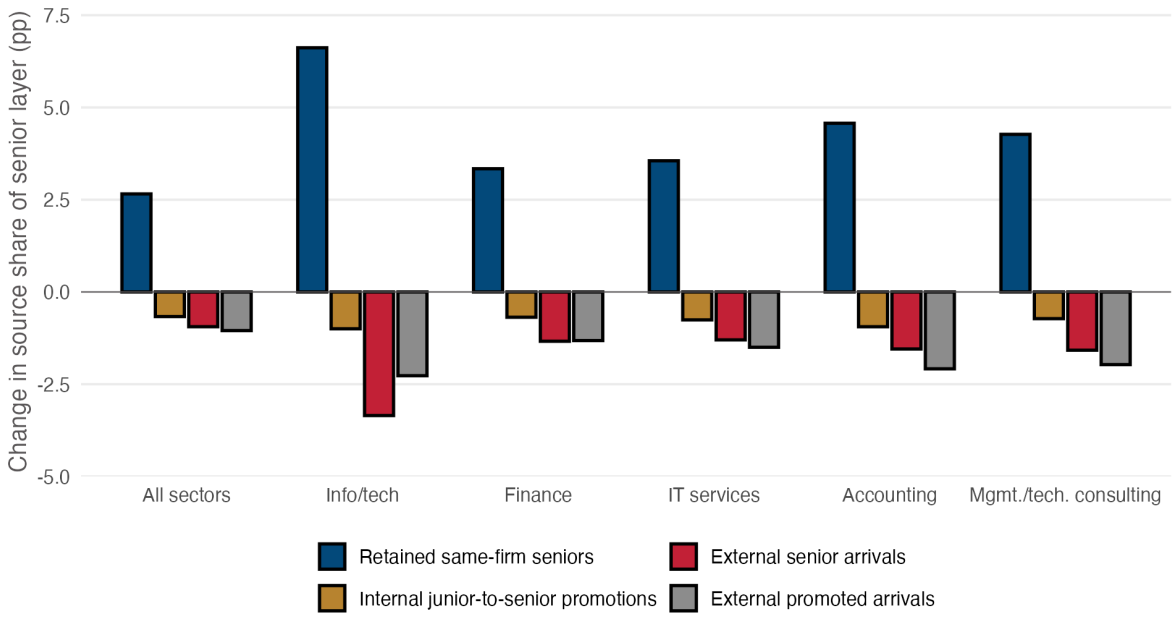


Figure 7: *BGI worker profiles.* Change in the composition of the senior layer. Change from 2021–2022 to 2023–2025 in each source's share of the senior layer: retained same-firm seniors, internal junior-to-senior promotions, external arrivals by already-senior workers, and external arrivals by workers who move up into senior roles. The retained share rises in every sector, while the internal-promotion and external-arrival shares fall. Computed from the worker-history panel; construction in Appendix C.2.

Falling promotion rates and the meaning of q. Table C.5 reports that internal junior-to-senior promotion rates fall in every focus sector after 2022, from 1.11 to 0.85 percent overall. At first glance, this cuts against a learning shock: if q rises, more juniors should qualify for senior positions. The tension can be dissolved once one distinguishes what q is in the model from what a promotion is in the data. In the model, q governs skill acquisition, and promotion follows automatically because the labor market clears. Every junior who acquires the skill is absorbed into the senior layer. In the data, though, a promotion is a realized title change, which requires a senior position to fill. However, as discussed before, Figure 7 shows precisely that these positions are increasingly occupied by retained incumbents. Faster skill acquisition combined with a senior layer that empties more slowly can therefore produce qualified juniors who wait. So, even if the learning rate goes up because of AI, the realized promotion rates go down. Our model cannot generate this effect, because promotion is frictionless. We argue that the joint pattern of Figure 7 and Table C.5 indicates that slot constraints of the kind studied by Ke, Li and Powell (2018) bind in the data. We read the falling measured promotion rate as a limitation of the frictionless-promotion assumption rather than as evidence against a rising q, and we return to this

margin in the concluding section.

Spans and wages together. Figure 8 plots, for each focus sector, the change in the realized junior/senior span against the change in the realized senior/junior wage ratio. Every sector lands in the southwest corner, in which a narrower junior/senior span is accompanied by a compressed senior wage premium. It is helpful to look at Appendix Figure D.3 to understand that this is not a mechanical effect. Actually, junior person-years fall for most sectors while senior person-years rise. Appendix Figure D.4 adds the levels behind the ratio and shows that median junior real wages rise by roughly four to nine percent across sectors, while senior wages move little. Consequently, the compression comes from the bottom.

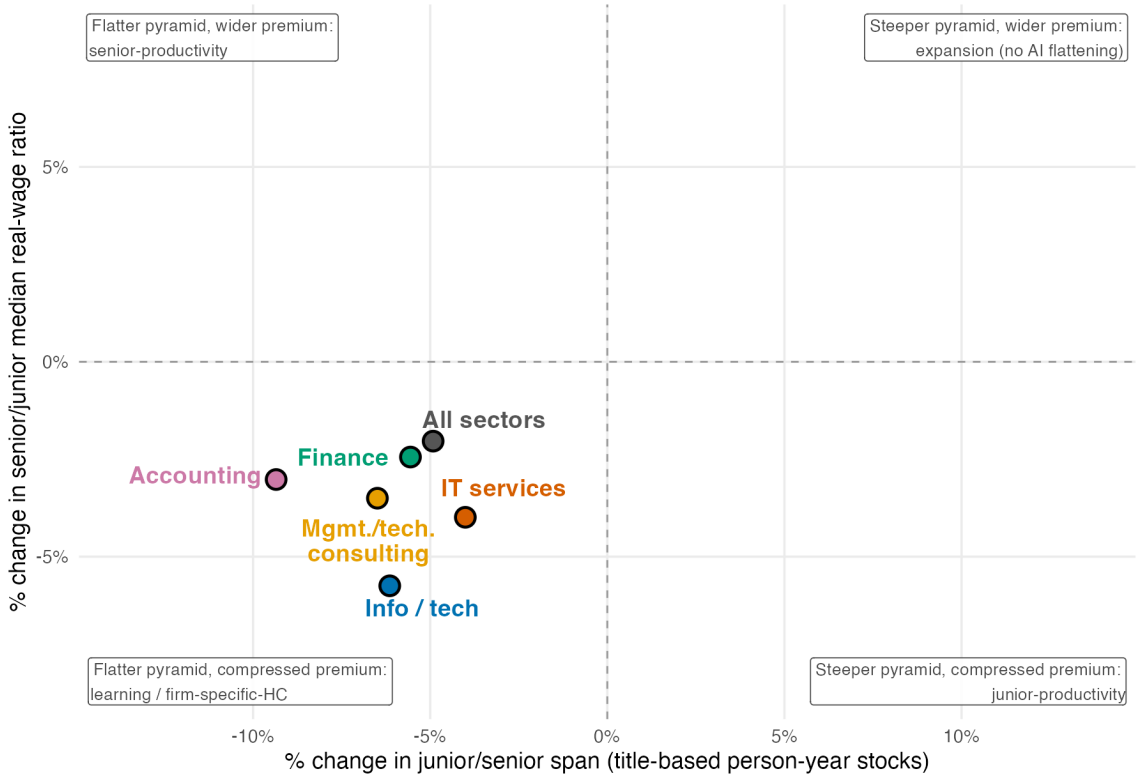


Figure 8: *BGI worker profiles with wages.* Change in the hierarchy span against change in the wage premium, by sector. Each point is a sector; the horizontal axis is the percent change in the junior/senior title-based span and the vertical axis the percent change in the senior/junior median real-wage ratio, both from 2021–2022 to 2023–2025. Computed from the worker-history panel; wages are the median of panel `REAL_WAGE`, about half of which are OEWS imputations. Construction in Appendix C.3.

What channel fits the data best? No single shock in the baseline model matches all three margins in steady state. Recall: a senior-productivity shock leaves the span unchanged and increases the senior wage premium (Proposition 2(i)); empirically this is wrong on

both margins. A junior-productivity shock compresses the senior wage premium with a rising junior wage but leaves the long-run span unchanged (Proposition 2(ii)); it only narrows on the transition path (Proposition 3(i)). A junior learning shock reduces the span (Proposition 2(iii)), leaves the wage ratio constant and predicts that both wage levels fall. This is not in line with the empirical ratio and the observation that junior wages increase.

The extensions do better. Under the Cobb-Douglas technology, a learning shock narrows the span, compresses the wage ratio, raises the junior wage, and lowers the senior wage (Table 1 and Proposition B.2). Here, we have all four observed margins, with the correct signs. With firm-specific human capital, a learning shock narrows the span and compresses the wage ratio (Lemma 3 and Corollary 1), and this extension additionally rationalizes the retain margin of Figure 7 directly, since internal seniors carry rents that firms do not surrender. Either extension therefore independently delivers the joint pattern; the data do not require both at once (we have actually not analyzed a model that combines them).

Timing. The posting data add a fourth discriminating moment. A pure learning shock leaves employment unchanged on impact and moves quantities only with a one-period delay (Proposition 3(i)). In Figure 1, however, the junior-to-senior posting ratio turns down promptly after 2022Q4. An immediate contraction of the junior layer is the signature of a productivity shock (Proposition 3(i)).

Summarizing: Two candidate stories fail cleanly. A pure senior-productivity story widens the premium in the baseline and leaves both wages and the ratio unchanged under Cobb-Douglas, so it cannot produce the observed compression under either technology. A pure learning shock in the baseline model fails on wage levels and on impact timing.

The natural reading of the full set of facts is therefore a joint shock: a productivity component that accounts for the immediate contraction in junior hiring, and a learning component which could operate through complementarity or firm-specific capital or both of them. This accounts for the compressed premium and the shift toward retention. We have characterized each shock in isolation; the stability condition $q\frac{\gamma^h}{\gamma^l} < \frac{2-d_h}{1-d_l}$ already shows that the two interact non-additively, and a formal analysis of the joint shock is a natural next step.

Caveats and ways forward. These patterns come with several caveats. A hiring freeze could mechanically raise the observed junior median wage by changing the composition of the junior pool: with fewer new entrants, the remaining juniors may be more experienced, longer-tenured, and therefore better paid even if wages for comparable workers do not rise. Roughly half of the wage observations are occupation-by-location imputations (Ap-

pendix C.3), and the post-2022 period contains confounds beyond AI, including monetary tightening and sector-specific retrenchment in technology.

These concerns can be assessed descriptively through two complementary checks. First, Appendix Figure D.4 shows that the premium compression is not an artifact of the hot 2021–2022 baseline. Next, Appendix Figures D.5 and D.6 reweight junior and senior wage observations within sector and year to hold fixed their 2021–2022 joint distribution of total labor-market experience and current-job tenure. After this adjustment, junior median wages still rise in all five focus sectors (3.8–10.4%), premium compression becomes 0.53–1.35 percentage points larger, and every focus sector remains in the southwest quadrant of Figure 8.

None of this substitutes for identification: pinning down the causal channel, for instance through within-sector variation in AI exposure, remains the central task for future work. And, finally, the model's most distinctive theoretical result, oscillation between pyramid-like and diamond-like structures during the transition, is not tested here, and cannot be with this window. A full cycle requires observing an initial hiring cutback, a rebuilding cohort, and a second cutback; at the annual frequency of the profile panel, that sequence extends beyond a three-year post-period. There seems to be an uptick in junior postings in 2025Q4 (Figure 1). This could be consistent with the first rebuilding phase but this is no more than suggestive.

The oscillation is, however, the model's falsifiable prediction going forward: cohort-level reversals in junior hiring, sharpest in the firms and occupations where the senior-junior productivity gap is widest and learning fastest. We regard this as the natural test as the post-ChatGPT panel lengthens.

9 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 is to acquire 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 around 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 effects of junior productivity, senior productivity, 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[5] and may need complementary analysis. We have likewise abstracted from firing costs. For some parameter ranges neither margin alters the steady state, since the long-run structure depends only on flow rates and productivity parameters, but both make the transition considerably harder to characterize. We view them as important extensions, and note that our central observation, that today's junior hire determines tomorrow's stock of seniors, operates independently of them. 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 that this model is a good start for thinking about how white collar careers and the management of firms' talent pool will be affected by AI. These are important issues and we see it as an advantage of our model that it is embedded in a labor market, lends itself to being combined with the large-scale data of BGI, and offers reflections on the future of organizations and junior wages. We are also confident that we can allow, in next steps, 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

[5]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 have investigated the role of managers' incentives (Hoffman and Tadelis, 2021; Friebel, Heinz and Zubanov, 2022; Minni, 2026; Haegele, 2024; Friebel and Raith, 2026).

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. 4

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, 4

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

Bar-Isaac, Heski, and Clare Leaver. 2021. "Investment in Human Capital Under Endogenous Asymmetric Information." CEPR Discussion Paper No. DP16305. 36

Beckmann, Martin. 1978. *Rank in organizations.* Berlin:Springer. 5

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. 4, 5

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. 36

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, 4

Fairburn, James A, and James M Malcomson. 1994. "Rewarding performance by promotion to a different job." *European Economic Review*, 38(3-4): 683–690. 36

Freeman, Richard B. 1975. "Supply and Salary Adjustments to the Changing Science Manpower Market: Physics, 1948–1973." *American Economic Review*, 65(1): 27–39. 3

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. 5

Friebel, Guido, and Michael Raith. 2026. "Talent Management: The Role of Bosses." *Journal of Labor Economics*. Forthcoming. 4, 36

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. 36

Fudenberg, Drew, and Luis Rayo. 2019. "Training and effort dynamics in apprenticeship." *American Economic Review*, 109(11): 3780–3812. 5

Garicano, Luis. 2000. "Hierarchies and the Organization of Knowledge in Production." *Journal of political economy*, 108(5): 874–904. 36

Garicano, Luis, and Esteban Rossi-Hansberg. 2015. "Knowledge-based hierarchies: Using organizations to understand the economy." *Annual Review of Economics*, 7(1): 1–30. 36

Garicano, Luis, and Luis Rayo. 2017. "Relational knowledge transfers." *American Economic Review*, 107(9): 2695–2730. 5

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. 5

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. 4

Gibbons, Robert, and Michael Waldman. 1999*a*. "Careers in organizations: Theory and evidence." In *Handbook of Labor Economics*. Vol. 3, 2373–2437. Elsevier. 1, 4

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. 4, 30

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*. 36

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. 36

Huang, Yao, Jin Li, and Yuchen Luo. 2026. "Artificial Intelligence, On-the-Job Learning, and Labor Market Effects." Working paper. 36

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. 5

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*. 5

Karabarbounis, Marios, and Santiago Pinto. 2018. "What Can We Learn from Online Wage Postings? Evidence from Glassdoor." *Economic Quarterly*, 104(4Q): 173–189. 67

Ke, Rongzhu, Jin Li, and Michael Powell. 2018. "Managing careers in organizations." *Journal of Labor Economics*, 36(1): 197–252. 4, 5, 32

Kräkel, Matthias, and Anja Schöttner. 2012. "Internal labor markets and worker rents." *Journal of Economic Behavior & Organization*, 84(2): 491–509. 4

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. 4

Lazear, Edward P, and Sherwin Rosen. 1981. "Rank-order tournaments as optimum labor contracts." *Journal of Political Economy*, 89(5): 841–864. 4

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

Lucas, Robert E., Jr. 1978. "On the Size Distribution of Business Firms." *The Bell Journal of Economics*, 9(2): 508–523. 22

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

Malcomson, James M. 1984. "Work incentives, hierarchy, and internal labor markets." *Journal of political economy*, 92(3): 486–507. 4

Martellini, Paolo, Todd Schoellman, and Jason Sockin. 2024. "The Global Distribution of College Graduate Quality." *Journal of Political Economy*, 132(2): 434–483. 67, 70

Milgrom, Paul, and Sharon Oster. 1987. "Job discrimination, market forces, and the invisibility hypothesis." *The Quarterly Journal of Economics*, 102(3): 453–476. 36

Minni, Virginia. 2026. "Making the invisible hand visible: Managers and the allocation of workers to jobs." *The Quarterly Journal of Economics*, 141(3): 1871–1920. 36

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. 5

Pastorino, Elena. 2024. "Careers in firms: The role of learning about ability and human capital acquisition." *Journal of Political Economy*, 132(6): 1994–2073. 4

Prendergast, Canice. 1993. "The role of promotion in inducing specific human capital acquisition." *The Quarterly Journal of Economics*, 108(2): 523–534. 4

Prendergast, Canice, and Robert H Topel. 1996. "Favoritism in organizations." *Journal of Political Economy*, 104(5): 958–978. 36

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. 5

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

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. 5

Sockin, Jason. 2022. "Show Me the Amenity: Are Higher-Paying Firms Better All Around?" CESifo CESifo Working Paper 9842. 67, 70

Stewman, Shelby, and Suresh Konda. 1983. "Careers and organizational labor markets: Demographic models of organizational behavior." *American Journal of Sociology*, 88(4): 637–685. 5

Waldman, Michael. 1984. "Job assignments, signalling, and efficiency." *The Rand journal of economics*, 15(2): 255–267. 36

Waldman, Michael. 2003. "Ex ante versus ex post optimal promotion rules: The case of internal promotion." *Economic Inquiry*, 41(1): 27–41. 4

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, 4

White, Harrison C. 1970. *Chains of opportunity: System models of mobility in organizations.* Harvard University Press. 5

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. 4

Appendix

A Proofs for the Baseline Model

Proof of Lemma 1. Fix a firm i and period t. First, equilibrium wages cannot be negative. If $w_t^j < 0$ for either worker type $j \in \{l, h\}$, the firm could hire arbitrarily many workers of that type without assigning them additional tasks, making its profit unbounded. Hence, $w_t^l, w_t^h \geq 0$.

Consider either worker type $j \in \{l, h\}$. If $\gamma^j n_{i,t}^j > z_{i,t}^j$, the firm can reduce employment to $z_{i,t}^j/\gamma^j$ without changing its task allocation. This preserves feasibility and weakly increases profit by

$$w_t^j \left(n_{i,t}^j - \frac{z_{i,t}^j}{\gamma^j} \right) \geq 0.$$

If $w_t^j > 0$, the increase is strict, so the original employment choice cannot be optimal. If $w_t^j = 0$, the two employment choices are equally profitable, and we assume that the tie-breaking condition selects the one using fewer workers. Therefore,

$$z_{i,t}^l = \gamma^l n_{i,t}^l \qquad \text{and} \qquad z_{i,t}^h = \gamma^h n_{i,t}^h.$$

It remains to show that the firm completes all z tasks. There is a sufficiently large pool of identical potential junior workers. Entry into junior employment is perfectly elastic at the reservation value $\underline{v}$. Thus, whenever junior employment is positive, free entry implies $v_t^l = \underline{v}$. Moreover, $v_{t+1}^l, v_{t+1}^h \geq \underline{v}$. From the junior worker's value function (VF^l),

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

Since $0 \leq \widetilde{q}_{i,t} \leq 1$, it follows that

$$w_t^l \leq \underline{v} - \delta(1 - d_l)\underline{v} - \delta d_l \underline{v} = (1 - \delta)\underline{v}.$$

The assumption $\gamma^l y > (1 - \delta)\underline{v}$ implies that $y - \frac{w_t^l}{\gamma^l} > 0$. Suppose, contrary to the claim, that

$$z_{i,t}^l + z_{i,t}^h < z.$$

For a sufficiently small $\varepsilon > 0$, the firm can increase $z_{i,t}^l$ by ε and $n_{i,t}^l$ by ε/γ^l. This deviation

is feasible and increases profit by

$$\varepsilon\left(y-\frac{w_t^l}{\gamma^l}\right)>0,$$

contradicting optimality. Hence, $z_{i,t}^l+z_{i,t}^h=z$. Therefore, all three capacity constraints bind. ■

Proof of Lemma 2. By Lemma 1, the capacity constraints bind, so

$$z_{i,t}^l=\gamma^l n_{i,t}^l \qquad \text{and} \qquad z_{i,t}^h=\gamma^h n_{i,t}^h.$$

Because $n_{i,t}^l>0$ and $n_{i,t}^h>0$, both $z_{i,t}^l$ and $z_{i,t}^h$ are strictly positive. Suppose, toward a contradiction, that

$$\frac{w_t^h}{\gamma^h}>\frac{w_t^l}{\gamma^l}.$$

For a sufficiently small $\varepsilon\in(0,z_{i,t}^h)$, consider a deviation in which the firm reallocates ε tasks from seniors to juniors:

$$\widehat{z}_{i,t}^l=z_{i,t}^l+\varepsilon, \qquad \widehat{z}_{i,t}^h=z_{i,t}^h-\varepsilon,$$

and adjusts employment accordingly:

$$\widehat{n}_{i,t}^l=n_{i,t}^l+\frac{\varepsilon}{\gamma^l}, \qquad \widehat{n}_{i,t}^h=n_{i,t}^h-\frac{\varepsilon}{\gamma^h}.$$

The deviation preserves the total number of completed tasks and satisfies both capacity constraints. It therefore leaves revenue unchanged while increasing profit by

$$\varepsilon\left(\frac{w_t^h}{\gamma^h}-\frac{w_t^l}{\gamma^l}\right)>0,$$

contradicting optimality. The reverse inequality can be ruled out symmetrically by reallocating a sufficiently small number of tasks from juniors to seniors. Therefore,

$$\frac{w_t^l}{\gamma^l}=\frac{w_t^h}{\gamma^h}.$$

■

Proof of Proposition 1. Part (i): In a symmetric equilibrium, the flow condition for a

representative firm is given by

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

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

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

Hence, we can get $w^{l*} = \underline{v}(1-\delta)\left[\frac{1-\delta(1-d_h)+\delta q(1-d_l)}{1-\delta(1-d_h)+\frac{\gamma^h}{\gamma^l}\delta q(1-d_l)}\right]$. Since all terms in the numerator and denominator are positive, w^{l*} and w^{h*} are positive.

Moreover, the steady state senior value satisfies

$$v^{h*} - \underline{v} = \frac{\left(\frac{\gamma^h}{\gamma^l} - 1\right)(1-\delta)\underline{v}}{1-\delta(1-d_h) + \frac{\gamma^h}{\gamma^l}\delta q(1-d_l)} > 0,$$

so the senior participation condition is satisfied.

The steady-state flow condition and the binding task constraint uniquely determine aggregate employment, while the no-arbitrage condition and the worker value functions uniquely determine wages and worker values. Hence, the aggregate steady state is unique. ■

Proof of Proposition 2. The results follow from the steady-state expressions in Proposition 1:

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

It is useful to write the wage expressions as

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

and

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

Part (i): The steady-state span does not depend on γ^h, so

$$\frac{\partial S^*}{\partial \gamma^h} = 0.$$

The employment expressions imply

$$\frac{\partial N^{h*}}{\partial \gamma^h} = -\frac{Z}{\left(\gamma^l S^* + \gamma^h\right)^2} < 0$$

and

$$\frac{\partial N^{l*}}{\partial \gamma^h} = -\frac{S^* Z}{\left(\gamma^l S^* + \gamma^h\right)^2} < 0.$$

In the expression for w^{l*}, an increase in γ^h increases the denominator without changing the numerator. Hence,

$$\frac{\partial w^{l*}}{\partial \gamma^h} < 0.$$

In the expression for w^{h*}, differentiating with respect to γ^h gives

$$\frac{\partial w^{h*}}{\partial \gamma^h} = \underline{v}(1-\delta)\left[1-\delta(1-d_h)+\delta q(1-d_l)\right]\frac{\gamma^l \left[1-\delta(1-d_h)\right]}{\left\{\gamma^l \left[1-\delta(1-d_h)\right]+\gamma^h \delta q(1-d_l)\right\}^2} > 0.$$

Part (ii): The steady-state span also does not depend on γ^l, so

$$\frac{\partial S^*}{\partial \gamma^l} = 0.$$

Furthermore,

$$\frac{\partial N^{h*}}{\partial \gamma^l} = -\frac{S^* Z}{\left(\gamma^l S^* + \gamma^h\right)^2} < 0$$

and

$$\frac{\partial N^{l*}}{\partial \gamma^l} = -\frac{(S^*)^2 Z}{(\gamma^l S^* + \gamma^h)^2} < 0.$$

In the expression for w^{h*}, an increase in γ^l increases the denominator without changing the numerator. Hence,

$$\frac{\partial w^{h*}}{\partial \gamma^l} < 0.$$

Differentiating the junior wage gives

$$\frac{\partial w^{l*}}{\partial \gamma^l} = \underline{v}(1-\delta)\left[1-\delta(1-d_h)+\delta q(1-d_l)\right]\frac{\gamma^h \delta q(1-d_l)}{\left\{\gamma^l\left[1-\delta(1-d_h)\right]+\gamma^h\delta q(1-d_l)\right\}^2} > 0.$$

Part (iii): An increase in q reduces the steady-state span:

$$\frac{\partial S^*}{\partial q} = -\frac{S^*}{q} < 0.$$

It follows that

$$\frac{\partial N^{h*}}{\partial q} = -\frac{\gamma^l Z}{(\gamma^l S^* + \gamma^h)^2}\frac{\partial S^*}{\partial q} > 0,$$

whereas

$$\frac{\partial N^{l*}}{\partial q} = \frac{\gamma^h Z}{(\gamma^l S^* + \gamma^h)^2}\frac{\partial S^*}{\partial q} < 0.$$

Differentiating the junior wage with respect to q gives

$$\frac{\partial w^{l*}}{\partial q} = -\underline{v}(1-\delta)\frac{\delta(1-d_l)\left[1-\delta(1-d_h)\right]\left(\frac{\gamma^h}{\gamma^l}-1\right)}{\left[1-\delta(1-d_h)+\frac{\gamma^h}{\gamma^l}\delta q(1-d_l)\right]^2} < 0.$$

Since

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

and the productivity ratio does not depend on q, we also have

$$\frac{\partial w^{h*}}{\partial q} < 0.$$

Finally, because the steady-state wages are strictly positive, the unit-cost equality from Lemma 2 can be written as

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

The steady-state wage ratio therefore does not depend on q. ■

Proof of Proposition 3. Throughout the proof, parameters take their post-shock values unless subscripted by 0, and N_{AI}^{h*}, N_{AI}^{l*}, and S_{AI}^{*} denote the post-shock steady-state values.

Part (i): *Impact.* At the beginning of $t = 1$, the senior stock is inherited from the pre-shock steady state, so $N_1^h = N_0^{h*}$.

Under a productivity shock, the binding capacity constraints (Lemma 1) give

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

The assumption $\gamma_{AI}^h N_0^{h*} < Z$ ensures $N_1^l > 0$. Because $\gamma_{AI}^h \geq \gamma_0^h$ and $\gamma_{AI}^l \geq \gamma_0^l$ with at least one strict inequality,

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

Junior employment therefore falls at $t = 1$ while the senior stock is unchanged.

Under a learning shock, productivities are unchanged, so the binding capacity constraints give $N_1^l = N_0^{l*}$: employment is unchanged at $t = 1$. The flow condition and the pre-shock steady-state flow balance give

$$N_2^h - N_0^{h*} = (q_{AI} - q_0)(1 - d_l)N_0^{l*} > 0, \tag{4}$$

so the senior stock rises at $t = 2$.

Part (ii): *Convergence.* Substituting $N_t^l = (Z - \gamma^h N_t^h)/\gamma^l$ into the flow condition yields

$$N_{t+1}^h = \zeta N_t^h + \frac{q(1 - d_l)Z}{\gamma^l}, \qquad t \geq 1.$$

The post-shock steady state satisfies this equation, and subtracting it from the above equation gives (1), whose solution is

$$N_t^h - N_{AI}^{h*} = \zeta^{t-1}\left(N_1^h - N_{AI}^{h*}\right). \tag{5}$$

Because $d_h > 0$ and $q(1 - d_l)\gamma^h/\gamma^l > 0$, we have $\zeta < 1$; the stability condition gives $\zeta > -1$. The deviation therefore shrinks by the factor $|\zeta|$ each period, alternates in sign if $\zeta < 0$, and keeps its sign if $\zeta \geq 0$.

To translate the senior dynamics into junior employment and the span, compare the binding capacity constraint $\gamma^l N_t^l + \gamma^h N_t^h = Z$ with its steady-state counterpart $\gamma^l N_{AI}^{l*} +$

$\gamma^h N_{AI}^{h*} = Z$. Subtracting gives

$$N_t^l - N_{AI}^{l*} = -\frac{\gamma^h}{\gamma^l}\left(N_t^h - N_{AI}^{h*}\right).$$

Moreover, the span can be written as $S_t = \frac{Z}{\gamma^l N_t^h} - \frac{\gamma^h}{\gamma^l}$, and hence

$$S_t - S_{AI}^* = -\frac{Z}{\gamma^l N_t^h N_{AI}^{h*}}\left(N_t^h - N_{AI}^{h*}\right). \tag{6}$$

Junior employment and the span therefore always deviate from their steady-state levels in the direction opposite to the senior stock. Finally, Proposition 2 implies $N_{AI}^{h*} < N_1^h$ after a productivity shock and $N_{AI}^{h*} > N_1^h$ after a learning shock, which fixes the side on which each path starts. Together with (5)–(6), this delivers the oscillating path when $\zeta < 0$ and the monotonic path when $\zeta \geq 0$.

Positivity of junior employment. Under a productivity shock, $N_1^h > N_{AI}^{h*}$ and the shrinking deviation imply $N_t^h \leq N_1^h < Z/\gamma_{AI}^h$ for all t, so $N_t^l > 0$. Under a learning shock, the pre-shock capacity constraint gives $\frac{Z}{\gamma^h} - N_0^{h*} = \frac{\gamma^l}{\gamma^h} N_0^{l*}$. The assumption $(q_{AI} - q_0)(1 - d_l)\gamma^h/\gamma^l < 1$ and (4) then imply

$$N_2^h - N_0^{h*} = (q_{AI} - q_0)(1 - d_l)N_0^{l*} < \frac{\gamma^l}{\gamma^h} N_0^{l*} = \frac{Z}{\gamma^h} - N_0^{h*},$$

hence $N_2^h < Z/\gamma^h$. If $\zeta \in [0, 1)$, the senior stock rises toward N_{AI}^{h*} from below, so $N_t^h \leq N_{AI}^{h*} < Z/\gamma^h$ for all t. If $\zeta \in (-1, 0)$, every deviation after $t = 2$ is smaller in absolute value than the deviation at $t = 2$, so $N_t^h \leq \max\{N_2^h, N_{AI}^{h*}\} < Z/\gamma^h$ for all t. In all cases, $N_t^l = (Z - \gamma^h N_t^h)/\gamma^l > 0$. ■

B Extensions

B.1 General Production Technology

The baseline model imposes two restrictions. Senior and junior workers are perfect substitutes, and each firm faces a fixed task count. This appendix relaxes both. Senior and junior labor may be complements, and firm scale is determined by the CEO's managerial technology. We first set up the environment and characterize the steady state, then state the long-run effect of AI for a general technology and for the Cobb-Douglas case, and close with the transition. Proofs are collected at the end.

Setup. In period t, firm i hires $n^l_{i,t}$ juniors and $n^h_{i,t}$ seniors. Effective labor inputs are $\gamma^l n^l_{i,t}$ and $\gamma^h n^h_{i,t}$. These inputs produce intermediate output through a production function f, which is increasing, concave, twice continuously differentiable, and homogeneous of degree one. The CEO converts intermediate output into final output through an increasing, twice continuously differentiable, and strictly concave function g:

$$y_{i,t} = g\big[f(\gamma^l n^l_{i,t}, \gamma^h n^h_{i,t})\big].$$

Constant returns to scale imply

$$f(\gamma^l n^l_{i,t}, \gamma^h n^h_{i,t}) = \gamma^h n^h_{i,t}\, \phi(\tilde{s}_{i,t}) \quad \text{and} \quad \tilde{s}_{i,t} \equiv \frac{\gamma^l n^l_{i,t}}{\gamma^h n^h_{i,t}} = \frac{\gamma^l}{\gamma^h} s_{i,t},$$

where $\phi(\tilde{s}) \equiv f(\tilde{s}, 1)$. We call $\tilde{s}_{i,t}$ the *effective span*.

Let $\sigma \equiv \frac{f_l f_h}{f f_{lh}}$ denote the *elasticity of substitution* between the two (effective) labor inputs, where $\sigma = 1$ is the Cobb-Douglas case. A larger σ means juniors and seniors are more substitutable. Let $\epsilon_g \equiv -g'(x)/[x\, g''(x)] > 0$ denote the *scale elasticity* of the managerial technology. A larger ϵ_g means the firm expands more when the marginal cost of intermediate output falls.

Given wages (w^l_t, w^h_t), firm i solves

$$\max_{n^l_{i,t},\, n^h_{i,t}} \; g\big[f(\gamma^l n^l_{i,t}, \gamma^h n^h_{i,t})\big] - w^l_t n^l_{i,t} - w^h_t n^h_{i,t}.$$

The firm's first-order conditions link wages to the span, as the following lemma shows.

Lemma B.1. *When firms hire both types of workers, the wage ratio equals*

$$\frac{w^h_t}{w^l_t} = \frac{\gamma^h}{\gamma^l}\frac{f_h}{f_l} = \frac{\gamma^h}{\gamma^l} \cdot \frac{\phi(\tilde{s}_t) - \tilde{s}_t \phi'(\tilde{s}_t)}{\phi'(\tilde{s}_t)}. \tag{7}$$

The wage ratio now depends on both the relative productivity and the span. The span enters through f_h/f_l, the marginal product of senior labor relative to junior labor. If the two types are perfect substitutes, as in the baseline, then $f_h/f_l = 1$ and the wage ratio reduces to γ^h/γ^l. Under the general technology, f_h/f_l is no longer fixed. A wider span raises the marginal product of seniors relative to juniors and therefore raises the senior premium; a narrower span compresses it.

The next lemma characterizes the steady-state equilibrium.

Lemma B.2. *The economy has a unique steady-state equilibrium:*

(*i*) *the span* $s^* = d_h/[q(1-d_l)]$*, and the effective span* $\tilde{s}^* = (\gamma^l/\gamma^h)s^*$*;*

(*ii*) *the wage ratio* $\frac{w^{h*}}{w^{l*}} = \frac{\gamma^h}{\gamma^l}\frac{\phi(\tilde{s}^*)-\tilde{s}^*\phi'(\tilde{s}^*)}{\phi'(\tilde{s}^*)} \equiv \omega^*$*;*

(*iii*) *wages* $w^{l*} = (1-\delta)\underline{v}\frac{1-\delta(1-d_h)+\delta q(1-d_l)}{1-\delta(1-d_h)+\omega^*\delta q(1-d_l)}$ *and* $w^{h*} = \omega^* w^{l*}$*;*

(*iv*) *employment levels* $n^{h*} = \frac{(g')^{-1}\left(\frac{w^{l*}}{\gamma^l\phi'(\tilde{s}^*)}\right)}{\gamma^h\phi(\tilde{s}^*)}$ *and* $n^{l*} = s^* n^{h*}$*.*

In the long run, flow rates pin down the span. The span sets the senior premium, which together with free entry sets wage levels. Wages then set the firm's scale and employment.

The long-run effect of AI. We now compare steady states. In the statement below, $\sigma_x^h, \sigma_x^l, \bar{\sigma}$, and σ_x^q are constant thresholds in σ, and their expressions are given in the proof.

Proposition B.1 (General Technology)**.** *Consider the steady-state response to a small increase in productivity or in the learning rate.*

Part A (productivity). *Following a rise in* γ^h *or* γ^l*:*

(*i*) *the span* s^* *is unchanged;*

(*ii*) *if* $\sigma > 1$*, a rise in* γ^h *raises the wage ratio* ω^* *and the senior wage* w^{h*}*, and lowers the junior wage* w^{l*}*; a rise in* γ^l *has the opposite effects. If* $\sigma < 1$*, all these signs reverse. If* $\sigma = 1$*, the wage ratio and both wages are unchanged;*

(*iii*) *output scale rises after a rise in* γ^h *if and only if* $\sigma > \sigma_x^h$*, and after a rise in* γ^l *if and only if* $\sigma > \sigma_x^l$*, where* $\sigma_x^h, \sigma_x^l < 1$*;*

(*iv*) *senior and junior employment move together, both rising or both falling.*

Part B (learning). *Following a rise in* q*, and assuming seniors earn a premium,* $\omega^* > 1$*:*

(*i*) *the span* s^* *falls;*

(*ii*) *the wage ratio* ω^* *and the senior wage* w^{h*} *both fall, while the junior wage* w^{l*} *rises if* $\sigma < \bar{\sigma}$*, falls if* $\sigma > \bar{\sigma}$*, and is unchanged if* $\sigma = \bar{\sigma}$*, where* $\bar{\sigma} > 1$*;*

(*iii*) *output scale rises if and only if* $\sigma > \sigma_x^q$*;*

(*iv*) *senior and junior employment need not move together.*

Proposition B.1 shows that the shape result is the same as in the baseline. Productivity does not enter the flow equation, so s^* is pinned down by flow rates alone, and only faster learning moves it.

The wage responses to productivity depend on substitutability. When the two types are substitutes ($\sigma > 1$), a productivity gain for one type lowers the relative marginal product of the other, and wages move as in the baseline. When they are complements ($\sigma < 1$), a productivity gain for one type raises the relative marginal product of the other, and the wage responses reverse.

The response to faster learning is different, and the junior wage is where the new force appears. Under faster learning, juniors reach senior status sooner, but the senior wage they reach is lower, because the narrower span compresses the senior premium. The first effect is the same as in the baseline: faster promotion raises the value of a junior job and lets firms pay less today. The second effect is new: a lower senior wage reduces the value of reaching the senior layer, which forces firms to pay more today. The second effect dominates unless substitution is strong, so the junior wage rises whenever $\sigma < \bar{\sigma}$.

The scale response is new relative to the baseline, which fixes output. The firm expands when the shock lowers the marginal cost of intermediate output and contracts when it raises that cost. Employment then combines scale with shape: under a productivity shock the span is fixed, so both layers move together; under a learning shock the fall in the span shifts the workforce toward seniors, so the two layers need not.

We next consider Cobb–Douglas production, the boundary between substitutes and complements, where the elasticity of substitution equals one.

Proposition B.2 (Cobb–Douglas)**.** *Let $f(\gamma^l n^l, \gamma^h n^h) = (\gamma^l n^l)^{1-\alpha}(\gamma^h n^h)^{\alpha}$, where $\alpha \in (0,1)$ is the output elasticity of senior labor. In steady state:*

(*i*) *The span, the wage ratio, and the junior wage are*

$$s^* = \frac{d_h}{q(1-d_l)}, \qquad \omega^* = \frac{\alpha}{1-\alpha}\, s^*, \qquad w^{l*} = (1-\delta)\underline{v}\, \frac{1-\delta(1-d_h)+\delta q(1-d_l)}{1-\delta(1-d_h)+\frac{\alpha}{1-\alpha}\delta d_h}.$$

(*ii*) *A rise in either γ^h or γ^l leaves the span, the wage ratio, and both wages unchanged. It raises output scale. Senior and junior employment both rise if $\epsilon_g > 1$, both fall if $\epsilon_g < 1$, and remain unchanged if $\epsilon_g = 1$.*

(*iii*) *A rise in q lowers the span, the wage ratio, and the senior wage, but raises the junior wage. Output scale expands if and only if α is sufficiently large. Senior employment can rise even if output contracts, whereas junior employment rises only if output expands sufficiently.*

The proposition under Cobb-Douglas makes a sharper distinction between productivity and learning shocks. Under Cobb–Douglas, a productivity gain raises output but leaves the firm's shape and wage structure unchanged. This is because a gain for either type raises the marginal products of both types in the same proportion, so the wage ratio and both wages remain unchanged. Moreover, the lower marginal cost of production raises output, but employment rises only if output grows enough to offset the reduction in labor required per unit of output.

Faster learning instead narrows the span, lowers the senior wage, and raises the junior wage. Under Cobb-Douglas, the lower senior wage outweighs the effect of faster learning. Moreover, when production is sufficiently senior-intensive, the fall in the senior wage dominates and output expands. Otherwise, output contracts. Employment also reflects the change in span. Faster learning shifts employment from juniors toward seniors. Senior employment can therefore rise even if output contracts moderately, while junior employment rises only if output expands sufficiently.

Transition. Along a transition path, the wage ratio moves with the contemporaneous span through Lemma B.1. The senior premium is high when the span is wide and low when it is narrow. Pipeline fluctuations therefore also pass into wages, rather than wages jumping immediately and permanently to their new steady-state levels as in the baseline.

Together, these wage and structure dynamics shape workers' careers. In particular, when the span oscillates, cohorts alternate between hard entry with a high senior premium after promotion and easy entry with a low senior premium after promotion. This follows because a cohort hired during a hiring freeze faces few junior openings, but those who are promoted reach the senior layer when seniors are relatively scarce and the wage premium is high. By contrast, a cohort hired during a hiring boom enters easily, but those who are promoted reach the senior layer when seniors are relatively abundant and the wage premium is low.

Furthermore, when workers are forward-looking, the resulting wage responses can dampen these potential oscillations in firm structure. During a potential hiring boom, juniors expect a lower senior wage premium after promotion and therefore require a higher wage today. The firm then hires fewer juniors than it otherwise would, which moderates both the current boom and the next period's wave of new seniors. During a potential hiring freeze, juniors expect a higher senior wage premium after promotion and therefore accept a lower wage today, and the firm hires more of them. Forward-looking wages therefore tend to smooth the transition in firm structure.

Proofs. Write $\gamma \equiv \gamma^h/\gamma^l$ and $s \equiv n^l/n^h$. Then $\tilde{s} = s/\gamma$. By constant returns to scale, $f(\gamma^l n^l, \gamma^h n^h) = \gamma^h n^h \, \phi(\tilde{s})$. The marginal products of the two effective inputs are $f_l = \phi'(\tilde{s})$ and $f_h = \phi(\tilde{s}) - \tilde{s}\,\phi'(\tilde{s})$.

The firm's first-order conditions are

$$g'(f)\,\gamma^l \phi'(\tilde{s}) = w^l, \qquad g'(f)\,\gamma^h \big[\phi(\tilde{s}) - \tilde{s}\,\phi'(\tilde{s})\big] = w^h. \tag{8}$$

For later use, define the effective labor shares $\alpha_l \equiv \tilde{s}\phi'(\tilde{s})/\phi(\tilde{s})$ and $\alpha_h \equiv 1 - \alpha_l$. Also define $H(\tilde{s}) \equiv [\phi(\tilde{s}) - \tilde{s}\phi'(\tilde{s})]/\phi'(\tilde{s}) = (\alpha_h/\alpha_l)\tilde{s}$. Under $\phi'' < 0$, $H'(\tilde{s}) = -\phi(\tilde{s})\phi''(\tilde{s})/[\phi'(\tilde{s})]^2 > 0$. Moreover, $\sigma = H(\tilde{s})/[\tilde{s}H'(\tilde{s})]$, which implies $-\tilde{s}\phi''(\tilde{s})/\phi'(\tilde{s}) = \alpha_h/\sigma$. Hence

$$\frac{d\ln H}{d\ln \tilde{s}} = \frac{1}{\sigma}, \qquad \frac{d\ln \phi'}{d\ln \tilde{s}} = -\frac{\alpha_h}{\sigma}, \qquad \frac{d\ln \phi}{d\ln \tilde{s}} = \alpha_l. \tag{9}$$

Proof of Lemma B.1. Dividing the two conditions in (8) gives (7). ■

Proof of Lemma B.2. (*i*) In steady state, the flow condition gives $d_h n^{h*} = q(1-d_l)n^{l*}$. Therefore, $s^* = d_h/[q(1-d_l)]$ and $\tilde{s}^* = (\gamma^l/\gamma^h)s^*$. (*ii*) Lemma B.1 evaluated at $\tilde{s}^*$ gives $\omega^* = (\gamma^h/\gamma^l)H(\tilde{s}^*)$.

(*iii*) Free entry and the worker value equations imply

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

Substituting the second equation into the first and using $w^{h*} = \omega^* w^{l*}$ yields the wage expressions in the statement.

(*iv*) Let x^* denote steady-state intermediate output. The junior first-order condition gives $g'(x^*) = w^{l*}/[\gamma^l \phi'(\tilde{s}^*)]$, where $x^* = \gamma^h n^{h*}\phi(\tilde{s}^*)$. Applying $(g')^{-1}$ gives the expression for n^{h*} in the statement, and $n^{l*} = s^* n^{h*}$. ■

Proof of Proposition B.1, Part A. Part (*i*) follows immediately from $s^* = d_h/[q(1-d_l)]$.

(*ii*) Because $\tilde{s}^* = s^*/\gamma$, $\omega^* = \gamma H(\tilde{s}^*)$, and $\partial \ln \tilde{s}^*/\partial \ln \gamma = -1$, equation (9) gives $\partial \ln \omega^*/\partial \ln \gamma = (\sigma - 1)/\sigma$. Let

$$\varepsilon_l \equiv -\frac{\partial \ln w^{l*}}{\partial \ln \omega^*} = \frac{\delta q(1-d_l)\omega^*}{1-\delta(1-d_h)+\delta q(1-d_l)\omega^*} \in (0,1).$$

It follows that $\partial \ln w^{l*}/\partial \ln \gamma = -\varepsilon_l(\sigma-1)/\sigma$ and $\partial \ln w^{h*}/\partial \ln \gamma = (1-\varepsilon_l)(\sigma-1)/\sigma$. Thus, if $\sigma > 1$, the wage ratio and senior wage rise with γ, while the junior wage falls. If $\sigma < 1$, these signs reverse. If $\sigma = 1$, the wage ratio and both wages are unchanged. The stated

results then follow.

(*iii*) The junior first-order condition gives $g'(x^*) = w^{l*}/[\gamma^l \phi'(\tilde{s}^*)]$. By part (*ii*), equation (9), and $d\ln g'(x^*)/d\ln x^* = -1/\epsilon_g$, we have

$$\frac{\partial \ln x^*}{\partial \ln \gamma^h} = \frac{\epsilon_g}{\sigma}\big[\varepsilon_l(\sigma-1)+\alpha_h\big], \qquad \frac{\partial \ln x^*}{\partial \ln \gamma^l} = \frac{\epsilon_g}{\sigma}\big[(1-\varepsilon_l)(\sigma-1)+\alpha_l\big].$$

The first is positive if and only if $\sigma > \sigma_x^h \equiv 1-\alpha_h/\varepsilon_l$, and the second is positive if and only if $\sigma > \sigma_x^l \equiv 1-\alpha_l/(1-\varepsilon_l)$. Both thresholds are below one.

(*iv*) Because $n^{h*} = x^*/[\gamma^h \phi(\tilde{s}^*)]$, equation (9) gives

$$\frac{\partial \ln n^{h*}}{\partial \ln \gamma^h} = \frac{\partial \ln x^*}{\partial \ln \gamma^h} - \alpha_h, \qquad \frac{\partial \ln n^{h*}}{\partial \ln \gamma^l} = \frac{\partial \ln x^*}{\partial \ln \gamma^l} - \alpha_l.$$

Each expression consists of an output-scale effect and a labor-saving effect. Employment rises if and only if the former outweighs the latter. Because s^* is unchanged, $n^{l*} = s^* n^{h*}$ moves with n^{h*}. ■

Proof of Proposition B.1, Part B. Part (*i*) follows immediately from $s^* = d_h/[q(1-d_l)]$.

(*ii*) Because $\partial \ln \tilde{s}^*/\partial \ln q = -1$, $\omega^* = \gamma H(\tilde{s}^*)$ and equation (9) give $\partial \ln \omega^*/\partial \ln q = -1/\sigma < 0$. Let $\varepsilon_q \equiv \delta q(1-d_l)/[1-\delta(1-d_h)+\delta q(1-d_l)] \in (0,1)$. The wage expressions in Lemma B.2 then imply

$$\frac{\partial \ln w^{l*}}{\partial \ln q} = \varepsilon_q - \varepsilon_l\left(1-\frac{1}{\sigma}\right), \qquad \frac{\partial \ln w^{h*}}{\partial \ln q} = \varepsilon_q - \varepsilon_l - \frac{1-\varepsilon_l}{\sigma}.$$

When $\omega^* > 1$, $\varepsilon_l > \varepsilon_q$, so the senior wage falls. The junior wage rises if $\sigma < \bar{\sigma} \equiv \varepsilon_l/(\varepsilon_l - \varepsilon_q) > 1$, falls if $\sigma > \bar{\sigma}$, and is unchanged if $\sigma = \bar{\sigma}$.

(*iii*) The junior first-order condition, the wage derivative above, and equation (9) give

$$\frac{\partial \ln x^*}{\partial \ln q} = \epsilon_g\left[\varepsilon_l - \varepsilon_q + \frac{\alpha_h - \varepsilon_l}{\sigma}\right].$$

Thus, output rises if and only if $\sigma > \sigma_x^q \equiv (\varepsilon_l - \alpha_h)/(\varepsilon_l - \varepsilon_q)$.

(*iv*) Because $n^{h*} = x^*/[\gamma^h \phi(\tilde{s}^*)]$ and $n^{l*} = s^* n^{h*}$, equation (9) gives

$$\frac{\partial \ln n^{h*}}{\partial \ln q} = \frac{\partial \ln x^*}{\partial \ln q} + \alpha_l, \qquad \frac{\partial \ln n^{l*}}{\partial \ln q} = \frac{\partial \ln x^*}{\partial \ln q} - \alpha_h.$$

The smaller span shifts employment toward seniors, while the output-scale effect moves both layers in the same direction. The two layers therefore need not move together. ■

Proof of Proposition B.2. Under Cobb–Douglas production, $\phi(\tilde{s}) = \tilde{s}^{\,1-\alpha}$, so $H(\tilde{s}) = \alpha\tilde{s}/(1-\alpha)$, $\alpha_l = 1-\alpha$, $\alpha_h = \alpha$, and $\sigma = 1$.

(*i*) Lemma B.2 gives $s^* = d_h/[q(1-d_l)]$ and $\omega^* = \alpha s^*/(1-\alpha)$. Substituting these expressions into the wage formulas in the lemma gives the stated wage levels.

(*ii*) Setting $\sigma = 1$ in Part A(ii) of Proposition B.1 shows that the wage ratio and both wages are unchanged after either productivity shock. Part A(iii) gives $\partial \ln x^*/\partial \ln \gamma^h = \epsilon_g \alpha > 0$ and $\partial \ln x^*/\partial \ln \gamma^l = \epsilon_g(1-\alpha) > 0$. Part A(iv) then gives $\partial \ln n^{h*}/\partial \ln \gamma^h = \alpha(\epsilon_g - 1)$ and $\partial \ln n^{h*}/\partial \ln \gamma^l = (1-\alpha)(\epsilon_g - 1)$. Thus, output rises after either shock, while employment rises if $\epsilon_g > 1$ and falls if $\epsilon_g < 1$. Because the span is fixed, junior employment moves with senior employment.

(*iii*) Part B(i) gives the fall in the span. Setting $\sigma = 1$ in Parts B(ii)–(iii) gives $\partial \ln \omega^*/\partial \ln q = -1$, $\partial \ln w^{l*}/\partial \ln q = \varepsilon_q$, $\partial \ln w^{h*}/\partial \ln q = \varepsilon_q - 1$, and $\partial \ln x^*/\partial \ln q = \epsilon_g(\alpha - \varepsilon_q)$. Because $\varepsilon_q \in (0,1)$, the junior wage rises and the senior wage falls. Output rises if and only if $\alpha > \varepsilon_q$. Finally, Part B(iv) gives $\partial \ln n^{h*}/\partial \ln q = \epsilon_g(\alpha - \varepsilon_q) + (1-\alpha)$ and $\partial \ln n^{l*}/\partial \ln q = \epsilon_g(\alpha - \varepsilon_q) - \alpha$, which give the stated employment conditions. ■

B.2 Firm-Specific Human Capital

Proof of Lemma 3. Let $\lambda_i, \mu_i \geq 0$ denote the stationary multipliers on the capacity and talent constraints. The relevant Kuhn–Tucker conditions are

$$w^l \geq \lambda_i \gamma^l + \mu_i q(1-d_l), \qquad w^h \geq \lambda_i \gamma^h + \mu_i(1-d_h), \qquad \mu_i \geq \delta[-w^h + \lambda_i(\gamma^h + \kappa) + \mu_i(1-d_h)], \tag{10}$$

with equality in each condition when the corresponding employment is strictly positive. Senior participation implies $w^h > 0$.

(*i*) Suppose that external hiring is positive. Consider a firm i that hires external seniors. Its talent constraint must bind. Otherwise, $\mu_i = 0$. Because $n_i^{h,ex} > 0$, (10) gives $w^h = \lambda_i \gamma^h$ and $w^h \geq \lambda_i(\gamma^h + \kappa)$, contradicting $\kappa > 0$ because $w^h > 0$.

Because the talent constraint binds, stationarity and $n_i^{h,ex} > 0$ imply $n_i^{h,in} > 0$. Hence, both senior conditions bind and yield $\mu_i = \delta \lambda_i \kappa$. Together with the junior condition, this implies

$$\frac{w^h}{w^l} \leq \frac{\gamma^h + \delta\kappa(1-d_h)}{\gamma^l + \delta\kappa q(1-d_l)} < \frac{\gamma^h + \kappa}{\gamma^l}.$$

Market clearing requires some firm j to release internal seniors, so its talent constraint is slack and $\mu_j = 0$. This firm cannot hire externally, and stationarity implies $n_j^l > 0$. Hence, (10) gives $w^l = \lambda_j \gamma^l$ and $w^h \geq \lambda_j(\gamma^h + \kappa)$, and therefore $w^h/w^l \geq (\gamma^h + \kappa)/\gamma^l$, contradicting the above inequality. Thus, no external seniors are hired.

With no external hiring, market clearing requires the talent constraint to bind $d_h n_i^{h,in} = q(1-d_l)n_i^l$. Together with the capacity constraint, this implies $n_i^{h,in}, n_i^l > 0$. If $\lambda_i = 0$, (10) gives $[1-\delta(1-d_h)]\mu_i = -\delta w^h < 0$, a contradiction. Hence, $\lambda_i > 0$, so the capacity constraint also binds. The two binding constraints give the stated span and employment levels, which are identical across firms.

(*ii*) Under the competitive outside-market wage, the external-hiring condition binds. Together with the internal-senior condition, it gives $\mu_i = \delta\lambda_i\kappa$. The junior condition then yields

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

Finally, free entry and the worker value equations imply

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

Substituting $w^{h*} = \Gamma w^{l*}$ gives the stated wage levels. ■

Proof of Corollary 1. Part (*i*) follows immediately from the expression for Γ.

(*ii*) Substituting Γ into the senior wage gives

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

Differentiating shows that $\partial w^{h*}/\partial q$ has the sign of $[1-\delta(1-d_h)](\gamma^l - \gamma^h - \kappa) < 0$.

(*iii*) Substituting Γ into the junior wage gives

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

Its derivative with respect to q has the sign of

$$\begin{aligned}&[1-\delta(1-d_h)]\gamma^l[\gamma^l - \gamma^h - \delta(1-d_h)\kappa] \\ &\quad + 2\delta q(1-d_l)\kappa[1-\delta(1-d_h)]\gamma^l + [\delta q(1-d_l)]^2\kappa(\gamma^h+\kappa).\end{aligned}$$

This expression is quadratic in κ. The quadratic term has a positive coefficient. Its constant term is $[1-\delta(1-d_h)]\gamma^l(\gamma^l - \gamma^h) < 0$. It therefore has one negative root and one positive root. Let $\kappa^\dagger > 0$ denote the positive root. The expression is negative when $\kappa < \kappa^\dagger$ and positive when $\kappa > \kappa^\dagger$. Therefore, the junior wage decreases in q below the threshold and increases in q above it. ■

To establish the transition result used in the main text, we first characterize the firm's dynamic hiring policy.

Lemma B.3. *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* ($n_t^l = \frac{z-(\gamma^h+\kappa)n_t^{h,in}}{\gamma^l}$).

Proof. The Kuhn–Tucker conditions are those in (10) with time subscripts. The internal-senior condition is $\mu_t \geq \delta[-w_{t+1}^h + \lambda_{t+1}(\gamma^h + \kappa) + \mu_{t+1}(1 - d_h)]$.

Under the competitive outside-market wage, the external-senior condition binds. Because every firm employs juniors and internal seniors along the equilibrium path, their conditions also bind. The two senior conditions give $\mu_t = \delta\kappa\lambda_{t+1}$. The external-senior and junior conditions then give $w_t^h = \lambda_t\gamma^h + \delta(1 - d_h)\kappa\lambda_{t+1}$ and $w_t^l = \lambda_t\gamma^l + \delta q(1 - d_l)\kappa\lambda_{t+1}$.

Free entry and the worker value equations give $w_t^l = (1 - \delta)\underline{v} - \delta q(1 - d_l)(v_{t+1}^h - \underline{v})$ and $v_t^h - \underline{v} = w_t^h - (1 - \delta)\underline{v} + \delta(1 - d_h)(v_{t+1}^h - \underline{v})$. Let $x_t \equiv v_t^h - \underline{v} + \kappa\lambda_t$. Combining these equations gives

$$x_t = (1 - \delta)\underline{v}\frac{\gamma^h + \kappa - \gamma^l}{\gamma^l} + \delta\left[1 - d_h - q(1 - d_l)\frac{\gamma^h + \kappa}{\gamma^l}\right]x_{t+1}.$$

The stability condition makes the absolute value of the coefficient on x_{t+1} smaller than one. Let x^* be the constant solution. If the coefficient is zero, the equation directly gives $x_t = x^*$. Otherwise,

$$x_{t+1} - x^* = \frac{x_t - x^*}{\delta\left[1 - d_h - q(1 - d_l)\frac{\gamma^h+\kappa}{\gamma^l}\right]},$$

where the multiplier on $x_t - x^*$ has absolute value greater than one. Any nonzero deviation would therefore grow without bound, so boundedness requires $x_t = x^*$ in every period.

The junior condition and free entry give $\gamma^l\lambda_t = (1 - \delta)\underline{v} - \delta q(1 - d_l)x_{t+1}$. Thus, λ_t is constant. The definition of x_t then implies that v_t^h is constant as well. The wage expressions therefore imply that wages are constant over time and equal to their steady-state levels. These wages are strictly positive, so $\lambda > 0$ and $\mu = \delta\kappa\lambda > 0$.

Because $\mu > 0$, the talent constraint binds in every period. Firms therefore retain all internal seniors. In a symmetric equilibrium, external senior supply comes only from internal seniors that firms do not retain. Thus, $n_t^{h,ex} = 0$. Because $\lambda > 0$, the capacity constraint also binds. Hence, $n_t^l = [z - (\gamma^h + \kappa)n_t^{h,in}]/\gamma^l$. ■

Corollary B.1. *A larger κ expands the parameter region in which the transition oscillates.*

Proof. By Lemma B.3, the flow condition in Proposition 3 applies after replacing γ_{AI}^h with

$\gamma^h_{AI} + \kappa$. The transition therefore oscillates when

$$\frac{\gamma^h_{AI} + \kappa}{\gamma^l_{AI}} > \frac{1 - d_h}{q(1 - d_l)}.$$

A larger κ expands the parameter region satisfying this condition. ■

Proof of Proposition 4. By Lemma B.3 wages satisfy $w^h = \lambda[\gamma^h + \delta\kappa(1 - d_h)]$ and $w^l = \lambda[\gamma^l + \delta\kappa q(1 - d_l)]$, and both constraints bind in every period. Substituting the proposed value function and these wages 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 binding talent constraint, the right-hand side becomes $zy - \lambda[\gamma^h n_t^{h,in} + \gamma^l n_t^l] + \delta z(y - \lambda)/(1 - \delta)$. Because capacity binds, $\gamma^h n_t^{h,in} + \gamma^l n_t^l = z - \kappa n_t^{h,in}$, so the right-hand side equals $\lambda\kappa n_t^{h,in} + z(y - \lambda)/(1 - \delta)$. The internal senior stock lies in a compact set and $\delta < 1$, so the bounded solution to the Bellman equation is unique.

(*i*) The proof of Lemma B.3 gives

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

This expression strictly decreases in γ^h and γ^l. Its derivative with respect to q has the sign of $[1 - \delta(1 - d_h)](\gamma^l - \gamma^h - \kappa)$, which is negative. Therefore, each AI shock lowers λ.

Let V_0 and λ_0 denote the value function and shadow price before the shock. Let V_1 and λ_1 denote their values after the shock. Because the internal senior stock is predetermined at $t = 1$, $n_1^{h,in} = n_0^{h,in}$. Equation (2) therefore gives $V_1(n_1^{h,in}) - V_0(n_0^{h,in}) = (\lambda_1 - \lambda_0)\left(\kappa n_0^{h,in} - z/(1 - \delta)\right)$. Because $\lambda_1 < \lambda_0$ and $\kappa n_0^{h,in} < z < z/(1 - \delta)$, the firm's value strictly increases at $t = 1$.

(*ii*) After $t = 1$, the parameters and λ remain fixed. Equation (2) implies $V(n_t^{h,in}) - V(n^{h,in*}) = \lambda\kappa(n_t^{h,in} - n^{h,in*})$. Because $\lambda\kappa > 0$, the firm's value comoves with its internal senior stock. ■

B.3 Free Entry

In steady state, the span, wages, and employment per firm are determined as in the baseline model. Free entry determines the mass of firms. Because $w^{h*}/w^{l*} = \gamma^h/\gamma^l$, the labor cost per task is w^{l*}/γ^l. The zero-profit condition is therefore

$$y(Q^*) = c^* \equiv \frac{w^{l*}}{\gamma^l} + \frac{F}{z}. \tag{11}$$

Because $y(\cdot)$ is strictly decreasing, this condition determines aggregate output Q^* and the mass of firms $M^* = Q^*/z$. Let $\eta \equiv -\frac{y(Q^*)}{y'(Q^*)Q^*}$ denote the price elasticity of demand.

Proof of Proposition 5. Substituting the junior wage in Proposition 1 into (11) gives

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

An increase in either productivity lowers only the first term. An increase in q also lowers this term because

$$\frac{\partial c^*}{\partial q} = -\frac{\underline{v}(1-\delta)\delta(1-d_l)[1-\delta(1-d_h)](\gamma^h-\gamma^l)}{\{\gamma^l[1-\delta(1-d_h)]+\gamma^h\delta q(1-d_l)\}^2} < 0.$$

Thus, every AI shock lowers c^*. Because $y(Q^*) = c^*$ and $y'(Q) < 0$, both Q^* and $M^* = Q^*/z$ increase.

We next consider a marginal shock. Differentiating $y(Q^*) = c^*$ with respect to any parameter x gives $\partial \ln Q^*/\partial x = -\eta\, \partial \ln c^*/\partial x$.

For $x \in \{\gamma^h, \gamma^l\}$, the span S^* does not change and $N^{h*} = Q^*/(\gamma^l S^* + \gamma^h)$ and $N^{l*} = S^* N^{h*}$. Hence,

$$\frac{\partial \ln N^{h*}}{\partial x} = -\eta\frac{1}{c^*}\frac{\partial c^*}{\partial x} - \frac{\partial \ln\left(\gamma^l S^* + \gamma^h\right)}{\partial x}.$$

The first term is positive and increases with η. The second term is negative and does not depend on η. Senior employment therefore rises if and only if demand is sufficiently elastic. Junior employment has the same sign because S^* is fixed.

For a learning shock, the employment expressions can be written as

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

A higher q raises Q^*, so senior employment rises. For junior employment,

$$\frac{\partial \ln N^{l*}}{\partial q} = -\eta\frac{1}{c^*}\frac{\partial c^*}{\partial q} - \frac{\gamma^h(1-d_l)}{\gamma^l d_h + \gamma^h q(1-d_l)}.$$

This expression is strictly increasing in η, negative when η is sufficiently small, and positive when η is sufficiently large. Aggregate junior employment therefore rises if and only if demand is sufficiently elastic. ■

Proof of Proposition 6. Part (i). Wages depend only on the post-shock parameters and adjust at $t = 1$. The per-task cost therefore becomes c^*_{AI} immediately. Because free entry requires $y(M_t z) = c^*_{AI}$, we have $M_t = M^*_{AI} > M_0$ for every $t \geq 1$.

Part (ii). Because $N_1^h = N_0^{h*}$, the aggregate capacity constraint gives $N_1^l = (Q_{AI}^* - \gamma_{AI}^h N_0^{h*})/\gamma_{AI}^l$. Under a learning shock, productivities do not change, so $N_1^l - N_0^{l*} = (Q_{AI}^* - Q_0^*)/\gamma_0^l > 0$.

Under a productivity shock, the old and new steady-state spans are both S^*. Using the post-shock capacity constraint and $N_0^{l*} = S^* N_0^{h*}$ gives

$$N_1^l - N_0^{l*} = \left(S^* + \frac{\gamma_{AI}^h}{\gamma_{AI}^l}\right)\left(N_{AI}^{h*} - N_0^{h*}\right).$$

Junior hiring therefore increases if and only if long-run senior employment increases. Proposition 5 shows that this occurs if and only if demand is sufficiently elastic.

Part (iii). For $t \geq 1$, the flow and capacity conditions imply

$$N_{t+1}^h - N_{AI}^{h*} = \left[1 - d_h - q_{AI}(1-d_l)\frac{\gamma_{AI}^h}{\gamma_{AI}^l}\right]\left(N_t^h - N_{AI}^{h*}\right).$$

The stability condition makes the coefficient in $(-1, 1)$. Employment therefore converges. The coefficient is negative if and only if $q_{AI}\gamma_{AI}^h/\gamma_{AI}^l > (1-d_h)/(1-d_l)$. So the path oscillates if $\zeta < 0$. Otherwise it is monotonic. Junior employment moves in the opposite direction because $N_t^l - N_{AI}^{l*} = -(N_t^h - N_{AI}^{h*})\gamma_{AI}^h/\gamma_{AI}^l$. ■

B.4 Learning and the Span of Control

Assumption B.1. The function $q : (0, \infty) \to (0, \bar{q}]$ satisfies (i) q is continuously differentiable; (ii) $q'(s) < 0$ for all $s > 0$; (iii) the elasticity $\eta_q(s) \equiv -sq'(s)/q(s)$ satisfies $\eta_q(s) < 1$ for all $s > 0$; and (iv) $\lim_{s\to 0^+} q(s)s = 0$ and $\lim_{s\to\infty} q(s)s = \infty$.

We assume that wages remain specific to worker type and do not depend on a firm's span.[6] The static firm problem and the no-arbitrage condition therefore remain unchanged. As in the baseline, we focus on the symmetric equilibrium, in which every firm has span S_t.

The aggregate senior stock evolves according to $N_{t+1}^h = (1-d_h)N_t^h + (1-d_l)q(S_t)N_t^l$. In steady state, this condition gives (3). Moreover, $d[q(s)s]/ds = q(s)[1 - \eta_q(s)] > 0$. Together with Assumption B.1(iv), this shows that the steady-state span is unique. Because (3) contains neither productivity nor capacity, productivity shocks leave the span unchanged. Given S^*, employment and wages are determined as in Proposition 1, with q replaced by $q(S^*)$.

[6] If junior wages could depend on a firm's span, a narrower span could allow the firm to pay a lower junior wage by offering faster learning. Span would then become a firm choice. We abstract from this to focus on the talent-pipeline channel.

Proof of Proposition 7. Let S^*_χ denote the post-shock steady-state span when learning depends on the span. It satisfies $(1-d_l)\chi q(S^*_\chi)S^*_\chi = d_h$. At the constant-$q$ post-shock span S^*/χ, the left-hand side equals $(1-d_l)q(S^*/\chi)S^* > d_h$, because $S^*/\chi < S^*$ and q is decreasing. Because $q(s)s$ is strictly increasing, $S^*_\chi < S^*/\chi$.

The effective post-shock learning rate is $\chi q(S^*_\chi)$. Because $S^*_\chi < S^*$, this rate is strictly greater than $\chi q(S^*)$, the post-shock rate in the constant-q baseline. Proposition 2 shows that both wages are strictly decreasing in the learning rate. Both wages therefore fall by more under span-dependent learning. ■

B.5 Rapid Learning

We assume the wage floor $w_{\min} \in (0, (1-\delta)\underline{v}]$.

Proof of Proposition 8. *Instability.* When $N^h_t < Z/\gamma^h$, firms employ all seniors and juniors complete the remaining tasks. Lemma 1 gives $N^l_t = (Z - \gamma^h N^h_t)/\gamma^l$. Substituting this expression into (FC) gives

$$N^h_{t+1} = \left[1 - d_h - q(1-d_l)\frac{\gamma^h}{\gamma^l}\right] N^h_t + \frac{q(1-d_l)}{\gamma^l} Z. \tag{12}$$

Subtracting the steady-state equation gives $N^h_{t+1} - N^{h*} = [1 - d_h - q(1-d_l)\gamma^h/\gamma^l](N^h_t - N^{h*})$. The condition in the proposition makes the coefficient ζ strictly smaller than -1. Any nonzero deviation therefore alternates in sign and grows in magnitude as long as $\gamma^h N^h_t < Z$. The right-hand side of (12) is decreasing in N^h_t and equals $(1-d_h)Z/\gamma^h > 0$ at $N^h_t = Z/\gamma^h$. The senior stock cannot become negative before reaching the corner. It must therefore reach $N^h_T \geq Z/\gamma^h$ in finite time. If the initial senior stock already satisfies this inequality, the economy begins at the corner.

The two-period cycle. At the corner, firms employ Z/γ^h seniors and hire no juniors. Unemployed seniors lose their skills. After senior departures, $(1-d_h)Z/\gamma^h$ seniors remain in the next period. Firms then hire $d_h Z/\gamma^l$ juniors to complete the remaining tasks. Because firms employ both types of workers, Lemma 2 gives $w^h_S/w^l_S = \gamma^h/\gamma^l$.

The senior stock in the following period is

$$N^h_O = (1-d_h)^2\frac{Z}{\gamma^h} + q(1-d_l)\frac{d_h Z}{\gamma^l}, \quad \gamma^h N^h_O - Z = d_h Z\left[q(1-d_l)\frac{\gamma^h}{\gamma^l} - (2-d_h)\right] > 0.$$

The economy therefore returns to the oversupply state. Because this state again leaves exactly Z/γ^h seniors employed and no juniors hired, the same shortage state follows. The two states repeat indefinitely. In the oversupply state, a senior is employed with probability

$P = (Z/\gamma^h)/N_O^h < 1$. Excess senior supply makes the wage floor bind, so employed seniors earn $w_{\min}$.

Wages. Let v_O^h and v_S^h denote a senior's value at the beginning of the oversupply and shortage states. The value functions and junior free-entry condition then give

$$\begin{aligned} v_O^h &= P\left[w_{\min} + \delta(1-d_h)v_S^h + \delta d_h \underline{v}\right] + (1-P)\underline{v}, \\ v_S^h &= w_S^h + \delta(1-d_h)v_O^h + \delta d_h \underline{v}, \\ \underline{v} &= w_S^l + \delta(1-d_l)\left[q v_O^h + (1-q)\underline{v}\right] + \delta d_l \underline{v}. \end{aligned}$$

Solving the system and using $w_S^h = (\gamma^h/\gamma^l)w_S^l$ shows that $w_S^l - w^{l*}$ has the same sign as

$$(1-\delta)\underline{v}\left[\left(\frac{\gamma^h}{\gamma^l}-1\right) + P\left\{1+\delta\frac{\gamma^h}{\gamma^l}\left[q(1-d_l)-(1-d_h)\right]\right\}\right] - Pw_{\min}\left[1-\delta(1-d_h)+\delta\frac{\gamma^h}{\gamma^l}q(1-d_l)\right].$$

Because $w_{\min} \leq (1-\delta)\underline{v}$, this expression is bounded below by $(1-\delta)\underline{v}\left(\gamma^h/\gamma^l - 1\right)\left[1 - P\delta(1-d_h)\right] > 0$, so $w_S^l > w^{l*}$. Then $w_S^h = (\gamma^h/\gamma^l)w_S^l > (\gamma^h/\gamma^l)w^{l*} = w^{h*}$. ■

C Data and Sample Construction

In this section, we detail the data and construction choices behind the empirical figures in Sections 2 and 8 and Appendix D. We define the samples, how we classify workers and postings, the sector bins, the comparison windows, and the exact object plotted in each figure.

We use two independent data sources from the Burning Glass Institute (BGI). *Job postings* measure advertised vacancies and proxy for labor demand: they show what firms advertise at a point in time, including employer-stated experience requirements and total posting volume. We complement these data with *worker profiles*, which measure realized worker histories: they show titles, employers, promotion histories, and wage estimates, described below in Appendix C.3, for person-years observed in the profile data.

C.1 Job Postings

Source. The posting data come from BGI's March 2026 U.S. job-postings archive. All posting numbers in the paper are computed from full-population sector-by-quarter or firm-by-quarter aggregates.

Universe and filters. We keep job postings from January 1, 2018 through December 31, 2025. Based on the recommendation of the BGI data team, we keep only postings from company job boards to reduce noise in our sample, as postings scraped from non-corporate job boards tend to be lower in quality. This restriction keeps employer-posted vacancies and drops postings from aggregation websites that are more likely to be duplicates. Each posting is assigned to a calendar quarter using its posting date and identified by a `JOB_ID` that uniquely identifies a posting. A firm enters a sector-quarter as soon as it has at least one company-source posting in that sector-quarter.

Seniority classification. Job postings frequently state the seniority level firms are hiring for. In the BGI data, this is documented by `MIN_YEARS_EXPERIENCE`, a variable BGI constructs by parsing the job posting text. Each posting's seniority is read from its stated minimum required years of experience:

$$\text{rank} = \begin{cases} \text{junior} & \text{if minimum experience} \leq 2 \text{ years}, \\ \text{middle} & \text{if minimum experience} \in \{3, 4, 5\} \text{ years}, \\ \text{senior} & \text{if minimum experience} \geq 6 \text{ years}, \\ \text{unknown} & \text{if minimum experience is missing.} \end{cases}$$

We keep job postings with missing `MIN_YEARS_EXPERIENCE` and do not assign them to junior, middle, or senior tags. Across all company-source postings, 49.7% are rank-classifiable; within the five focus sectors for this paper, 62.7% are rank-classifiable. Throughout the paper, junior/senior ratios use postings where a seniority rank is classifiable, whereas figures showing total job postings include both seniority-rank-classifiable and unclassifiable postings.

Sectors. The main posting figures use five NAICS-based industry bins that remain our focus sectors throughout this paper. The broad bins are tech/platform (51) and finance (52). The narrower bins are IT services (5415), accounting (5412), and Mgmt./tech. consulting (54161). Postings outside these bins remain in all-postings counts but are excluded from sector-specific figures.

Posting counts. For each sector-quarter, we count junior, middle, senior, unknown, and total postings. Sector-quarter total postings are the basis for the bars in Figure 1.

Sample sizes. Table C.1 reports the full company-source posting sample used for the posting-side figures. The "Firms (panel)" column is the distinct standardized-firm universe observed at least once over the 2018–2025 panel, not the number of firms active in a given quarter. This distinction matters because most firms post vacancies intermittently; other than a small number of mega-posters, firms need not appear in every month or quarter.

Table C.1: Postings sample, by sector (company-source, 2018-Q1–2025-Q4)

Sector	Postings	Rank-classifiable	Classifiable %	Firms (panel)
Tech/platform (NAICS 51)	2,013,929	1,220,915	60.6	4,721
Finance (NAICS 52)	3,918,078	2,583,362	65.9	5,785
IT services (NAICS 5415)	1,401,944	933,119	66.6	6,118
Accounting (NAICS 5412)	463,427	285,181	61.5	1,673
Mgmt./tech. consulting (NAICS 54161)	1,112,734	565,194	50.8	3,987
All five focus sectors	8,910,112	5,587,771	62.7	22,284
All company-source postings	61,742,621	30,660,449	49.7	185,611

BGI job postings. Company-source U.S. job postings, 2018–2025. "Rank-classifiable" counts postings with non-missing minimum required experience. "Firms (panel)" is the distinct standardized-firm universe observed at least once over the full panel, not a per-quarter active count.

C.2 Worker Profiles

Source. The worker profiles data come from the Burning Glass Institute worker-history panel, May 2026 vintage. These data are constructed from LinkedIn profile histories processed by BGI. Each observation in the data is a person-year. The full data span 2012 through 2025, although the main figures in the paper focus on 2021–2025, with 2021–2022 considered pre-ChatGPT and 2023–2025 considered post-ChatGPT. The worker profiles panel is U.S.-focused.

Inclusion and exclusion. A person-year enters the profile analysis if it has a non-missing person identifier, a non-missing employer name, and a usable job title. We drop person-years flagged as intern, apprentice, student, founder, owner, retired, or volunteer before building the main title ladder. The retained panel totals 600.4 million person-years across 2012–2025: 350.3 million junior person-years and 250.1 million senior person-years. Annual retained volume is about 45.5–46.7 million person-years over 2020–2025.

Title ladder and seniority. Each person-year is assigned to a title rank using an ensemble-based classifier applied to the raw job title stated on worker profiles. More senior patterns take precedence over less senior patterns. The ranks are:

> R0 intern/trainee; R1 entry or unleveled individual contributor; R2 senior individual contributor (for example, senior, lead, principal, or staff titles); R3 manager/supervisor;

R4 director/head; R5 vice president; R6 executive (chief, president, partner, or managing director titles).

The main empirical figures in the paper group junior workers as R0–R1 and senior workers as R2 and above. Note that profile rank is title-based, while junior versus senior rank in Figure 1, which uses job posting data, is experience-based. Importantly, we keep these independent ways of classifying seniority separate: profile seniority is never assigned from posting experience requirements, and posting seniority is never assigned from profile titles.

Title-rank validation. The title classifier is transparent, although it has known limitations. A validation pass across several text-embedding models puts per-category precision at roughly 87–99% for most rungs, with manager titles the weakest at about 80%. The main likely errors are ambiguous title words: for example, "principal" can mean a senior individual contributor in firms but a school principal outside the sectors of interest; "staff" is senior only in staff-plus technical compounds, not in titles such as staff accountant. While these errors are a reason to read the rank classifications cautiously, they are less likely to create the common post-2022 movement across sectors unless title language itself changes sharply at the same time.

Sectors. We use the same five focus sectors described in Appendix C.1. For worker profiles, Mgmt./tech. consulting is measured using the broader NAICS 5416 group available in the profile panel.

For the firm-level profile figures, a firm enters a sector-year if it has at least 100 worker-profile person-years with usable title ranks. Within each firm-year, we count junior person-years (R0–R1) and senior person-years (R2+). Figure 2 uses these counts to measure whether a large firm is top-heavy, while Appendix Figure D.1 compares the median large firm's junior/senior span with the size-weighted aggregate span.

Promotions and employer moves. To empirically construct flows into new positions from worker profiles, we use our worker panel of consecutive-year same-person observations. A *promotion* is an increase in title rank between year $t-1$ and year t. An *internal junior-to-senior promotion* is a move from R0–R1 to R2+ at the same employer. An *external promoted arrival* is a worker who was below senior in the prior year, appears at a different employer in the current year, and is senior in the current year. An external senior arrival is a worker who was already senior in the prior year and appears at a different employer in the current year.

Figure 7 applies these origin definitions to matched senior person-years and plots the change in each origin's share of the senior layer from 2021–2022 to 2023–2025. The

shares are computed using window-level sums. Because the denominator is matched senior person-years, the figure describes the observed profile flow into the senior layer, not every senior worker in the economy. Appendix Figure D.2 reports the same senior-origin categories in average annual person-years rather than shares.

Sample sizes. Table C.2 reports profile person-years by sector. The pooled total includes all sectors, not only the five focus sectors shown in the table.

Table C.2: Profile person-years, by sector (2012–2025 pooled)

Sector	Person-years	Junior (R0–R1)	Senior (R2+)
Tech/platform (NAICS 51)	37,350,702	18,434,347	18,916,355
Finance (NAICS 52)	52,923,032	25,725,204	27,197,828
IT services (NAICS 5415)	17,090,520	8,192,597	8,897,923
Accounting (NAICS 5412)	5,623,008	2,362,545	3,260,463
Mgmt./tech. consulting (NAICS 5416)	9,949,332	4,580,013	5,369,319
All sectors (pooled)	600,377,417	350,324,860	250,052,557

BGI worker profiles. Worker-history panel person-years from 2012–2025, after the exclusions described in the text. Junior and senior groups follow the ensemble-based title classification described in Appendix C.2: junior workers are ranks R0–R1 and senior workers are ranks R2 and above. The pooled total exceeds the sum of the listed rows because most profile person-years fall outside the five focus sectors.

C.3 Wages

Profile wages are annual wage estimates, not advertised salary ranges and not wages stated in profile text. We use `REAL_WAGE`, measured in 2025 dollars, from the worker-history panel. The wage field combines two components. The first is a profile-linked wage-model estimate built from Glassdoor wage distributions: BGI's wage pipeline matches profile records to wage distributions defined by combinations of job title, employer, geography, occupation, year, and experience, moving to coarser cells when exact matches are sparse. Glassdoor wage and review data have been used in prior work to capture meaningful variation across industries, firms, and high-skill labor markets (Karabarbounis and Pinto, 2018; Sockin, 2022; Martellini, Schoellman and Sockin, 2024). The second is an occupation-by-location wage imputation from the Occupational Employment and Wage Statistics (OEWS) survey. Each person-year carries an `OEWS` flag: `OEWS`$= 0$ denotes the profile-linked Glassdoor-based wage-model estimate and `OEWS`$= 1$ denotes the OEWS imputation; missing flags are treated as imputed.

The senior/junior wage premium is the ratio of the median REAL_WAGE among senior person-years (R2+) to the median REAL_WAGE among junior person-years (R0–R1), computed over person-years with positive wages. The annual wage ratio is computed within sector-year, and the window value used in Figure 8 is the unweighted mean of the annual wage ratios inside the window.

Figure 8 combines this wage premium with the realized hierarchy span. The span is junior person-years (R0–R1) divided by senior person-years (R2+) within a sector-year, and the window value is the unweighted mean of annual spans inside the window. Appendix Figure D.3 shows the junior and senior person-year stocks behind the span change, while Appendix Figure D.4 shows wage levels and wage-premium sensitivity to the baseline window.

Two coverage facts bound the wage evidence. First, about half of panel person-years carry an OEWS-imputed wage rather than a profile-linked wage-model estimate: the profile-linked wage share is approximately 0.49–0.55 over 2012–2024. Second, profile-linked wage-model coverage ends in 2024; in 2025 the profile-linked wage share is zero, so the 2025 all-estimates series is entirely OEWS-imputed. For that reason, the wage-premium facts are useful for the joint quantity-price pattern but should not be read as precise measured wage changes.

For the wage-premium baseline check in Appendix Figure D.4, we compare the 2023–2025 wage premium against two baselines: 2018–2020 and 2021–2022. Anchoring to 2018–2020 rather than the hot 2021–2022 labor market produces at least as much premium compression in every focus sector. For example, in tech/platform the senior/junior premium falls by 9.8% relative to 2018–2020 and by 6.1% relative to 2021–2022. Thus, the compression is not created by choosing 2021–2022 as the only baseline.

Experience and current-job-tenure adjustment. To address the possibility that the junior hiring contraction mechanically tilts the post-period junior pool toward more experienced or longer-tenure workers, we use the same positive-REAL_WAGE person-years for which total labor-market experience and current-job tenure can be measured. Total experience is grouped into 0–2, 3–5, 6–10, 11–15, 16–20, 21–30, and 31 or more years. Current-job tenure is the profile year minus the job start year and is grouped into 0, 1, 2, 3–5, 6–10, and 11 or more years. We retain tenure values from 0 to 60 and exclude observations for which current-job tenure exceeds total experience by more than one year. This measure is current-job tenure, not literal years in the worker's current rank.

Separately within every sector, rank group (junior or senior), and year, we reweight the joint experience-by-tenure cells to the corresponding sector-by-rank group's pooled

2021–2022 distribution. We then recompute annual weighted median wages and aggregate them over the same 2021–2022 and 2023–2025 windows used in the paper. Appendix Figure D.5 reports the wage-level and premium results, and Appendix Figure D.6 returns the adjusted premium to the channel map. The published and same-sample unadjusted changes differ by no more than about 0.01 percentage point. As a secondary diagnostic, restricting to profile-linked Glassdoor-based estimates (`OEWS`$= 0$) and ending the post window in 2024 yields flat or negative adjusted junior-wage changes; because that exercise changes the wage source, matched sample, and endpoint simultaneously, we treat it as a comparability caveat rather than as a competing preferred estimate.

C.4 Comparison Windows

Unless stated otherwise, before/after comparisons use 2021–2022 as the pre-period and 2023–2025 as the post-period. The pre-period is the two years immediately before the public release of ChatGPT in late 2022. The post-period is the first three calendar years after that release, and deliberately retains 2023 rather than starting in 2024. Window values are unweighted means of the annual values inside the window unless the figure note below states that the object is built from window-level sums. Figure 1, for example, uses quarterly data from 2018-Q1 through 2025-Q4 because its role is to show the long-term time trend of job postings in key sectors rather than a two-window comparison.

C.5 Known Limitations

This section states the main limitations of the job postings and worker profiles data and, where possible, how they affect interpretation of our empirical results.

1. **Postings are advertised flows.** The high-frequency job postings measure vacancies and are a good proxy for employer demand. However, while a decline in junior postings can reflect lower employer demand, it can also partly reflect changes in the advertising practices of employers or a shift in recruiting channels wherein employers who did not post job ads on corporate job boards start doing so at a larger rate.

2. **Missing experience requirements in job postings.** Note that we use job postings data in only one figure in the paper, Figure 1. As detailed in Appendix C.1, experience requirements in the job postings data come from the variable `MIN_YEARS_EXPERIENCE`. Roughly half of all job postings on company job boards, and about one-third of the job postings in the key focus sectors, lack a stated minimum experience requirement. If firms change how often they state such experience requirements, our empirical

measures of junior-versus-senior ratios can move even without changes in employer demand for those positions. To address such issues, Table C.3 shows the missing-experience share in the five focus sectors by year. The share rises from 36.1% in 2021–2022 to 38.1% in 2023–2025. Therefore, while reporting changes are a limitation in that most postings do not have a junior or senior tag, the aggregate movement in the reported shares is likely not a sharp enough break to mechanically explain the common post-2022 patterns.

3. **Profile ranks are title-based.** As discussed in Appendix C.2, the profile rank ladder is title-based. The classifier is ensemble-based and checked over multiple validation runs. Nevertheless, it can misclassify ambiguous titles. While showing a direct year-by-year validation would require exposing proprietary classifier details and hand-validation data, we do not find evidence that this kind of misclassification becomes especially sharp in particular years of our sample.

4. **Profile flows require observed consecutive years.** In our empirical exercises, promotion and senior-origin measures use workers observed in consecutive years. If a senior worker appears in the panel without an observed prior-year record, that worker is not assigned an origin. Table C.4 shows that this matched senior layer covers roughly 96–98% of senior person-years in the relevant windows.

5. **Wage imputation.** About half of profile wages are occupation-by-location imputations, and the 2025 profile wage series is entirely imputed. Prior work using Glassdoor data gives us some confidence that these measures capture relevant variation in the high-skill labor markets we study (Sockin, 2022; Martellini, Schoellman and Sockin, 2024). Appendix C.3 describes the wage construction and the `OEWS` flag in more detail.

6. **Profile recency.** The newest profile years have had less time to appear in the source data. This can mechanically reduce late-period person-year counts, especially for new entrants in the worker profile sample. While the conditioning on observed consecutive-year histories for the promotion and senior-origin facts helps in this regard, it does not entirely eliminate this issue. Reassuringly, Table C.5 shows that the internal junior-to-senior promotion rate falls in every focus sector when the post-period is over an earlier time horizon, i.e., 2023–2024, as well as when the time period includes the last year in the sample, i.e., 2023–2025.

Table C.3: Missing experience requirements in focus-sector postings, by year

Year	Total postings	Missing experience	Missing share
2018	947,846	355,619	37.5%
2019	1,143,917	420,655	36.8%
2020	809,684	307,029	37.9%
2021	1,063,726	372,297	35.0%
2022	1,407,482	520,104	37.0%
2023	1,135,717	448,181	39.5%
2024	1,144,486	430,873	37.6%
2025	1,257,254	467,583	37.2%
2021–2022	2,471,208	892,401	36.1%
2023–2025	3,537,457	1,346,637	38.1%

BGI job postings. Company-source postings in the five focus sectors. Missing experience means the posting lacks a stated minimum experience requirement, so it remains in total posting counts but is not assigned to the junior, middle, or senior posting categories. Annual rows sum the underlying sector-quarter posting aggregates.

Table C.4: Coverage of the matched senior layer

Sector	2021–2022	2023–2025	Change
All sectors	97.2%	98.4%	+1.2 pp
Tech/platform	96.5%	98.1%	+1.5 pp
Finance	97.7%	98.6%	+0.9 pp
IT services	96.6%	98.1%	+1.5 pp
Accounting	96.9%	98.4%	+1.5 pp
Mgmt./tech. consulting	96.2%	97.9%	+1.6 pp

BGI worker profiles. Share of senior person-years (R2+) that can be linked to a prior-year profile record. The table compares matched senior person-years from the senior-flow decomposition with total senior person-years from the profile stock aggregate. Figure 7 is computed on the matched senior layer.

Table C.5: Internal junior-to-senior promotion rates under alternative post windows

Sector	2021–2022	2023–2024	2023–2025	Change to 2023–2024	Change to 2023–2025
All sectors	1.11%	0.92%	0.85%	-0.20 pp	-0.26 pp
Tech/platform	1.98%	1.48%	1.40%	-0.50 pp	-0.58 pp
Finance	1.59%	1.29%	1.24%	-0.30 pp	-0.35 pp
IT services	1.34%	1.03%	0.95%	-0.30 pp	-0.39 pp
Accounting	3.37%	3.05%	2.80%	-0.32 pp	-0.57 pp
Mgmt./tech. consulting	2.19%	1.89%	1.77%	-0.29 pp	-0.42 pp

BGI worker profiles. Internal junior-to-senior promotions are moves from R0–R1 to R2+ at the same employer among workers observed in consecutive years. The table reports window-average annual rates and shows that the post-2022 decline is present whether the post window ends in 2024 or 2025.

D Supplementary Figures

This appendix collects only the supporting figures referenced in the main text. Figure D.1 supports the motivation figure by showing that realized hierarchy flattening is visible for the typical large firm, not only in the size-weighted aggregate. Figure D.2 supports the retain-versus-rebuild result by showing the senior-layer decomposition in levels. Figure D.3 supports the quadrant plot by decomposing the span axis into junior and senior stocks. Figure D.4 supports the quadrant plot by showing the wage-premium result under two baselines. Before/after comparisons use 2021–2022 versus 2023–2025 unless stated otherwise. Appendix C gives the construction details. Figure D.5 shows the wage-level and premium results after holding experience and current-job-tenure composition fixed. Figure D.6 then returns the adjusted premium to the channel map.

Shape Diagnostics

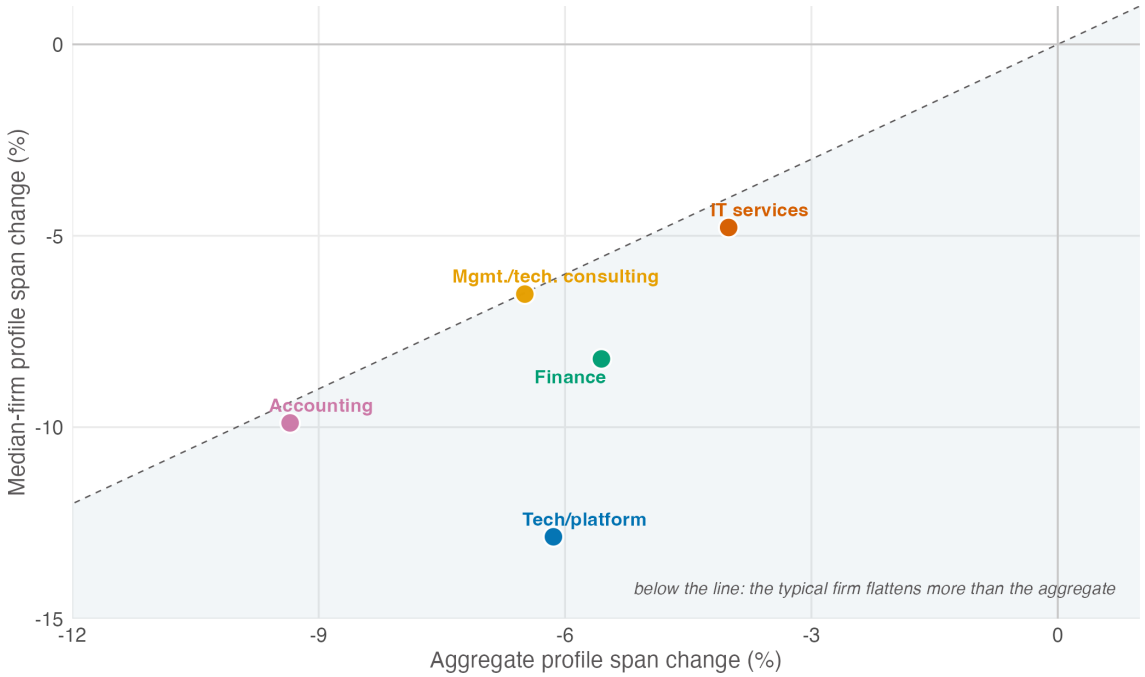


Figure D.1: *BGI worker profiles.* Typical firm versus aggregate profile span. Each point plots a sector's aggregate worker-stock span change against the median large firm's span change, from 2021–2022 to 2023–2025. Firms enter the median if they have at least 100 worker-profile person-years. Dots at or below the equality line mean the typical firm flattens at least as much as the size-weighted aggregate. Construction is described in Appendix C.2.

Retain-versus-Rebuild Robustness

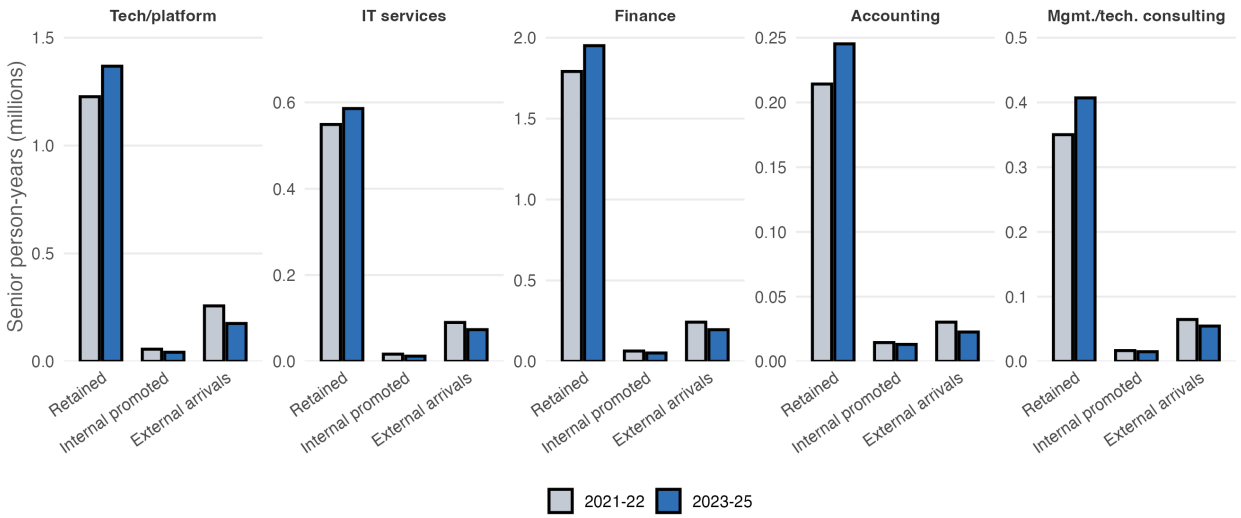


Figure D.2: *BGI worker profiles.* Senior-layer composition in levels. The decomposition of Figure 7 expressed in average annual senior person-years rather than shares, by sector, for 2021–2022 (gray) and 2023–2025 (blue). For the technology industry, for example, retained person-years rise from 1.23M to 1.37M. On the other hand, internal promotions fall from 56k to 42k, whereas external arrivals fall from 257k to 176k. Construction details are in Appendix C.2.

Span Robustness

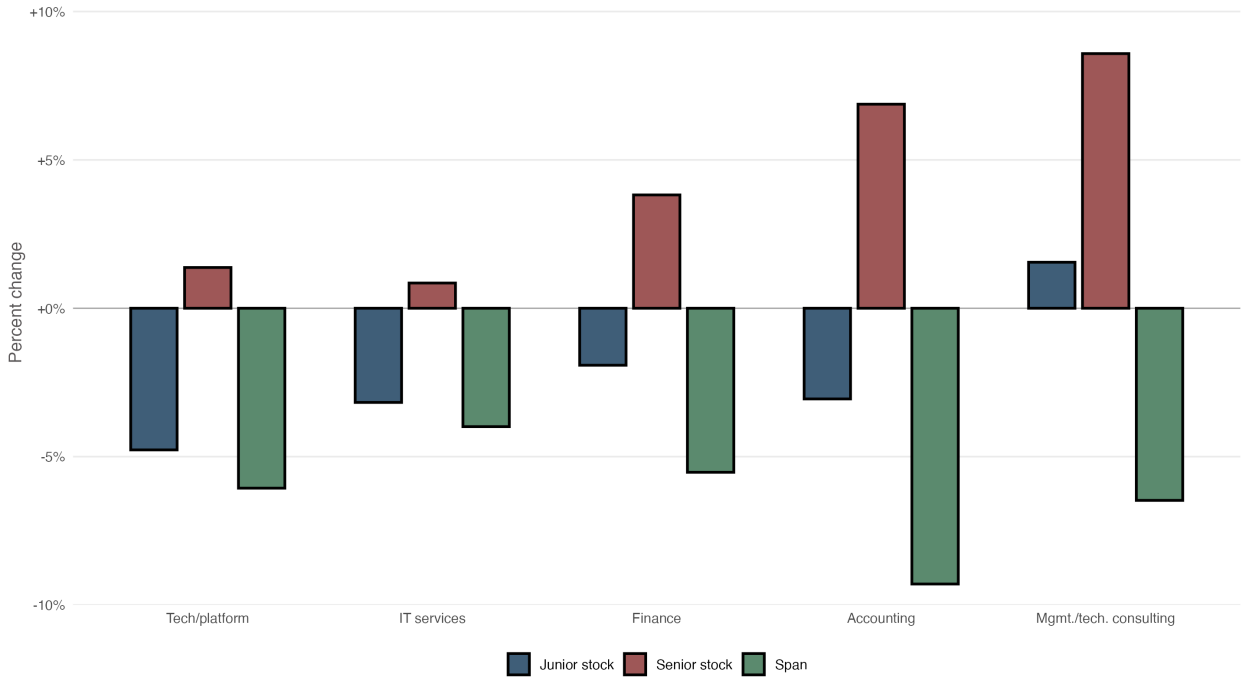


Figure D.3: *BGI worker profiles.* Profile stocks behind the span. Percent change in average annual junior person-years, senior person-years, and the junior/senior span from 2021–2022 to 2023–2025, by sector. Junior ranks are R0–R1 and senior ranks are R2 and above. The figure decomposes the span axis in Figure 8: in most focus sectors, junior person-years fall while senior person-years rise, so the span decline is not only a mechanical ratio artifact. Computed from the worker-history panel; construction in Appendix C.2.

Wage-Premium Robustness

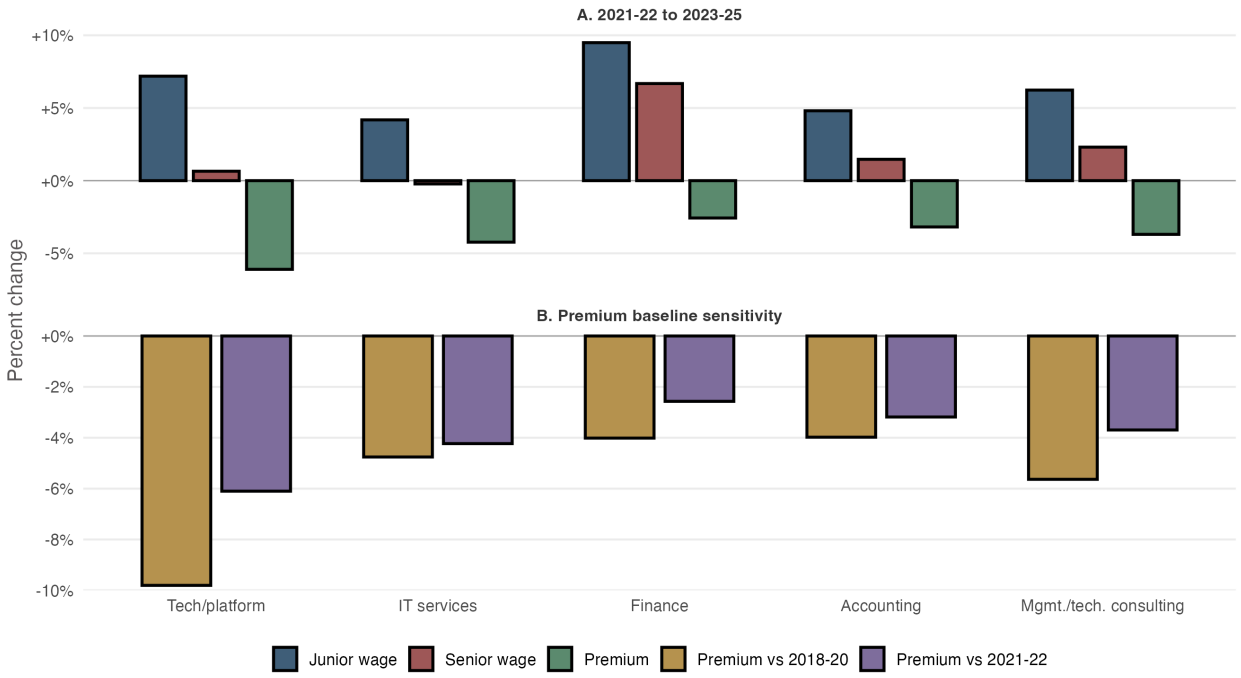


Figure D.4: *BGI worker profiles with wages.* Wage levels and the wage premium under two baselines. Top panel: percent change in median real junior wages, median real senior wages, and their ratio from 2021–2022 to 2023–2025, by sector. Bottom panel: the percent change in the senior/junior wage ratio measured against a 2018–2020 baseline (gold) and a 2021–2022 baseline (purple). Wages are the median of panel `REAL_WAGE`, about half of which are OEWS imputations. Construction details are in Appendix C.3.

Experience- and Tenure-Composition Robustness

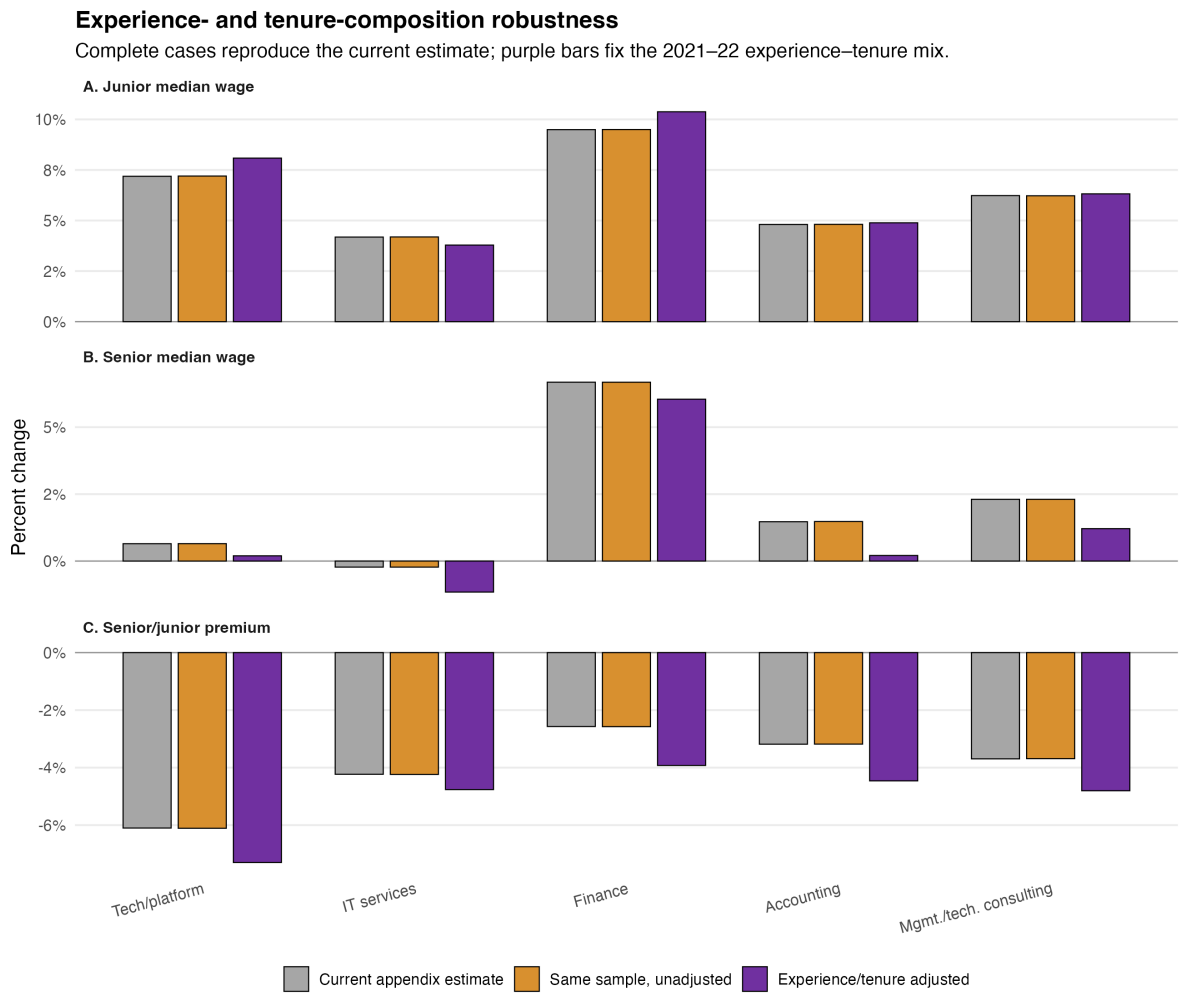


Figure D.5: Experience- and current-job-tenure composition robustness

Notes: All estimates compare 2021–2022 with 2023–2025 and use the paper's full `REAL_WAGE` construction. Gray bars reproduce the current appendix estimate. Orange bars use the same complete-case sample required for the composition adjustment but retain equal observation weights. Purple bars reweight junior and senior observations separately within every sector-year to their sector-by-rank group's joint 2021–2022 distribution of total labor-market experience and current-job tenure. The orange and gray estimates are nearly identical. After adjustment, junior median wages still rise in all five focus sectors and the senior/junior premium compresses more in every sector. Construction details are in Appendix C.3.

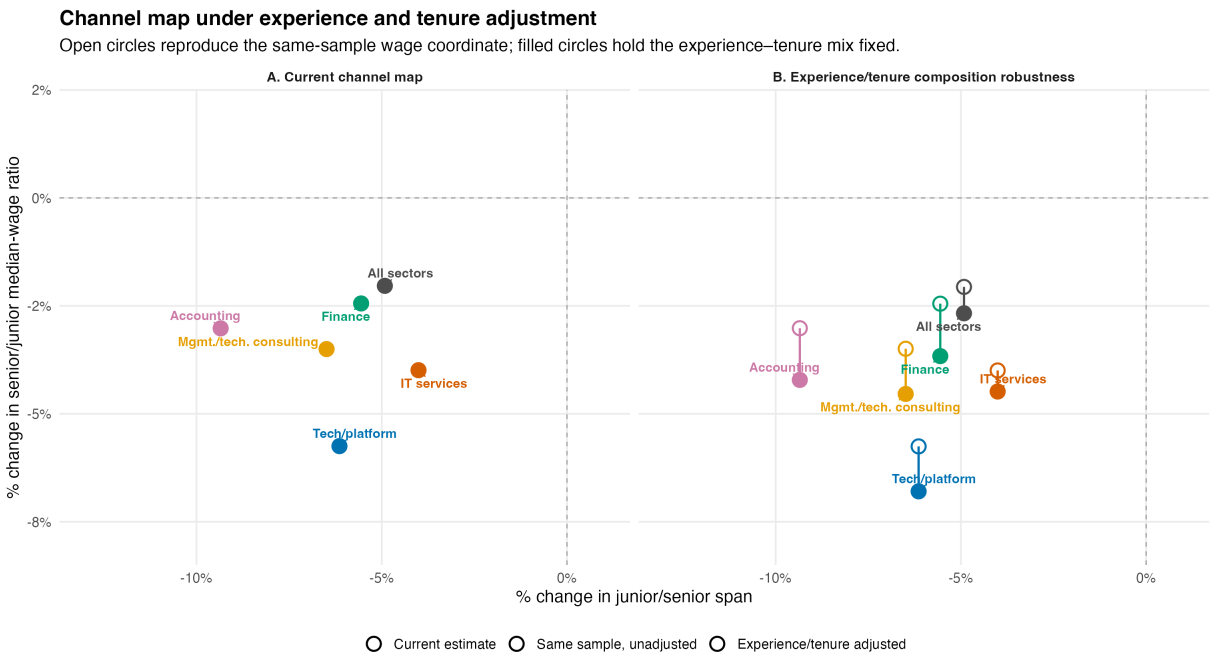
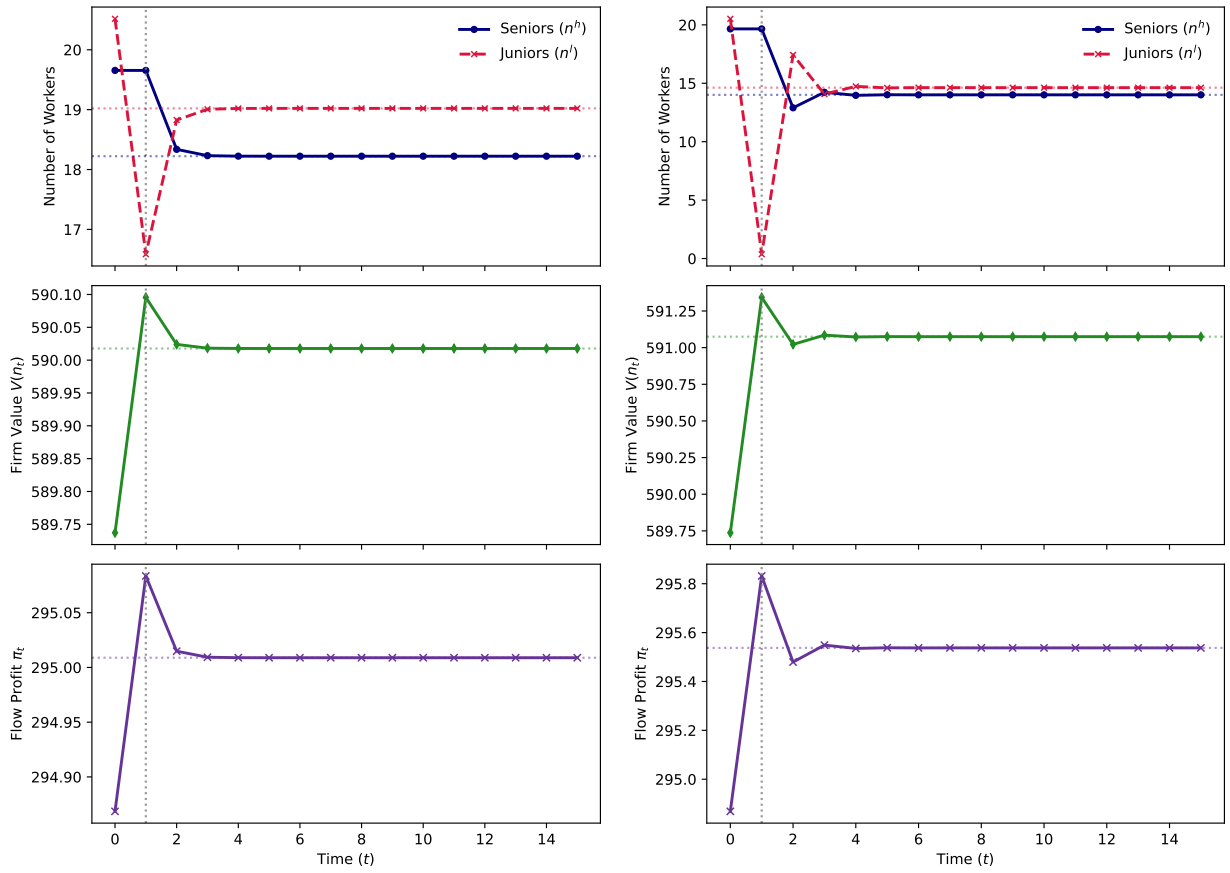


Figure D.6: Channel map under experience and current-job-tenure adjustment

Notes: Both panels compare 2021–2022 with 2023–2025. Panel A reproduces Figure 8. In Panel B, open circles show the unadjusted complete-case wage coordinate and filled circles show the composition-adjusted wage coordinate; segments connect the two. Reweighting changes only the wage-premium coordinate, not the hierarchy-span coordinate. Every focus sector remains in the southwest quadrant, and the adjustment moves the points downward, indicating greater premium compression after holding experience and current-job-tenure composition fixed.

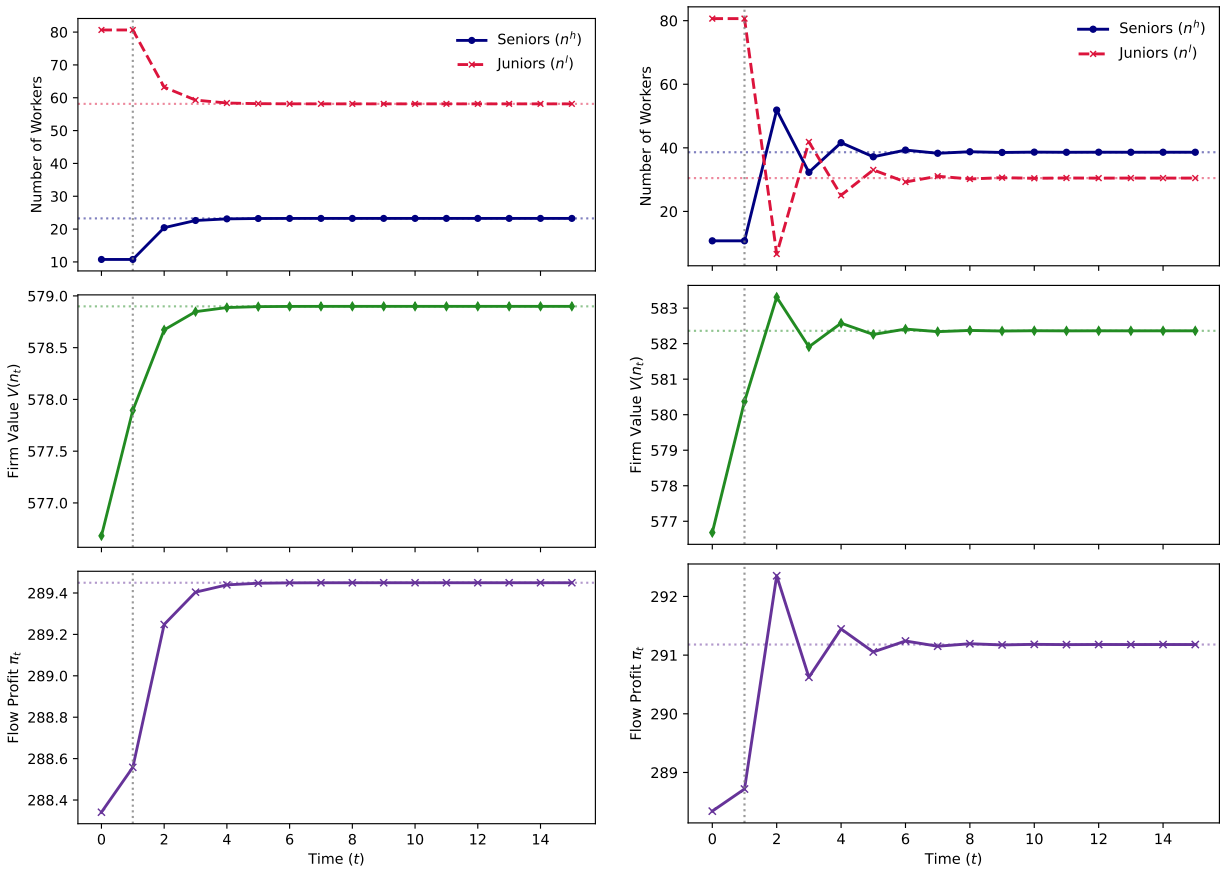
E Figures



(**a**) Smooth Transition.

(**b**) Oscillation.

Figure E.1: Simulation of the Transition of Discounted Firm Value and Current Profit. Parameters: $y = 3$, $\delta = 0.5$, $\kappa = 0.99$, $\underline{v} = 0.25$, and the remaining parameters are the same as in Figure 4.



(**a**) Smooth Transition.

(**b**) Oscillation.

Figure E.2: Simulation of the Transition of Discounted Firm Value and Current Profit. Parameters: $y = 3$, $\delta = 0.5$, $\kappa = 0.7$, $\underline{v} = 0.25$, and the remaining parameters are the same as in Figure 5.