When Replication Kills Initiative: A Theory of Venture Scaling, Restructuring and Innovation

Mu-Jeung Yang*

July 10, 2026

Abstract

Three stylized facts in representative Canadian firm-level panel data present a joint puzzle for existing theories of venture scaling: new establishments of multi-unit firms show flat growth unlike startups, organizational restructuring tracks firm size rather than age, and innovation in service ventures collapses past 150 employees while manufacturing innovation rises monotonically. I develop a formal theory in which startups grow through two channels, innovation and replication, and show that these can be either substitutes or complements, depending on whether frontline workers' initiative is essential for innovation. When it is, as in services, the standardization required for replication kills the initiative that fuels innovation. When it is not, as in manufacturing, the two channels are dynamic complements. The model matches the data across both sectors without relying on uncertainty or irreversible investments, and helps to understand when and why VC-backed startups centralize faster and innovate less over their lifecycle.

*Leeds School of Business, University of Colorado, Boulder. Email: mj.yang@colorado.edu. Financial support by the Leeds School of Business is gratefully acknowledged.

1 Introduction

Understanding startup scaling is a fundamentally strategic question in the sense of Leiblein et al. (2018): the decisions involved in scaling, including organizational restructuring, innovation strategy, incentive design are not just interdependent with each other, but also have persistent effects on organizational performance. Startup scaling has been defined in various ways, including "organizational transformation that accompanies growth" (DeSantola and Gulati, 2017). But while recent empirical work has studied how hierarchy affects startup performance (Lee, 2022) and when startups begin scaling (Lee and Kim, 2024), organizational structure is treated as fixed in both cases. This gap is especially problematic for understanding startup scaling as organizational structure is malleable for young ventures and is typically being adapted as the firm grows from a nascent venture into a full startup and eventually a mature company.

In this paper, I build a new dynamic theory that accounts for three stylized facts about the startup lifecycle documented in Kueng et al. (2017). First, life-cycle growth is a startup phenomenon: new establishments of existing multi-unit firms show flat growth profiles, suggesting that replication and organic startup growth are fundamentally different processes. Second, organizational restructuring tracks firm size rather than age, implying that structure is endogenous to growth rather than a fixed characteristic that evolves with time. Third, innovation in service ventures is hump-shaped, rising to a peak around 15–50 employees and collapsing past 150, a pattern strikingly absent in manufacturing, where innovation increases monotonically with size. To explain these patterns, my theory builds on the insight that firm growth can be achieved through two distinct mechanisms: innovation on the one hand, and replication, or the standardization of operations and processes to reproduce a business model at scale (Winter and Szulanski, 2001; Rivkin, 2001; Lawrence, 2020), on the other hand. The key idea of my model is that these two growth channels can be either substitutes or complements, depending on a single structural condition: whether the workers who deliver the firm's core operations are also the workers who generate its innovations. In services, where they are the same people, replication and innovation are in direct conflict, as centralizing operations kills the initiative required for innovation. In manufacturing, where R&D can be organizationally separated from production, the two channels are complements: centralizing operations funds the growth that increases future innovation. This single distinction generates the divergent innovation profiles observed in the data. The model's core mechanism, centralization killing initiative, is consistent with Winter and Szulanski (2001)'s idea of codification and standardization, and I show in the appendix that the results are robust to explicitly modeling replication as the opening of new establishments that copy the firm's first establishment.

This theory contributes to two literatures. The first is the literature on *venture scaling*. The model endogenizes organizational structure: rather than treating hierarchy as fixed, structure, innovation, and growth are simultaneously determined along a unique equilibrium path. This resolves a circularity[1] in existing informational theories of scaling (Winter and Szulanski, 2001; Gans et al., 2019; Lee and Kim, 2024). From this endogenization, two further results follow. Innovation collapse is not a sign of organizational failure but the signature of successful replication, demanding a careful separation of innovation from growth that prior work has conflated. And the finding that VC-backed startups that scale early are more likely to fail (Lee and Kim, 2024) admits a structural reinterpretation: impatient ownership directly accelerates centralization, suppressing the initiative that generates experimentation.

The second is the literature on *organizations and innovation*. The model identifies the deep primitives, such as ownership patience, separability of operations from innovation, and incentive alignment, which determine when sacrificing innovation for replication becomes optimal, building on the insight that organizational structure mediates a tradeoff between exploration and exploitation (March, 1991; Aghion and Tirole, 1994; Lee, 2022). The paper also complements an empirical literature on R&D organization in mature multi-divisional

[1]Informational theories prescribe that firms should scale only after sufficient experimentation, but experimentation requires employee initiative, which depends on an organizational structure that scaling itself transforms.

firms (Argyres and Silverman, 2004; Argyres et al., 2020, 2025), which documents that across-establishment R&D structure, i.e. the degree of R&D concentration at a corporate central laboratory, is a first-order determinant of innovation outcomes in firms drawn from the Industrial Research Institute and the Cattell directories of corporate R&D. The centralization variable in the present paper is different: it captures the within-establishment allocation of decision authority between managers and frontline workers during the scaling of a focal startup's original establishment, before any across-establishment R&D distribution exists to measure. A five-employee firm at age zero has no R&D facilities to distribute; the relevant organizational question is its within-establishment centralization. In contrast, mature sample firms in Argyres and Silverman (2004); Argyres et al. (2020) are by selection past the scaling trajectory I study. The two literatures therefore address sequential rather than competing questions.

2 Data and Stylized Facts

I build on the empirical analysis by Kueng et al. (2017) to establish the key stylized facts my model seeks to explain.

2.1 Data

Kueng et al. (2017) draw on the Workplace and Employee Survey (WES), a random stratified establishment-level panel conducted by Statistics Canada from 1999 to 2006, covering approximately 6,500 establishments per year across the entire private sector. They restrict attention to the subsample of approximately 5,500 for-profit business firms. The data and methodology are described in detail in Kueng et al. (2017); I summarize the key features here and present the empirical patterns that motivate the theory.

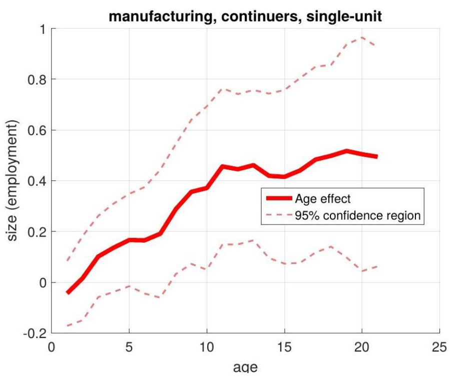
This data is especially useful for understanding venture scaling for three important reasons. First, the WES is *representative*: it covers the entire Canadian economy with a mandatory survey and approximately 90% response rate, avoiding the selection biases inherent in studies of publicly traded firms or single industries. Second, it measures *organizational structure and innovation in the same establishment*: centralization of decision-making, performance pay, occupational structure, and product/process innovation are all observed at the establishment level, allowing the researchers to trace how organizational form and innovation co-evolve with firm size. Third, it is a *panel*: the same establishments and firms are followed over 1999–2006, which permits establishment and firm fixed effects. This is critical for the age-versus-size distinction: fixed effects absorb cohort differences, so any remaining age effects are identified within-establishment. Other representative surveys exist, such as the Annual Business Survey (ABS) for the United States, but these either lack organizational structure variables entirely or, when they include items like structured management practices, collect them only occasionally rather than in every wave, precluding panel estimation with fixed effects.

The key organizational measures are the degree of centralization (the number of tasks decided exclusively by managers), performance pay adoption, the number of occupational layers, and the span of control. Innovation is measured following the Oslo-manual methodology: whether the establishment introduced new or improved products/services (product innovation), new or improved processes (process innovation), or incremental improvements. All results below are from regressions of the outcome on size-bin dummies, controlling for establishment fixed effects and year effects, with 1–4 employees as the omitted reference category.[2] Each plotted coefficient therefore represents how much an establishment of a given size differs from the smallest establishments, after controlling for age and all time-invariant establishment characteristics.

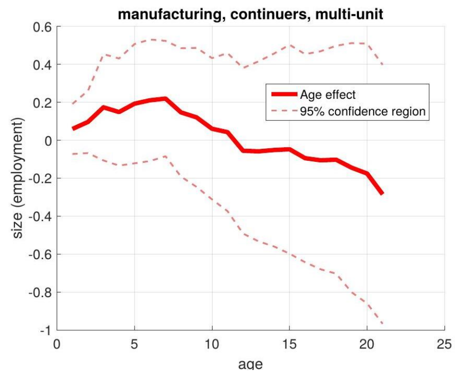
[2]The estimating equation is $x_{e,t} = \nu + \sum_s \kappa_s D_s + \sum_a \lambda_a D_a + \tau_t + \xi_e + \varepsilon_{e,t}$, where κ_s are size-bin coefficients, λ_a are age effects, and ξ_e are establishment fixed effects. Innovation outcomes are binary (linear probability model). Standard errors are clustered at the size–region–industry level. See Kueng et al. (2017) for details.

2.2 Growth is a startup phenomenon

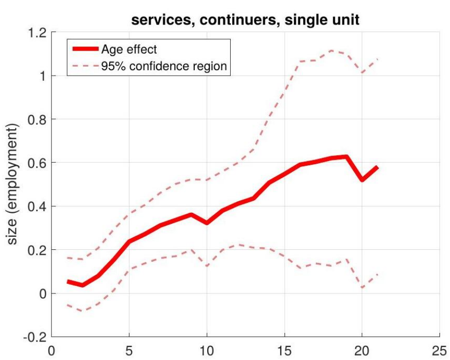
New establishments of existing multi-unit firms exhibit flat or declining size trajectories. Startups, by contrast, show the classic life-cycle growth pattern: steep early growth followed by gradual flattening. Figure 1 shows this contrast for both manufacturing and services.



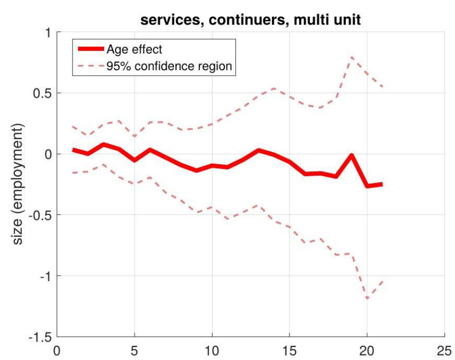
(a) Manufacturing, single-unit (startups)



(b) Manufacturing, multi-unit (new establishments)



(c) Services, single-unit (startups)



(d) Services, multi-unit (new establishments)

Figure 1: Age–size profiles for continuing establishments: startups (left) versus new establishments of existing firms (right). Source: Kueng et al. (2017), Figure 2.

This pattern is puzzling because many empirical studies do not distinguish between startups and new establishments as both are "new" and should follow the same growth dynamics. This is especially salient in the entrepreneurship literature, and its debate on whether franchising should be considered a form of entrepreneurship. Ketchen et al. (2011) argue that franchising involves genuine entrepreneurial opportunity recognition and risk-bearing by franchisees; Combs et al. (2011) counter that the standardized format leaves little room for entrepreneurial initiative. The data speak directly to this debate: replicated establishments do not exhibit the growth dynamics that characterize entrepreneurial ventures. Whatever organizational conditions produce startup growth (autonomy, experimentation, adaptation) are absent from replicated units from birth. In the language of Winter and Szulanski (2001), the replicated establishment is in the exploitation regime from day one; it never passes through exploration. This is also consistent with the sequential ambidexterity view of Nickerson and Zenger (2002), in which exploration generates innovations and opportunities that exploitation can build on to capture value.

2.3 Organizational restructuring tracks size, not age

Kueng et al. (2017) find that conditional on firm size, age has no statistically significant effect on any outcome, including centralization, performance pay, occupational layers and innovation, in either sector. A firm with 200 employees that reached that size in 5 years has the same degree of centralization as a firm with 200 employees that took 15 years. This result is identified through within-establishment variation (establishment fixed effects absorb cohort differences), so it cannot be driven by selection or survivor bias.[3]

This is inconsistent with theories that emphasize organizational maturation as a distinct process from growth. The "growing pains" literature (Genedy et al., 2024) and passive learning models such as Jovanovic (1982) predict that organizations need time to develop internal capabilities, independent of their growth rate. The data say otherwise: whatever organizational changes firms undergo, they are triggered by reaching a particular size, not by reaching a particular age. These findings potentially support the view of many VC investors that "blitz-scaling" can be useful for startups, since age by itself does not necessarily contribute to better outcomes. However, as we will analyze below, there are other theoretical reasons for premature scaling, not based on firm age alone.

[3]The age coefficients are jointly insignificant conditional on size in every specification. This is not a power issue: the age coefficients are precisely estimated and tightly centered on zero.

2.4 Innovation collapses in services but not in manufacturing

Figure 2 shows conditional size effects on innovation for both sectors. In manufacturing (left panels), innovation rises monotonically with firm size across all three types, product, process, and incremental. In services (right panels), innovation rises to approximately 15–50 employees and then declines sharply. Firms with 500+ employees are less innovative than firms with fewer than 5. The decline is parallel across all three innovation types: large service firms are less innovative across the board.

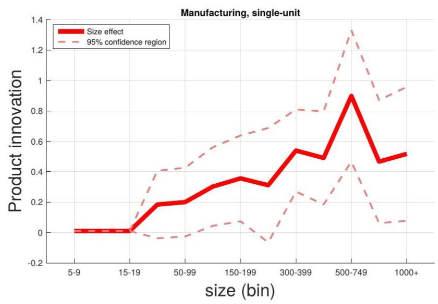
This is the most puzzling pattern in the data. Indeed, innovation theories going back at least to Schumpeter (1942) predict that large firms innovate more as they have more resources, and capture returns more easily. This holds in manufacturing. But in services, the relationship reverses: a firm with 500 employees is *less likely* to report any innovation than a firm with 3 employees. Whatever organizational changes accompany scaling in services, they do not just slow innovation but crush it.

2.5 The organizational signature of scaling

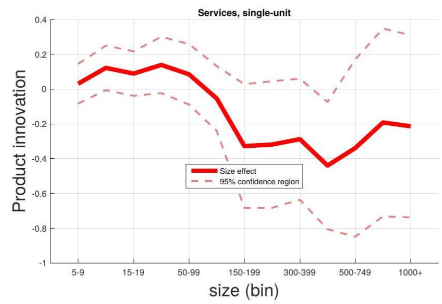
The innovation collapse in services coincides with a distinctive organizational transformation. Figure 3 shows that as service firms scale, centralization increases with size and performance pay adoption rises. Figure 4 shows that occupational layers follow an inverted U (rising initially, then delayering at large sizes) while the span of control widens (fewer managers per employee). Manufacturing firms show weaker or absent versions of these patterns.

Taken together, these patterns describe a specific organizational trajectory: as service firms grow, they centralize decision-making, adopt performance pay, eliminate middle layers, and widen the span of control. This is the organizational signature of standardization for replication, the scaling startup converges on the same organizational form as a franchise branch. It is "franchising" itself from within.[4]

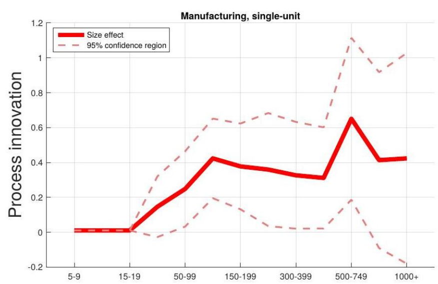
[4] Recent empirical work by Lawrence (2020) documents this template-based replication process directly in retail.



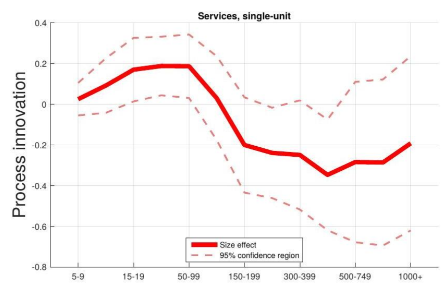
(a) Product innovation, manufacturing



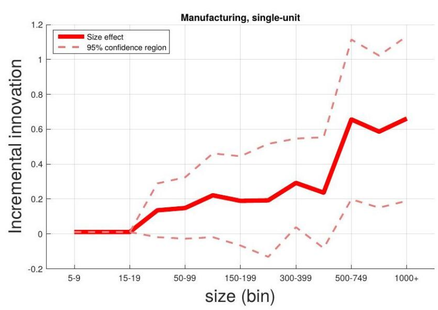
(b) Product innovation, services



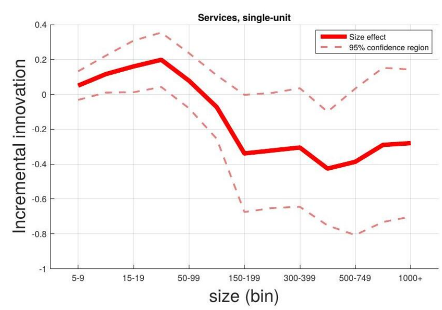
(c) Process innovation, manufacturing



(d) Process innovation, services

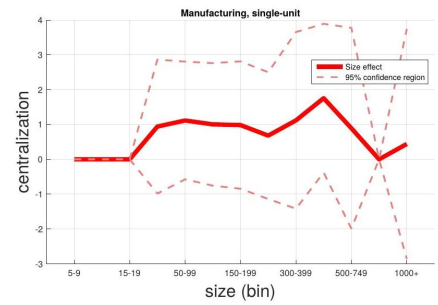


(e) Incremental innovation, manufacturing

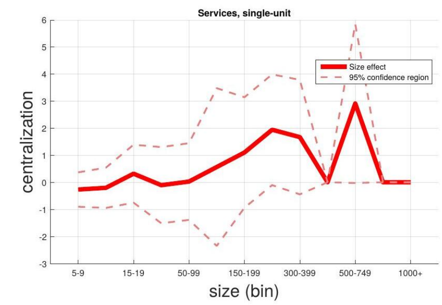


(f) Incremental innovation, services

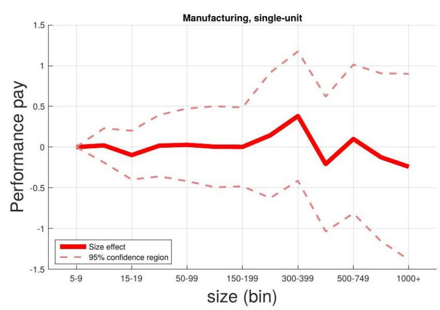
Figure 2: Conditional size effects on innovation: manufacturing (left) versus services (right). All effects are relative to establishments with 1–4 employees. Source: Kueng et al. (2017), Figure 8.



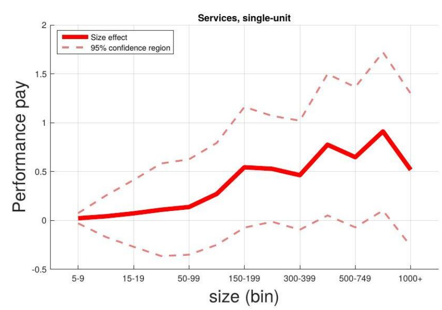
(a) Centralization, manufacturing



(b) Centralization, services

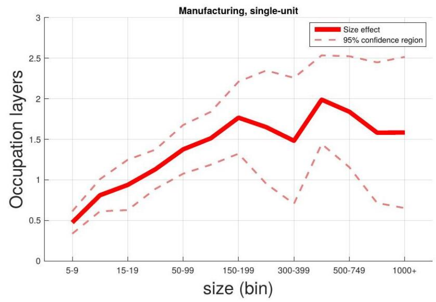


(c) Performance pay, manufacturing

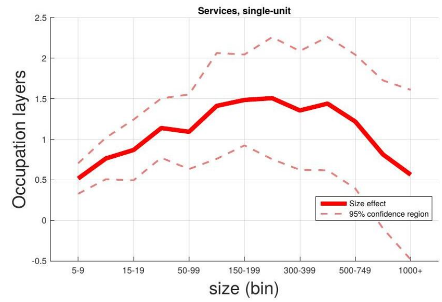


(d) Performance pay, services

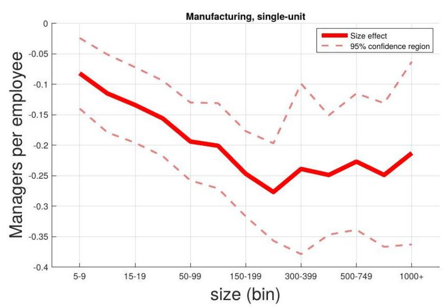
Figure 3: Conditional size effects on centralization (top) and performance pay (bottom): manufacturing (left) versus services (right). Some estimates for the largest size bins are suppressed by Statistics Canada for confidentiality but were reported to be positive and significant. Source: Kueng et al. (2017), Figures 6–7.



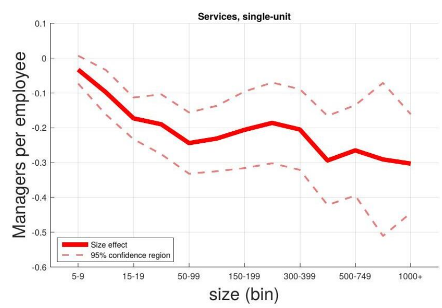
(a) Occupational layers, manufacturing



(b) Occupational layers, services



(c) Managers per employee, manufacturing



(d) Managers per employee, services

Figure 4: Conditional size effects on occupational layers (top) and span of control (bottom): manufacturing (left) versus services (right). Source: Kueng et al. (2017), Figures 6–7.

3 A Model of Scaling and Organizational Restructuring

The stylized facts call for a unified explanation. Why does the Lee (2022) hierarchy tradeoff bind so much harder in services? Why does Winter and Szulanski's (2001) replication strategy destroy innovative value in one sector but not the other? I adapt Aghion and Tirole (1994) to answer both questions with a single parameter.

3.1 The Tradeoff between Control and Initiative

I build on Aghion and Tirole (1994) and their application of the incomplete contracts framework of Grossman and Hart (1986) to innovation. On the one hand, innovation requires unobservable effort: A firm cannot verify whether a worker is actively seeking improvements, for example experimenting with a new approach to a client problem, exploring whether a process could be redesigned, testing an idea that might fail. On the other hand, innovation results are incontractible: the exact nature of an innovation cannot be specified in a contract before it is discovered. This is Arrow's (1962) fundamental paradox of information: the value of an innovation cannot be assessed before it is disclosed, but once disclosed, a potential user already has it and the basis for contracting is destroyed. For example, a consulting firm cannot contract with a consultant to 'develop a novel strategic framework', since if the firm could specify the framework, it would not need the consultant to innovate. Similarly, a software firm cannot contract with a developer to 'discover a better payment system architecture', since the improvement is unknown until the developer, working inside the system daily, notices a pattern that suggests a redesign.

What contracts can specify is the allocation of authority over whatever gets invented. Aghion and Tirole (1994) show that this allocation creates a fundamental tradeoff between initiative by the worker and control by the firm (F). Whoever has decision authority, i.e. the right to decide whether and how an innovation is deployed, will become residual claimant on the marginal returns from the invention. As a result, when the F retains authority (centralization), W has weaker incentives to exert the costly effort required for discovery. The opposite applies in the case of decentralization: when the worker (W) retains authority, effort is high because the worker captures the returns through autonomy but the firm loses control over what gets innovated and how it is used. Initiative and control are substitutes: there is no contract that elicits high effort while preserving full firm control, because the innovation that would motivate the effort cannot be described until it exists.

The degree of firm control varies continuously in practice, through the intensity of managerial oversight, the scope of codified procedures, or the fraction of tasks governed by standardized protocols, creating a governance spectrum between full decentralization and full centralization (Aghion and Tirole, 1997). I parameterize this spectrum by $\theta \in [0, 1]$, where $\theta = 0$ is full decentralization (maximum worker authority) and $\theta = 1$ is full centralization (maximum firm authority). The worker's return from innovating under governance θ is:

$$R(\theta) = (1 - \theta) \cdot r \tag{1}$$

where $r = v_I + b$ is the worker's total return under decentralization: the innovation value v_I that accrues to the firm, plus the worker's private benefit b (formalized in Section 3.2). This follows directly from the Aghion–Tirole property-rights structure: at $\theta = 0$, the worker is the residual claimant and captures r; at $\theta = 1$, the firm.

A key application of the Aghion and Tirole framework is the question of what happens when scaling forces firms to increase θ: the incontractibility of innovation means there is no way to centralize operations while preserving innovation incentives through compensation alone. The firm cannot write a contract that says "follow the manual but also innovate," because the innovation is undefined until it occurs.

Throughout this paper, θ is referred to as "centralization," but it more precisely captures the degree of firm control over how work is performed. This is consistent with Winter and Szulanski (2001), who argue that successful replication requires codifying the "Arrow core" (the essential knowledge underlying the business model) into transferable routines, and with Lawrence (2020), who shows empirically how this codification operates through templates that guide replicating units' learning. Centralization (θ) is the organizational mechanism through which this codification occurs: it is the degree to which workers follow prescribed procedures rather than exercising their own judgment. A high-θ firm is one in which workers follow the firm's preferred procedures, execute the firm's preferred projects, and conform to the firm's standardized methods. This distinction maps directly to Aghion and Tirole's (1997) concept of real authority: under high θ, the firm's preferences determine outcomes regardless of the formal allocation of decision rights. The empirical measure in the WES data ("who normally makes the decision" about production, marketing, and hiring) captures exactly this dimension: not whether the org chart assigns authority to headquarters, but whether headquarters' preferences are what actually gets implemented on the ground.

One might ask whether the model's results depend on this indirect representation of replication through centralization, rather than modeling replication explicitly. In Appendix C, I extend the model to allow the firm to open M replicated establishments, each earning profit proportional to the degree of standardization θ, subject to convex expansion costs. The extension yields a closed-form solution with the same structure as the baseline.[5] The baseline model is also the appropriate specification for the empirical exercise, since the WES data measure firm size within the first establishment rather than summing employment across all establishments of the same firm. The startup growth patterns documented in Section 2 therefore reflect the organizational transformation of the original unit, which is precisely what the baseline model describes.

[5] All comparative statics are preserved, and two new predictions emerge: firms with more profitable replication opportunities centralize more, and the optimal number of establishments rises with firm size through $\theta^*(n)$.

3.2 Initiative and Innovation

To formalize the Aghion-Tirole tradeoff in a scaling venture, consider a firm of size n in which each worker i performs operational tasks and may also exert effort that produces a probability e_i of generating an innovation. Following Aghion and Tirole (1994), innovations are ill-defined ex ante and effort is unobservable. The cost of achieving innovation probability e_i is $c(e_i) = \frac{k}{2}e_i^2$, where $k > 0$ captures the difficulty of innovating in the firm's environment. The worker's return from innovating under degree of centralization θ is $R(\theta) = (1 - \theta) \cdot r$, where $r = v_I + b$: the sum of the innovation value v_I (value from new offerings, cost savings from more efficient processes) and the worker's private benefit b (intrinsic interest, professional reputation, job satisfaction). Both components motivate effort, but only v_I accrues to the firm as profit. The private benefit b also drives project selection: under decentralization, workers choose projects maximizing $v_I + b$, which may differ from projects maximizing v_I alone.

The worker chooses innovation probability e to maximize expected returns net of cost:

$$\max_e \; e \cdot (1 - \theta) \cdot r - \frac{k}{2}e^2 \tag{2}$$

yielding:

$$e^*(\theta) = \frac{(1 - \theta) \cdot r}{k} \tag{3}$$

Innovation probability is decreasing in centralization ($de^*/d\theta = -r/k$) and decreasing in the difficulty of innovation ($de^*/dk < 0$). Since e^* is a probability, I require $e^*(0) = r/k \leq 1$, i.e., $k \geq r$.

3.3 Separability between Operations and Innovation

The empirical evidence in section 2 suggested a key difference between manufacturing and services. The critical difference is whether the workers who deliver the output are the same workers who would innovate. In manufacturing, innovation is performed by specialized workers (engineers, designers, R&D scientists) in organizationally distinct units (labs, prototyping facilities). The firm can centralize production while leaving R&D decentralized, i.e. two separate organizational structures for two separate activities. In services, innovation is embedded in service delivery: the consultant who develops a new framework, the chef who experiments with a dish, the logistics coordinator who discovers a better route. These workers cannot be split into an "operations division" and an "innovation division" because both activities are performed by the same person in the same task (Vargo and Lusch, 2004).

I capture this with a separability parameter $\sigma \in \{0, 1\}$. When $\sigma = 0$, innovation effort depends on a separate organizational instrument θ_{innov} that the firm can set independently of θ_{ops}:

$$\theta_{\text{innov}} = \sigma \cdot \theta_{\text{ops}} \tag{4}$$

Manufacturing ($\sigma = 0$): The firm sets θ_{ops} and θ_{innov} independently. R&D can stay decentralized ($\theta_{\text{innov}} = 0$), so innovation is independent of centralization. In manufacturing, the productive knowledge needed to design a product is distinct from the productive knowledge needed to assemble it: process development and production occupy different organizational locations, typically with different personnel, and process innovations are transferred from development facilities into manufacturing once a process is ready for volume production (Hatch and Mowery, 1998).[6]

Services ($\sigma = 1$): $\theta_{\text{innov}} = \theta_{\text{ops}} = \theta$. The firm has one instrument for two activities. In professional and personal services, the expertise that generates novelty is the same expertise that delivers the service to the customer. There is no R&D unit standing apart

[6] Hatch and Mowery (1998) take separability as their organizational starting point: they analyze the *transfer* of process technology from R&D to manufacturing facilities, and document frictions in that transfer: some learning-by-doing knowledge is specific to the production environment where the process was developed, so "some knowledge is effectively lost" when the process moves to manufacturing. The natural way to capture this in the present framework would be to relax $\sigma \in \{0, 1\}$ to $\sigma \in (0, 1)$: manufacturing sits near, but not exactly at, $\sigma = 0$, with the residual coupling reflecting context-specific knowledge that is lost when development and production are organizationally separated. For analytical clarity I retain the binary specification in what follows, but the qualitative results are preserved for any interior $\sigma_M < \sigma_S$.

from delivery because the locus of relevant knowledge is the frontline workforce itself (von Nordenflycht, 2010).[7] Centralizing how that workforce performs its work (through scripts, protocols, standardized workflows, or algorithmic dispatch) therefore directly constrains the discretion through which service innovations emerge. The single instrument θ in the services case is not a modeling shortcut; it reflects the fact that, in these settings, the firm has no organizational technology that separately disciplines delivery while preserving the experimental latitude of the people who perform it.

3.4 Firm-Level Innovation

While each worker independently generates innovations with probability e^*, the firm's total innovation output exhibits diminishing returns to scale. Not every worker-level insight translates into a distinct firm-level innovation: ideas overlap, the easiest improvements are found first, and organizational capacity to absorb innovations is limited. Using the separability parameter σ defined above, I model firm-level innovation as:

$$I(n, \theta, \sigma) = n^{\zeta} \cdot e^*(\sigma\theta) = \frac{n^{\zeta}(1 - \sigma\theta)(v_I + b)}{k} \tag{5}$$

where $\zeta \in (0, 1)$. When $\sigma = 1$ (services), centralization reduces innovation: $I = n^{\zeta}(1 - \theta)(v_I + b)/k$. When $\sigma = 0$ (manufacturing), innovation is independent of centralization: $I = n^{\zeta}(v_I + b)/k$. The parameter ζ captures firm-level diminishing returns to innovation headcount: doubling the number of workers less than doubles the number of distinct innovations. This is consistent with the knowledge production function literature (Jones and Williams, 1998), where R&D output typically shows diminishing returns to the number of researchers, sometimes also called a "congestion" or "stepping on toes" effect.

[7] von Nordenflycht (2010) argues that in knowledge-intensive firms, expertise resides not only in executive or support functions but "also among its 'frontline workers' ". This person-centric locus of expertise is precisely what rules out a separable R&D function in services: the consultant, the surgeon, and the chef cannot be split into an operations role and an innovation role because both activities are performed by the same person in the same task.

3.5 Model overview and timing

I model the firm's lifecycle as three strategic stages, where each stage may correspond to 5–10 years of real time. Figure 5 summarizes the timing structure. In each stage, the firm observes its size n_t (Row 1), which determines operating profits Π_O (Row 2). Together with predetermined innovation profit $v_I \cdot I_{t-1}$, these form total profits Π_t^{total} (Row 3), which enter the present value J_t (Row 4). The firm chooses centralization θ_t^* to maximize J_t (Row 5), and this choice determines how many innovations I_t are generated in stage t (Row 6). Innovation returns accrue one stage later: the firm captures v_I per innovation in the following stage, while the private benefit b accrues to the worker at the time of innovation. Innovations generated at stage t are:

$$I_t = \frac{n_t^{\zeta}(1 - \sigma\theta_t)(v_I + b)}{k}, \tag{6}$$

and the associated profit $v_I \cdot I_t$ enters the firm's total profits in stage $t+1$ alongside operating profits. I define total profits as:

$$\Pi_t^{\text{total}} = \Pi_O(n_t, \theta_t) + v_I \cdot I_{t-1}, \tag{7}$$

where $v_I \cdot I_{t-1}$ is predetermined at stage t (generated in stage $t-1$). For the first stage, I_0 denotes an initial innovation endowment that funds profit at stage 1. Total profits serve as both the measure of firm performance and the source of funds for growth.

In the terminal stage (Stage 3 of Figure 5), innovations are worthless because there is no future stage in which they would pay off. This is more than an analytical convenience. Each stage spans years, and the value of innovations generated late in the firm's lifecycle is genuinely lower for at least two reasons. First, founders and managers making decisions today heavily discount any future in which they no longer lead the firm, a founder at stage 3 who expects to retire or sell has little reason to invest in innovations whose returns she will not capture. Second, as the firm's market and product mature, the frontier of profitable innovation recedes: the easy improvements have been made, the client base is established,

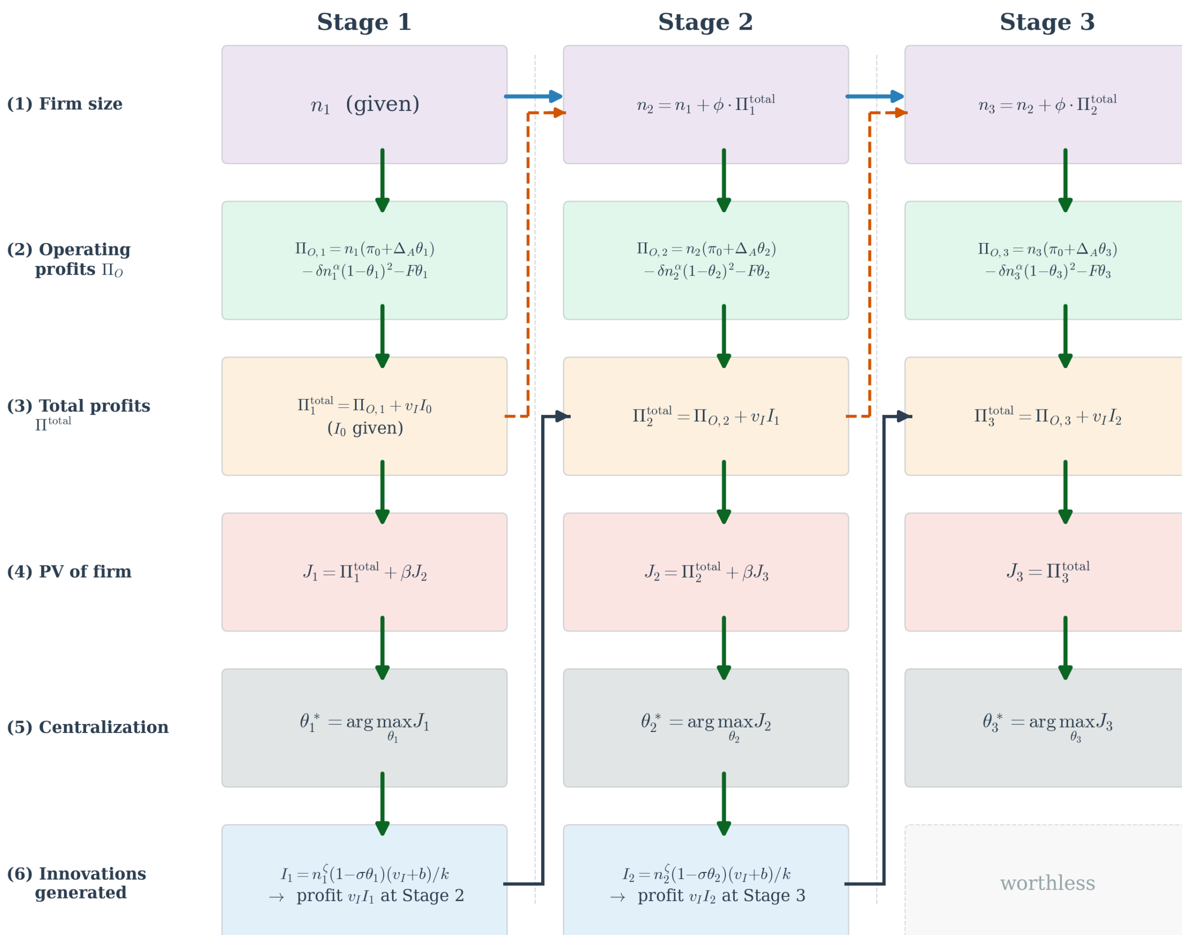


Figure 5: **Timing of the three-stage model.** Row (1): firm size n_t is a state variable (n_1 given). Row (2): operating profits Π_O depend on n_t and θ_t. Row (3): total profits $\Pi_t^{\text{total}} = \Pi_{O,t} + v_I I_{t-1}$ add predetermined innovation profit (I_0 given as initial endowment). Row (4): present value $J_t = \Pi_t^{\text{total}} + \beta J_{t+1}$. Row (5): the firm chooses θ_t^* to maximize J_t. Row (6): innovations I_t generated at stage t depend on n_t, θ_t, and σ; Stage-3 innovations are worthless (dashed box) because they would pay out at a non-existent Stage 4. Within each stage, the causal chain runs $n \to \Pi_O \to \Pi^{\text{total}} \to J \to \theta^* \to$ innovations. Across stages, innovations I_t generated at stage t pay out as profit $v_I \cdot I_t$ in Π_{t+1}^{total} at stage $t{+}1$, part of which is used to fund growth to n_{t+2} via retained earnings (innovation growth channel). Both channels exhibit a one-stage delay from profit to growth, but the innovation channel responds to centralization with a one-stage lag: θ_t controls Π_O at stage t contemporaneously, while it affects innovation profit only at stage $t{+}1$ via $v_I \cdot I_t$.

and the remaining innovations are increasingly marginal. The terminal-stage assumption captures both forces in reduced form.

Because innovation returns are delayed, the firm's willingness to decentralize depends on how much it values the future, captured by the discount factor $\beta \in (0, 1]$. With each stage spanning roughly 7 years, $\beta = (1 + \beta_A)^{-7}$ where β_A is the annual discount rate. Three forces determine β_A, and therefore β, in practice.

First, *VC investor time horizons.* Early-stage VCs target annual IRRs of 20–25% (Lerner and Leamon, 2023), implying $\beta_A = 0.20$–0.25 and therefore $\beta \approx 0.2$–0.3 per stage. At these rates, a dollar of innovation profit arriving one stage later is worth only 20–30 cents today. The VC's carried interest is based on exit proceeds, which reflect operating profits at the time of exit, not innovations that will generate profit after the exit. This creates a structural bias toward centralization: the VC-backed firm underweights the innovation returns that decentralization would produce.

Second, *founder tenure.* Azoulay et al. (2020) show that the mean age of founders of the fastest-growing new ventures is 45 (and approximately 42 for new firms overall). One extreme but clarifying assumption is that such founders may be relatively patient when it comes to the second stage of their ventures, i.e. over the next 20 years, but may heavily discount what happens to their ventures afterwards. This would fit a setup, in which they discount annual returns at a 3% annual rate that corresponds to the long-run average bond return for US treasuries, which implies a stage-wise discount factor of $\beta = (1.03)^{-7} \approx 0.8$. Additionally, since even this very patient founder will have to retire eventually[8], she will not care about profits afterwards, consistent with our 3-stage payoff assumptions.

Third, *capital market short-termism.* Sampson and Shi (2023) document that implied discount rates in U.S. public markets have risen substantially over 1980–2013. This matters for scaling-up startups pursuing an IPO to finance growth opportunities and young firms

[8] Indeed, founders are likely to experience higher stress levels than corporate CEOs, since the extent of idiosyncratic risk at startups is much higher than at diversified corporations. In this context, Borgschulte, Guenzel, Liu, and Malmendier document that stress significantly reduces the life expectancy of CEOs.

pursuing growth equity. Using the long-run average S&P 500 return as a benchmark, $\beta_A \approx 0.10$ gives $\beta \approx 0.51$: the public market values Stage-2 innovation returns at roughly half their face value. But operating profits are contractible and collateralizable while future innovation returns are uncertain and intangible, so capital-constrained firms effectively face a β_A for the innovation component that exceeds the market average, pushing β below 0.5 and creating a systematic bias toward centralization.

All examples provide a wide degree of variation for patience implied by ownership structure. The sixth and seventh rows of Figure 5 aggregate operating profits and innovation returns into current profits and the present value J_t of the firm. I now develop each remaining component formally, following the rows of the figure.

3.6 Operating Profits

Row (2) of Figure 5 shows operating profits Π_O, which accrue in the current stage and combine two benefits of centralization.

Standardization and Alignment What are the benefits of centralization? Under decentralization ($\theta = 0$), workers choose projects that maximize their total return $v_j + b_j$ from the set of available projects j. Under centralization ($\theta = 1$), the firm assigns projects that maximize firm value v_j. These project choices generally differ because the worker's private benefit b_j varies across projects: some projects are commercially valuable but routine (v_j high, b_j low); others are intellectually stimulating but less profitable (v_j moderate, b_j high).

Let v^W denote the firm value of the worker-preferred project (the project maximizing $v_j + b_j$) and v^F denote the firm value of the firm-preferred project (the project maximizing v_j). The alignment gap is:

$$\Delta_A \equiv v^F - v^W \geq 0 \tag{8}$$

This gap is non-negative because the firm-preferred project maximizes v_j by definition, so $v^F \geq v^W$. It is strictly positive whenever the worker's private benefit b varies across projects, whenever the worker would choose a different project than the firm would.[9]

Under intermediate governance $\theta \in (0, 1)$, the worker's effective project choice shifts continuously from the worker-preferred to the firm-preferred project. The firm value per unit of effort under governance θ is:

$$v(\theta) = v^W + \Delta_A \cdot \theta = v^W + (v^F - v^W)\theta \tag{9}$$

The operational gain from centralization is a specific causal mechanism: centralization redirects workers from private-benefit b driven projects to firm value driven projects that maximize v_j. The gain Δ_A is the firm-value difference between the project the firm would choose and the project the worker would choose.

Span of Control Centralization reduces coordination costs. As in Lucas (1978), the firm's owner has finite attention. Each autonomous worker generates decisions that require managerial oversight, such as resolving conflicts with other workers' choices, ensuring consistency with the firm's commitments, integrating individual adaptations into a coherent offering. The cost of this oversight grows in the number of autonomous workers because managerial attention is a fixed resource allocated across an increasing number of decisions (Levinthal and Wu, 2010). I model coordination costs as:

$$C(n, \theta) = \delta \cdot n^\alpha \cdot (1 - \theta)^2 \tag{10}$$

where $\delta > 0$ and $\alpha > 2$. Under centralization, workers follow directives rather than making autonomous choices, so the demand on managerial attention falls, the $(1-\theta)^2$ term captures this. The manager's job shifts from coordinating diverse autonomous decisions to monitoring

[9] Appendix A.8 works out a project-selection microfoundation underlying this reduced form. The microfoundation reveals an implicit upper bound: $\Delta_A \leq b - b_{j^F} \leq b$, where b is the worker's private benefit on their preferred project and $b_{j^F} \geq 0$ is their private benefit on the firm-preferred project. Equivalently, $\Delta_A < b$ is necessary for the alignment problem to exist: if $\Delta_A \geq b$, the worker would prefer the firm's project anyway and the principal–agent gap vanishes. This restriction is used to identify the sector-specific calibration in Appendix B.14.

compliance with a uniform standard, which is less costly. The dependence on n^α with $\alpha > 2$ reflects that coordination costs disproportionately increase in the number of workers: doubling the workforce more than doubles the coordination burden on the manager. This arises because coordination is fundamentally a combinatorial problem, each autonomous worker's decisions must be reconciled not only with the manager's preferences but with every other worker's choices, generating interactions that grow faster than linearly. These costs are an *operating* cost of the current workforce, not a cost of hiring: a firm that inherited 500 workers from past growth faces the coordination problem of managing 500 autonomous decision-makers today, regardless of when those workers were hired. This is why n functions as a state variable in the model: the coordination problem is determined by the firm's current size, which is itself the cumulative outcome of past centralization, profit, and growth decisions.

Overhead cost of centralization. Centralization itself requires infrastructure: operating manuals, monitoring systems, training programs, compliance procedures. These costs scale with the *degree* of centralization θ but not with firm size n, the manual is the same whether the firm has 10 or 500 workers. I model this as a linear cost $F\theta$, where $F > 0$. The per-worker overhead cost $F\theta/n$ vanishes as the firm grows, so centralization becomes cheaper per worker at scale. This creates a natural size threshold below which centralization is not worth the overhead, even absent any innovation consideration.

Combined operating profits:

$$\Pi_O(n, \theta) = n[\pi_0 + \Delta_A\theta] - \delta n^\alpha(1-\theta)^2 - F\theta \tag{11}$$

where $\delta > 0$ and $\alpha > 2$. Operating profits are affected by centralization through three channels: the alignment gain ($\Delta_A\theta$) which is the first benefit of centralization, the reduction of coordination costs ($(1-\theta)^2$), which is the second benefit of centralization. And the centralization overhead cost ($F\theta$) which is the main direct cost of centralization. For large firms the first two dominate; for small firms the overhead cost may outweigh the alignment gain. Additionally, the main indirect cost of centralization will be reduction in initiative and related innovative ideas by workers.

3.7 Growth and Firm Value

Firm size evolves by reinvesting total profits, which include both operating profits and innovation profits:

$$n_{t+1} = n_t + \phi \cdot \Pi_t^{\text{total}} = n_t + \phi \cdot \left[\Pi_O(n_t, \theta_t) + v_I \cdot I_{t-1}\right], \tag{12}$$

where $\phi > 0$ is the reinvestment rate and $v_I \cdot I_{t-1}$ is the innovation profit accruing in stage t from innovations generated in stage $t-1$. The firm grows through two channels: on the one hand, the *replication channel*, in which standardizing operations raises operating profits (Π_O), part of which is retained to fund hiring. On the other hand, the *innovation channel*, in which past innovations produce profit $v_I \cdot I_{t-1}$ that is also retained to fund hiring. Both channels feed into the same growth equation with a one-stage delay from profit to headcount.

Operating profits Π_O are net of all factor costs, including wages: π_0 represents the per-worker surplus after compensating workers at their market rate, so that Π_O is the residual cash flow available for reinvestment. Under this interpretation, the cost of hiring an additional worker is absorbed into π_0, and the growth equation states that the firm expands its workforce at a rate proportional to its net cash flow. The parameter ϕ converts retained earnings into headcount, reflecting the capital–labor complementarity in growing firms: higher profits fund both the equipment and the hiring that expansion requires.

A key modeling choice is that firm size n is a state variable, not a within-period choice. The firm does not freely select its workforce each stage in the model; instead, n accumulates over time as the outcome of past profit, innovation, and centralization decisions. This captures the idea that startup hiring is constrained by cash flow rather than by the labor market: a firm that earns more can hire more, but it cannot instantly jump to an arbitrary size. Treating n as a state variable rather than a static choice keeps the direct organizational link between firm size and centralization visible in the closed-form solution, rather than absorbing it into a deeper capital-stock or productivity primitive: standard operating procedures make it more effective to coordinate many workers, a key idea in Winter and Szulanski (2001). Because hiring is constrained by cash flow, firm size summarizes the cumulative history of past profit, innovation, and centralization decisions, and the organizational transformation is a gradual process that firms pass through as they grow. This is what generates the lifecycle dynamics and what allows the model to match panel data in which the same establishments are tracked over time.

Present value. Row (4) of Figure 5 shows the present value J_t of the firm at each stage. The firm maximizes total profits over all stages by backward induction:

$$J_t = \Pi_t^{\text{total}} + \beta \cdot J_{t+1} = \Pi_O(n_t, \theta_t) + v_I \cdot I_{t-1} + \beta \cdot J_{t+1} \tag{13}$$

Since $v_I \cdot I_{t-1}$ is predetermined, the firm's choice of θ_t affects J_t only through $\Pi_O(n_t, \theta_t)$ (directly) and through the innovations I_t generated this period, which will contribute $v_I \cdot I_t$ to J_{t+1} (with one-stage delay, discounted by β). This creates a fundamental timing asymmetry between the two growth channels. Both channels exhibit the same one-stage delay from profit to growth: the operating profit $\Pi_O(n_t, \theta_t)$ at stage t funds headcount n_{t+1}, and likewise the innovation profit $v_I \cdot I_{t-1}$ at stage t funds the same n_{t+1}. The asymmetry is upstream: θ_t controls $\Pi_O(n_t, \theta_t)$ contemporaneously, but it cannot affect $v_I \cdot I_{t-1}$, which is already set by the prior period's centralization choice θ_{t-1}. The innovation channel responds to centralization with a one-stage lag while the replication channel responds immediately. The firm's centralization decision trades off the immediate replication benefit (amplified by the growth multiplier Λ) against the lagged innovation benefit, which is discounted by β and, at any stage before the penultimate one, re-amplified by the following stage's growth multiplier, because innovation profits, once they arrive, are themselves reinvested (Appendix A.2 develops this recursion exactly).

4 Main Results

4.1 The lifecycle trajectory: boundary conditions of the theory

The model speaks to a specific empirical phenomenon: ventures born small and decentralized, growing through an interior regime, ending fully centralized. For this trajectory to be the model's prediction (as opposed to, say, firms being born already centralized or never reaching centralization), two substantive conditions on primitives and initial conditions must hold. I refer to these as the *boundary conditions* of the theory: the parametric region in which the model speaks to the empirical phenomenon. Establishing the lifecycle trajectory first sets the scope within which the subsequent propositions on the level of centralization, its drivers, and its implications for innovation are to be read.

Proposition 1 (Lifecycle Trajectory)**.** *Under the following boundary conditions, the firm's equilibrium centralization path traverses three regimes (full decentralization, interior, full centralization) in order:*

(B1) Born decentralized. *The firm's initial size n_1 does not exceed the full decentralization threshold: $n_1 \leq n_D^*$, where n_D^* is the largest size at which $\theta^* = 0$ (defined formally in Appendix A.5).*

(B2) Eventually saturating. *By the terminal stage, the firm has grown to $n_3 \geq \bar{n} = F/\Delta_A$, the size at which full centralization becomes optimal at the terminal stage.*

(B3) Interior passage. *The intermediate stage falls in the interior regime: $n_D^* < n_2 < n_C^*$, where n_C^* is the full centralization threshold (Appendix A.4). In a model with many* *short stages this would follow from (B1)–(B2) by continuity of the growth path; with three discrete stages it is a separate restriction on the growth rate, ruling out paths so slow that the firm is still at the decentralized corner at Stage 2 ($n_2 \leq n_D^*$) and paths so fast that it skips the interior regime entirely ($n_2 \geq n_C^*$).*

Under (B1)–(B3), the firm starts at $\theta_1^* = 0$, passes through an interior regime where $\theta_t^* \in (0, 1)$ is strictly increasing in n_t, and reaches $\theta_3^* = 1$ at the terminal stage (proof in Appendix A.6).

The three boundary conditions describe what the firm must look like for the model's empirical predictions to apply. (B1) requires the firm to be a small venture at birth, with the per-worker overhead F/n_1 and innovation force $n_1^{\zeta-1}$ both large enough that centralization is unprofitable. This condition naturally describes startups (born with 1–10 employees) but fails for new establishments of multi-unit firms, which are born at the parent firm's already-centralized θ. (B2) and (B3) jointly require the firm's growth dynamics to be vigorous enough that it traverses the entire interior regime within three stages and reaches the saturation threshold, but not so explosive that it skips the interior regime. These conditions rule out lifestyle businesses, which never grow beyond a particular size, stagnant ventures whose cumulative profit growth is insufficient, and (at the fast end) ventures that jump directly to franchise-like scale. When (B1)–(B3) hold, the firm follows the empirically observed pattern: a high-initiative phase in the decentralized regime, an organizational transition in the interior regime, and a mature replication phase in the saturated regime. The boundary conditions therefore delineate the model's scope: scaling ventures that are starting small and growth over their lifecycle.

4.2 Optimal centralization rises with firm size

For scaling ventures the assumption of an exogenous organizational structure is most questionable since organizational structure is malleable. My theory provides one formal way to understand organizational restructuring over the firm life cycle, consistent with the stylized facts of section 2:

Proposition 2 (Centralization and Firm Size)**.** *Given regularity conditions*[10]*, the optimal degree of centralization θ^* is uniquely determined by the first-order condition (Appendix A.3). At interior solutions:*

(a) Centralization is strictly increasing in firm size: $\partial\theta^*/\partial n > 0$.

(b) Firm age does not enter the first-order condition: conditional on size, two firms of different ages choose the same θ^*.

For firms that anticipate being large enough that $\theta_3^* = 1$ at the terminal stage, the growth multiplier $\Lambda(n_3) = 1 + \beta\phi \cdot dJ_3^*/dn_3$ reduces to the parameter $\Lambda = 1 + \beta\phi(\pi_0 + \Delta_A)$, and at the penultimate stage θ^* admits the closed form:

$$\theta^*(n) = \left[1 - \frac{\beta v_I(v_I + b)n^{\zeta-1}/(k\Lambda) \;-\; \Delta_A \;+\; F/n}{2\delta n^{\alpha-1}}\right]_0^1, \tag{14}$$

where $[\cdot]_0^1$ denotes clamping to $[0, 1]$.[11] *For manufacturing ($\sigma = 0$),*

$$\theta_M^*(n) = \left[1 - \frac{F/n \;-\; \Delta_A}{2\delta n^{\alpha-1}}\right]_0^1. \tag{15}$$

A single mechanism, replication killing initiative, simultaneously determines organizational structure, innovation, and growth along a unique equilibrium path. As the firm grows, several forces jointly push toward centralization: coordination costs rise (δn^α), the

[10]The uniqueness statement requires the second-order condition $\partial F/\partial\theta < 0$ to hold at the optimum. In the empirically relevant path with terminal-stage $\theta_3^* = 1$, this holds unconditionally because the future value function becomes linear in size and the reinvestment feedback vanishes. In the interior regime, under the boundary conditions of Proposition 1, a sufficient condition is the boundary-restricted form $n_D^* > n^{**}$, where the regularity threshold n^{**} is given explicitly in primitives in Appendix A.3. A more general condition $n_1 > n^{**}$ that does not require (B1) is also derived there.

[11]The comparative statics (a) and (b) are established directly from the first-order condition via the implicit function theorem (Appendix A.3), without requiring the closed form. The closed form (14) is the exact penultimate-stage policy under terminal-stage saturation; the Stage-1 policy takes the same form with the innovation weight β/Λ replaced by $\beta\Lambda/\Lambda_1$, where $\Lambda_1 > \Lambda$ is the Stage-1 growth multiplier, which preserves every comparative-static sign (Appendix A.4).

per-worker value of initiative falls ($n^{\zeta-1}$ with $\zeta < 1$), and overhead costs are amortized (F/n declines). Result (b) implies that organizational restructuring tracks size, not age. This endogenization also reveals a cost of replication not discussed by Winter and Szulanski (2001): centralization does not merely risk replicating an insufficiently understood business model, but it directly kills the initiative that generates innovation.

Several implications follow. First, the model provides a structural explanation for why organizational "growing pains" concentrate in the scaling zone between n_D^* and n_C^*: below the full decentralization threshold n_D^* startups exhibit full decentralization, and above the full centralization threshold n_C^* they have restructured toward $\theta^* = 1$. The contested region between these two thresholds (where the innovation–replication tradeoff is sharpest) is precisely where founders report the most organizational upheaval (DeSantola and Gulati, 2017). Second, the closed-form solution makes precise a claim that is often invoked informally: that startups are "naturally" flat (Lee, 2022). In the model, small firms are decentralized not because founders prefer flat structures, but because the overhead cost F/n and the innovation force $n^{\zeta-1}$ make centralization too expensive. Flatness is a consequence of size, not culture. Third, the age-irrelevance result offers a sharp empirical test: after controlling for size, age should have no additional predictive power for organizational structure. This distinguishes the model from lifecycle theories (Jovanovic, 1982) in which startup growth is driven by "passive learning" and age effects.

4.3 Drivers of Centralization during Scaling

The next proposition characterizes how the model's three central primitives shape centralization. Each primitive operates simultaneously through two channels: how aggressively the firm centralizes in the interior regime, and the size at which it first leaves the fully decentralized regime. The two manifestations express a single underlying mechanism: a primitive that makes centralization more attractive at any given size both raises θ^* where the firm is interior and brings forward the size at which centralization begins. Stating both manifestations together makes this unity explicit.

Proposition 3 (Drivers of Centralization)**.** *The degree of centralization and the timing of its onset are jointly governed by three primitives, each shaping both the interior centralization level θ^* and the full decentralization threshold n_D^*:*

(a) ***Ownership patience*** *(β): Patient owners centralize less aggressively in the interior regime ($\partial\theta^*/\partial\beta < 0$) and stay fully decentralized to larger sizes ($\partial n_D^*/\partial\beta > 0$).*

(b) ***Separability of operations and innovation*** *(σ): Services firms are less centralized than manufacturing firms of the same size ($\theta_S^*(n) < \theta_M^*(n)$ at any size where at least one sector's policy is interior, the two coinciding at the shared corners, with $\theta_S^* \to \theta_M^*$ as $\beta \to 0$) and stay fully decentralized to larger sizes ($n_{D,S}^* > n_{D,M}^*$).*

(c) ***Incentive misalignment*** *(Δ_A): Firms with greater misalignment centralize more in the interior ($\partial\theta^*/\partial\Delta_A > 0$) and exit full decentralization at smaller sizes ($\partial n_D^*/\partial\Delta_A < 0$).*

Proofs of the interior comparative statics are in Appendix A.3; proofs of the threshold comparative statics are in Appendix A.5.

Result (a) establishes that patience of ownership is a critical influence factor on organizational design. A VC-backed firm ($\beta \approx 0.2$) and a founder-led firm ($\beta \approx 0.7$) facing the same coordination problem adopt different organizational structures because they weight the future innovation stream differently. Patient owners preserve initiative because they value future innovations generated by this initiative. This shows up in two ways: in the interior regime they hold θ^* lower at any given size, and in the venture regime they stay fully decentralized to larger sizes. Impatient owners do the opposite: they centralize more aggressively when interior, and exit full decentralization earlier. This result sheds new light on the question of premature scaling. In traditional models of age effects, such as Jovanovic (1982), premature scaling by VCs is a consequence of not waiting long enough, since time directly leads to more information about the potential of a startup. However, my evidence in section 2, shows that age has no independent influence on innovation or organizational structure, once size is controlled for. This finding by itself suggests that aggressive scaling using VC funding can be effective for startups. More recent empirical work on premature scaling such as Lee and Kim (2024) finds that VC-backed startups that scale early are more likely to fail. Lee and Kim (2024) interpret this through the informational lens of insufficient experimentation. My model identifies a complementary channel that operates through organizational structure. Since VCs are relatively impatient investors (with 5–7 year time horizons) they push startups toward centralizing more aggressively and exiting the decentralized regime at smaller sizes, because the discounted value of future innovation is small. The resulting decline in innovation is consistent with lower experimentation, which implies insufficient information before making irreversible investments as in the Lee and Kim (2024) framework.

Result (b) identifies the main source of the manufacturing–services difference in the life-cycle path of innovation. In the model, the two sectors share the same workers, costs, and overhead; the only difference is whether centralizing operations also centralizes innovation ($\sigma = 1$ for services, $\sigma = 0$ for manufacturing). This single binary parameter generates the divergent innovation trajectories documented in the data: hump-shaped and ultimately declining in services, monotonically increasing in manufacturing. It also generates a difference in the timing of centralization onset: services ventures stay fully decentralized to larger sizes than manufacturing ventures because the innovation channel gives the small services firm an additional reason to remain decentralized that the manufacturing firm does not have. The mechanism is straightforward: in manufacturing, a firm can centralize its production processes, e.g. standardizing assembly lines, imposing quality control, writing operational manuals, without affecting the R&D team, which continues to operate autonomously. In services, there is no such separation. The consultant who follows a standardized methodology is the same person who would have developed a novel approach for the client. The customer service representative who follows the script is the same person who would have improvised a creative solution. Centralizing the delivery of the service necessarily centralizes the innovation process, because they are performed by the same people. This logic extends naturally to platform businesses: when Uber standardizes its driver protocols and customer interfaces across markets, it is centralizing the very activities through which local experimentation and adaptation would have occurred. The myopic limit ($\beta \to 0$) provides a sharp test of this mechanism. A service firm that does not value future innovation has no reason to protect initiative, as the only benefit of decentralization is the innovation it generates, and a myopic firm discounts that to zero. Such a firm organizes exactly like a manufacturing firm, centralizing purely on the basis of incentive alignment. The sectoral difference in organizational structure, in both the interior level and the timing of onset, is therefore entirely driven by the interaction of separability and patience: it exists only when firms are forward-looking enough to value the initiative that non-separability would destroy.

Result (c) identifies the organizational force pulling toward centralization. When Δ_A is large, the gap between what the firm wants and what workers would choose is wide, making centralization more valuable. This operates through two channels: a direct effect (higher Δ_A increases the marginal alignment gain $\partial\Pi_O/\partial\theta$) and an indirect effect through the growth multiplier (higher Δ_A raises Λ, amplifying the return to centralization through growth). Both channels reinforce: firms with greater misalignment centralize more in the interior regime and exit full decentralization at smaller sizes. The alignment parameter Δ_A captures something fundamental about the nature of the firm's innovation process. In some ventures, workers left to their own devices would pursue projects closely aligned with firm value, a machine learning engineer experimenting with model architectures at a deep-tech startup, or a chef testing new dishes at a restaurant where culinary creativity *is* the product. For these firms, Δ_A is small: the principal loses little by letting the agent choose, because the agent's preferred project is nearly as valuable as the principal's. Centralization buys little alignment and costs a great deal of initiative. In other ventures, autonomous workers would pursue projects that serve their own interests at the expense of firm value, a sales representative who builds personal client relationships rather than following the firm's account management system, or a consultant who develops idiosyncratic methods rather than the firm's proprietary framework. For these firms, Δ_A is large: the gap between the worker's preferred project (which builds the worker's human capital and outside options) and the firm's preferred project (which builds the firm's replicable systems) is wide. Centralization is valuable precisely because it closes this gap. This has a direct implication for which startups should centralize as they scale. The model predicts that centralization is least costly, and therefore most delayed, in ventures where worker initiative is naturally aligned with firm objectives. These are ventures where the product *is* the worker's creativity: research-driven biotechs, design studios, specialized consulting practices. Centralization is most beneficial, and therefore adopted earliest, in ventures where scaling requires workers to subordinate their own judgment to the firm's systems: franchise operations, standardized service delivery, logistics platforms. The practical implication for founders is that the case for centralization depends not just on firm size or investor patience, but on the nature of the agency problem within the firm. A founder whose workers are naturally aligned (Δ_A small) should resist centralization longer than conventional scaling advice suggests, because the innovation cost is high relative to the alignment gain. A founder whose workers would pursue divergent objectives (Δ_A large) should centralize earlier, because the alignment gain justifies the innovation sacrifice. Misdiagnosing Δ_A (assuming large misalignment where little exists, or vice versa) leads to organizational errors that compound over the growth path.

4.4 Innovation diverges from growth

Proposition 4 (Innovation and Growth)**.**

(a) Services innovation $I(n) = n^{\zeta}(1 - \theta^*(n))r/k$ *is hump-shaped: increasing for small* n *(where* $\theta^* = 0$*) and eventually decreasing for large* n *(where rising centralization overwhelms the scale effect), with* $I(n) = 0$ *for all* $n \geq n_C^*$*. Innovation eventually falls*

below the small-firm reference level.

(b) *Manufacturing innovation* $I(n) = n^{\zeta} r/k$ *is monotonically increasing in* n.

(c) *Operating profits* $\Pi_O(n, \theta^*(n))$ *are increasing in* n *for large firms: growth continues despite innovation collapse.*

(d) *New establishments of multi-unit firms inherit the parent's high* θ*, suppressing initiative and exhibiting flat life-cycle growth relative to de novo startups.*

Results (a) and (c) together establish the central insight: the same θ simultaneously causes innovation collapse and continued growth. This might seem counterintuitive since much theoretical and empirical work equates less innovation with organizational failure. The model offers a fundamentally different reading: innovation collapse is not necessarily a symptom of dysfunction but the *signature* of successful replication. The firm's per-worker surplus rises from π_0 (when workers pursue their own preferred projects) to $\pi_0 + \Delta_A$ (when workers execute the firm's preferred projects), and this surplus funds further expansion through the growth equation $n_{t+1} = n_t + \phi \cdot \Pi_t^{\text{total}}$. The very act that kills initiative is what makes replication profitable. A founder focusing on declining innovation in her scaling venture faces a diagnostic problem: is this healthy or a genuine organizational failure? The model provides a way to tell the difference. If the decline in innovation is accompanied by rising operating profits and continued growth, this can be the result of a successful replication strategy (Winter and Szulanski, 2001). However, if the decline is accompanied by falling profits, the firm has centralized beyond the optimum, destroying innovation without capturing sufficient alignment gains. The distinction matters because the prescriptions are opposite: in the first case, the founder should stay the course; in the second, she should decentralize.

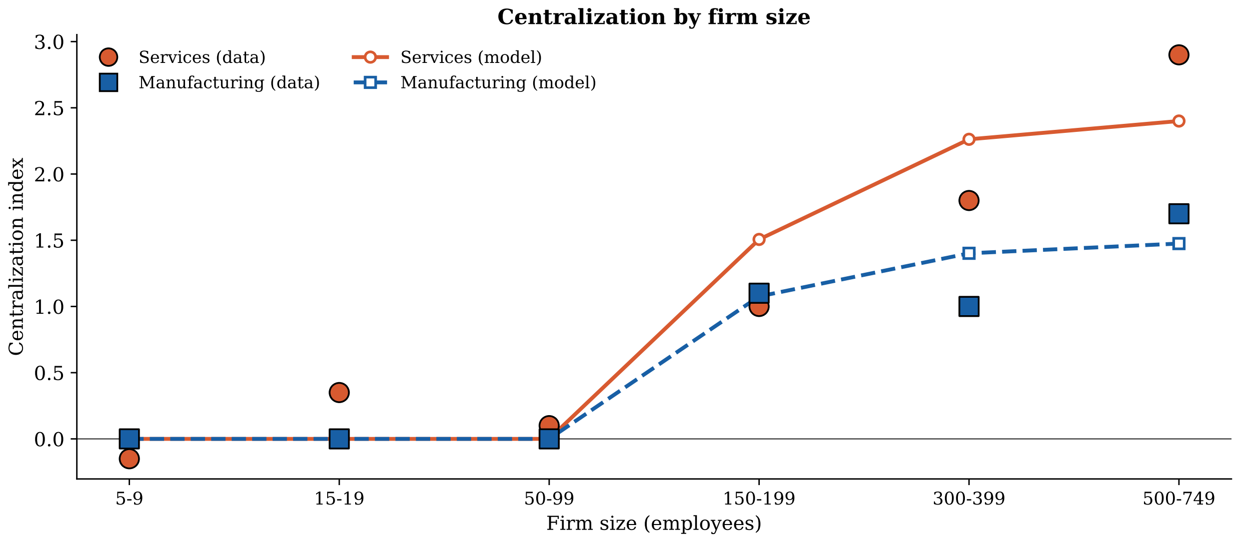
Result (d) extends this logic to multi-unit firms. A new establishment inherits its parent's high θ from day one. Unlike a de novo startup, which begins small and decentralized and thereby passing through the high-initiative phase that generates innovation-driven growth, the establishment starts in the replication regime. It was designed to replicate, not to innovate. Lawrence (2020) provides direct empirical evidence on this mechanism: replicating units in a Fortune 100 retail chain learn from designated templates rather than through independent experimentation. This is exactly the template-following, low-initiative behavior the model predicts for units born in the replication regime. This explains why establishment life-cycle growth profiles are flat compared to startups: it is not that establishments fail to grow, but that they grow through a different mechanism. Startups grow through innovation (low θ, high initiative, new products and processes); establishments grow through replication (high θ, low initiative, standardized operations scaled across markets). The flat growth profile of establishments is not a puzzle, it is exactly what the model predicts for organizational units that were never meant to innovate.

4.5 Extension: Is the Model Consistent with the Data?

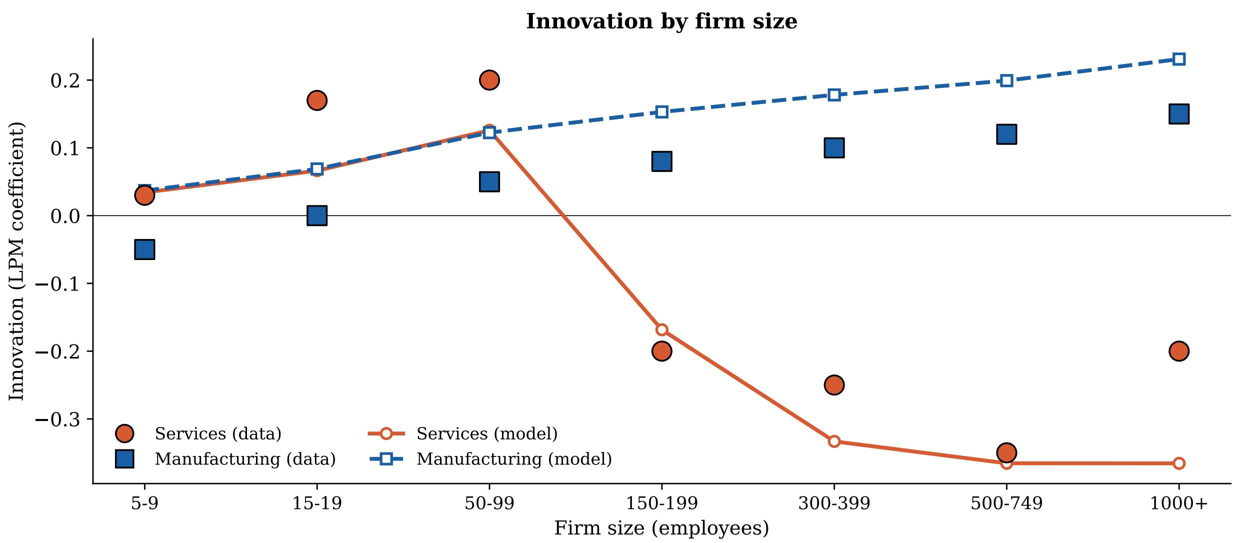
Mapping to the data. The four results correspond directly to the stylized facts documented in Section 2. Proposition 4(d) generates Stylized Fact 1: life-cycle growth is a startup phenomenon because only startups pass through the decentralized, high-initiative phase. Establishments, born centralized, never experience this phase and exhibit flat growth by construction. Proposition 2(b) generates Stylized Fact 2: organizational restructuring tracks size rather than age because $\theta^*(n)$ depends on n and structural parameters, with no role for firm age. A ten-year-old firm with 50 employees and a two-year-old firm with 50 employees face the same coordination problem and adopt the same organizational structure. Proposition 4(a) and 4(b) together generate Stylized Fact 3: the divergence between services and manufacturing innovation profiles. In services ($\sigma = 1$), innovation rises with the scale effect and then collapses as centralization takes hold, producing the hump shape. In manufacturing ($\sigma = 0$), centralization does not affect innovation, so the scale effect operates unimpeded and innovation rises monotonically. Finally, Proposition 4(c) resolves what might otherwise appear as a contradiction in the data: firms in the declining region of the services innovation profile are not shrinking or failing. They are growing because the centralization that killed their innovation is simultaneously generating the operating profits that fund expansion.

Figure 6 shows the empirical fit of the model across cross-section and lifecycle moments; the online appendix provides the full calibration.

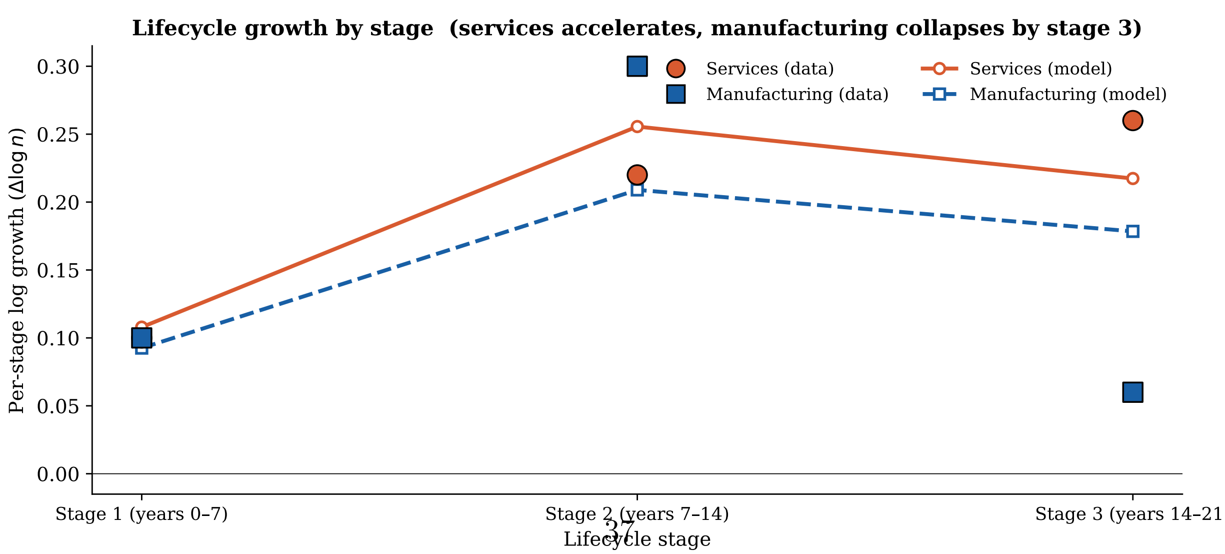
What the lifecycle panel captures, and what it does not. Panel (c) of Figure 6 extends the model's empirical test from the cross-section to the lifecycle. The data show a sharp sectoral divergence: services growth accelerates monotonically from stage 1 to stage 3 (per-stage log growth rising from 0.10 to 0.26), while manufacturing growth peaks at stage 2 (0.30) and collapses to 0.06 by stage 3. The calibration shown in Figure 6 augments the baseline with two reduced-form extensions discussed in detail in Appendix B.15: an initial innovation endowment I_{-1} representing the firm-value output of *nascent entrepreneurship* (the period of founder activity preceding incorporation, during which the venture's product, business model, and initial customer base are developed), and a stage-specific reinvestment rate ϕ_t representing financial-frictions variation across the firm lifecycle. With these extensions, the model reproduces the services lifecycle pattern in both shape and magnitude (model 0.11/0.24/0.23 vs. data 0.10/0.22/0.26) and captures the slow first stage in manufacturing exactly while only partially matching the stage-2 acceleration (model 0.09/0.19 vs. data 0.10/0.30 in stages 1 and 2). The model still does not reproduce the magnitude of the late-stage collapse in manufacturing (model 0.18 in stage 3 vs. data 0.06): once a firm centralizes, per-worker surplus rises from π_0 to $\pi_0 + \Delta_A$, and growth continues rather than decelerating. The manufacturing late-stage decline therefore points to a force outside the model, most plausibly demand-side saturation or capacity constraints that bind for mature manufacturing firms but not for services.



(a) Centralization by firm size.



(b) Innovation by firm size.



(c) Per-stage lifecycle growth (each stage spans roughly seven years).

4.6 Extension: Replication through Establishment Copies

The baseline model captures replication indirectly through centralization: standardizing operations raises profits, which fund growth. But replication in the sense of Winter and Szulanski (2001) involves producing copies of a standardized business model, which the baseline does not model explicitly. The extension introduces M replicated establishments alongside the original unit. Each establishment earns per-unit profit $\pi_M \cdot \theta$ (standardized operations are more profitable to copy), and opening establishments entails convex costs $c_M M^2/2$. The optimal number of establishments is $M^* = \pi_M \theta / c_M$, which rises linearly with centralization. Substituting M^* back into the objective yields a closed-form solution for θ^* with the same structure as the baseline, modified by replication terms $\pi_M^2/(c_M n)$ in both the numerator and denominator. These terms fade with firm size (number of employees), so replication matters most during the scaling transition and the extension converges to the baseline for large firms. At $\pi_M = 0$, the extension collapses exactly to the baseline model. The baseline's strong fit to the WES data likely reflects the fact that the WES measures firm size within the first establishment rather than summing across all establishments of the same firm, the data capture precisely the organizational transformation of the original unit, which is what the baseline describes. Firm-level datasets that aggregate employment across establishments, such as the Longitudinal Business Database (LBD) of the US Census Bureau, may require the replication extension to match the growth patterns of multi-unit firms.

5 Discussion and Conclusion

The value of a formal model lies in the distance between what it assumes and what it delivers. A key model result, that replication and innovation can be either substitutes or complements, follows from three assumptions: (i) coordination costs increase in firm size (δn^α with $\alpha > 2$), (ii) innovation runs into diminishing returns (n^ζ with $\zeta < 1$), and (iii) frontline employees' initiative drives innovation. From these assumptions alone, one might expect the firm to find a stable interior balance between innovation and replication, centralizing enough to fund growth but not so much as to destroy initiative. The formal model reveals something different. The balance point shifts dynamically toward centralization: for larger firms, the tradeoff resolves progressively in favor of replication. This pattern is reminiscent of Nickerson and Zenger (2002)'s theory of the pursuit of ambidexterity via the dynamic complementarity of exploration and exploitation (Boumgarden et al., 2012). In this context, the model reveals that the relationship between replication and innovation depends on a single structural parameter: whether centralizing operations also centralizes innovation (σ). In services ($\sigma = 1$), where the workers who deliver the service are the same workers who innovate, the two growth channels are simultaneous substitutes through organizational structure: the same θ that enables replication-driven growth kills innovation-driven growth. In manufacturing ($\sigma = 0$), where R&D can be organizationally separated from operations, the two channels are sequential complements: centralizing operations raises operating profits, which funds growth, which increases the workforce available for innovation, which generates future profit, which funds further growth. The same model, with the same parameters, generates both relationships. The separability parameter σ therefore serves as a boundary condition for the paper's central hypothesis, that "replication kills initiative." This applies if and only if the organizational structure of innovation and operations cannot be separated. When they can be separated, replication and innovation coexist, and centralization is unambiguously beneficial.

These results also illustrate how formal theory is sometimes helpful to push beyond what empirical analysis alone can deliver. Empirical work can document that organizational structure and innovation are negatively correlated, or that early scaling predicts failure. But without a formal model, empirical analysis alone often cannot identify the conditions under which the observed patterns would reverse. For example, Lee (2022) documents the tradeoff between creative and commercial success empirically and concludes that hierarchy is net beneficial for startups, entailing "a potentially marginal cost of creative success." What his empirical framework cannot do is explain *why* the tradeoff exists, *when* it is worth accepting, or *what determines* whether the creative cost is marginal or severe. My model answers all three questions. The tradeoff exists because centralization overrides the worker initiative that generates innovation: the Aghion and Tirole (1994) mechanism applied to scaling ventures. It is worth accepting when the firm is large enough (high coordination costs), impatient enough (low β), or sufficiently misaligned (high Δ_A) that the operating profit gains outweigh the discounted innovation losses. And the severity of the creative cost depends on whether frontline employees' initiative is essential for innovation (σ): marginal in manufacturing, where R&D can be organizationally separated from operations, but potentially severe in services, where it cannot. Lee's empirical setting, game development, happens to be one where the creative cost is moderate. Without the primitives β, σ, and Δ_A, a founder reading Lee (2022) has no way to assess whether her firm resembles game development (moderate cost) or consulting (severe cost). The model provides this diagnostic.

The model also clarifies what is, and what is not, required to generate the lifecycle dynamics that the scaling literature has sought to explain. A distinctive feature of this framework is that it does not require learning, experimentation, or uncertainty resolution to generate the exploration-to-exploitation transition. Informational theories (Winter and Szulanski, 2001; Gans et al., 2019; Lee and Kim, 2024) model scaling as a commitment decision that should follow sufficient learning about the business model or product-market fit. My model generates venture scaling and firm lifecycle dynamics purely from the organizational economics of innovation, without any role for information or uncertainty. The firm does not centralize because it has learned enough to stop experimenting; it centralizes because coordination costs have grown large enough to make initiative too expensive to maintain. This resolves the circularity in the prescription of informational theories: rather than requiring the firm to determine "when it knows enough" (a determination that itself depends on the organizational structure that scaling will change) the model pins down the entire trajectory of organizational structure, innovation, and growth simultaneously through the firm's optimal response to its own size and future opportunities. The timing of the transition is not a separate decision to be optimized; it is a consequence of the growth path.

Extension to digital platforms. The framework also extends naturally to digital platforms. Platforms such as Uber, Airbnb, Netflix, and YouTube grow by replicating standardized processes across markets. They exhibit a distinctive two-function organizational structure that the model's binary σ does not directly capture. In one function, engineering teams that develop the platform's algorithms and infrastructure are organizationally separated from frontline workers (drivers, hosts, content creators) who deliver the service, so centralizing operations does not kill engineering's initiative, and platform technology can continue innovating as the firm scales. In the other function, however, the platform's standardization of the frontline is exactly the centralization the model predicts kills initiative: a host's idiosyncratic add-ons for stays, a creator's novel content format, real frontline innovations that standardization suppresses or fails to transmit. Frontline standardization is rational because customer predictability is itself valuable: Δ_A is large, and the platform captures value through that predictability. Platforms thus resolve the centralization-initiative tradeoff differently across the two functions. The net effect on observed innovation depends on which function dominates: engineering (Netflix's recommendation algorithm matters more than any individual creator's idiosyncrasy) or frontline (early Airbnb's idiosyncratic stays). Testing this two-function structure sharply would require data on the size and innovation output of engineering teams versus the frontline, and ideally an accounting of frontline innovations that fail to propagate.

The model achieves this parsimony through deliberately stark simplifications: three discrete stages, a single organizational instrument θ, and no heterogeneity across workers or projects. These simplifications are strong, but the calibration exercise suggests they are not in direct conflict with the evidence. A narrow set of parameters simultaneously generates the correct qualitative patterns across four panels: hump-shaped services innovation, monotonic manufacturing innovation, rising centralization in both sectors. At the same time, the model fits the facts that age does not have effects on innovation or centralization beyond firm size and it naturally generates growth patterns over time for startups, while new establishments of existing firms do not significantly grow over time. The model's ability to match these patterns despite its simplicity suggests that the core mechanism, the interaction between innovation-driven and replication-driven growth, captures a first-order force in the data, and that richer models incorporating learning, uncertainty, or continuous time would refine rather than overturn the central predictions.

References

Aghion, P. and Tirole, J. (1994). The management of innovation. *The Quarterly Journal of Economics*, 109(4):1185–1209.

Aghion, P. and Tirole, J. (1997). Formal and real authority in organizations. *Journal of political economy*, 105(1):1–29.

Argyres, N., Rios, L. A., and Silverman, B. S. (2020). Organizational change and the dynamics of innovation: Formal R&D structure and intrafirm inventor networks. *Strategic Management Journal*, 41(11):2015–2049.

Argyres, N., Rios, L. A., and Silverman, B. S. (2025). On the heels of giants: Internal network structure and the race to build on prior innovation. *Strategic Management Journal*, 46(7):1728–1761.

Argyres, N. S. and Silverman, B. S. (2004). R&D, organization structure, and the development of corporate technological knowledge. *Strategic Management Journal*, 25(8–9):929–958.

Azoulay, P., Jones, B. F., Kim, J. D., and Miranda, J. (2020). Age and high-growth entrepreneurship. *American Economic Review: Insights*, 2(1):65–82.

Bandiera, O., Barankay, I., and Rasul, I. (2005). Social preferences and the response to incentives: Evidence from personnel data. *Quarterly Journal of Economics*, 120(3):917–962.

Boumgarden, P., Nickerson, J., and Zenger, T. R. (2012). Sailing into the wind: Exploring the relationships among ambidexterity, vacillation, and organizational performance. *Strategic management journal*, 33(6):587–610.

Combs, J. G., Ketchen, D. J., and Short, J. C. (2011). Franchising research: Major milestones, new directions, and its future within entrepreneurship. *Entrepreneurship Theory and Practice*, 35(3):413–425.

DeSantola, A. and Gulati, R. (2017). Scaling: Organizing and growth in entrepreneurial ventures. *Academy of Management Annals*, 11(2):640–668.

Gans, J. S., Stern, S., and Wu, J. (2019). Foundations of entrepreneurial strategy. *Strategic Management Journal*, 40(5):736–756.

Genedy, M., Hellerstedt, K., Naldi, L., and Wiklund, J. (2024). Growing pains in scale-ups: How scaling affects new venture employee burnout and job satisfaction. *Journal of Business Venturing*, 39(2):106367.

Grossman, S. J. and Hart, O. D. (1986). The costs and benefits of ownership: A theory of vertical and lateral integration. *Journal of Political Economy*, 94(4):691–719.

Hatch, N. W. and Mowery, D. C. (1998). Process innovation and learning by doing in semiconductor manufacturing. *Management Science*, 44(11, Part 1):1461–1477.

Ichniowski, C., Shaw, K., and Prennushi, G. (1997). The effects of human resource management practices on productivity: A study of steel finishing lines. *American Economic Review*, 87(3):291–313.

Jones, C. I. and Williams, J. C. (1998). Measuring the social return to r&d. *The Quarterly Journal of Economics*, 113(4):1119–1135.

Jovanovic, B. (1982). Selection and the evolution of industry. *Econometrica: Journal of the econometric society*, pages 649–670.

Ketchen, D. J., Short, J. C., and Combs, J. G. (2011). Is franchising entrepreneurship? Yes, no, and maybe so. *Entrepreneurship Theory and Practice*, 35(3):583–593.

Kueng, L., Yang, M.-J., and Hong, B. (2017). Firm life-cycle dynamics. NBER Working Paper.

Lawrence, M. (2020). Replication using templates: Does the unit learn from itself, the template, or both? *Strategic Management Journal*, 41(11):1955–1982.

Lazear, E. P. (2000). Performance pay and productivity. *American Economic Review*, 90(5):1346–1361.

Lee, S. (2022). The myth of the flat start-up: Reconsidering the organizational structure of start-ups. *Strategic Management Journal*, 43(1):58–92.

Lee, S. and Kim, J. D. (2024). When do startups scale? large-scale evidence from job postings. *Strategic Management Journal*, 45(9):1633–1669.

Leiblein, M. J., Reuer, J. J., and Zenger, T. (2018). What makes a decision strategic? *Strategy Science*, 3(4):558–573.

Lerner, J. and Leamon, A. (2023). *Venture capital, private equity, and the financing of entrepreneurship*. John Wiley & Sons.

Levinthal, D. A. and Wu, B. (2010). Opportunity costs and non-scale free capabilities: Profit maximization, corporate scope, and profit margins. *Strategic Management Journal*, 31(7):780–801.

Lucas, R. E. (1978). On the size distribution of business firms. *The Bell Journal of Economics*, 9(2):508–523.

March, J. G. (1991). Exploration and exploitation in organizational learning. *Organization Science*, 2(1):71–87.

Nickerson, J. A. and Zenger, T. R. (2002). Being efficiently fickle: A dynamic theory of organizational choice. *Organization Science*, 13(5):547–566.

Rivkin, J. W. (2001). Reproducing knowledge: Replication without imitation at moderate complexity. *Organization Science*, 12(3):274–293.

Sampson, R. C. and Shi, Y. (2023). Are us firms becoming more short-term oriented? evidence of shifting firm time horizons from implied discount rates, 1980–2013. *Strategic Management Journal*, 44(1):231–263.

Schumpeter, J. A. (1942). *Capitalism, Socialism and Democracy*. Harper & Brothers, New York.

Storn, R. and Price, K. (1997). Differential evolution – a simple and efficient heuristic for global optimization over continuous spaces. *Journal of Global Optimization*, 11(4):341–359.

Vargo, S. L. and Lusch, R. F. (2004). Evolving to a new dominant logic for marketing. *Journal of Marketing*, 68(1):1–17.

von Nordenflycht, A. (2010). What is a professional service firm? toward a theory and taxonomy of knowledge-intensive firms. *Academy of Management Review*, 35(1):155–174.

Winter, S. G. and Szulanski, G. (2001). Replication as strategy. *Organization Science*, 12(6):730–743.

ONLINE APPENDIX

A Model Setup and Proofs

Roadmap. This appendix builds up the technical apparatus needed to prove the four propositions in Section 2. Because the proofs share machinery, the appendix is organized by technical build-up rather than by proposition number: the order in which it proves propositions does not match the order in which the main text states them, and the table below maps each proposition to where its proof lives. The primitive setup and the backward-induction solution come first; the project-selection microfoundation for Δ_A and b, which underlies the primitives but is not needed for any proof, is deferred to the end (§A.8).

Organization.

- *Setup (§A.1–§A.2).* The centralization choice problem and its primitives (§A.1); the backward-induction solution that pins the saturated regime and the growth multiplier Λ (§A.2).

- *Technical apparatus (§A.3–§A.4).* The implicit function theorem framework (§A.3) yields the regularity condition $n > n^{**}$ and the comparative statics of θ^* in the interior regime. The large-n closed form (§A.4) specializes this to the saturated regime, in which Λ collapses to a parameter and θ^* admits a clamped algebraic expression.

- *Threshold and boundary results (§A.5–§A.6).* The full decentralization threshold n^*_D is defined in primitives and its comparative statics derived (§A.5); the boundary conditions (B1)–(B3) of Proposition 1 are then verified (§A.6).

- *Innovation–growth proofs (§A.7).* Proposition 4 is proved using the already-established properties of θ^*.

- *Microfoundation (§A.8).* The project-selection model underlying Δ_A, b, and v_I, supplying the restriction $\Delta_A \leq b$ used in the calibration.

Where each proposition's proof lives.

Proposition	Statement in main text	Proof location
1 (Lifecycle Trajectory)	§4.1	§A.6, building on §A.5
2 (Centralization and Firm Size)	§4.2	part (a), (b): §A.3; closed form: §A.4
3 (Drivers of Centralization)	§4.3	interior: §A.3; thresholds: §A.5
4 (Innovation and Growth)	§4.4	§A.7

A.1 Primitives and timing

The firm operates over three stages $t \in \{1, 2, 3\}$ with discount factor $\beta \in (0, 1]$, with each stage corresponding to 5-10 years as discussed in the main text. In each stage, the firm observes its size n_t, chooses centralization $\theta_t \in [0, 1]$, earns operating profits, and generates innovations. Innovations pay off both in the profits of the current stage and the growth rate to the following stage.

Operating profits.

$$\Pi_O(n, \theta) = n[\pi_0 + \Delta_A\theta] - \delta n^{\alpha}(1 - \theta)^2 - F\theta, \tag{16}$$

where π_0 is the baseline per-worker surplus, Δ_A is the alignment gain from centralization, $\delta n^{\alpha}(1 - \theta)^2$ is the coordination cost of decentralization, and F is the overhead cost of centralized infrastructure.

Innovation.

$$I(n, \theta, \sigma) = \frac{n^{\zeta}(1 - \sigma\theta)(v_I + b)}{k}, \tag{17}$$

which gives $I = n^{\zeta}(1-\theta)(v_I+b)/k$ for services ($\sigma = 1$) and $I = n^{\zeta}(v_I+b)/k$ for manufacturing ($\sigma = 0$). Worker initiative is $e^* = (1 - \sigma\theta)(v_I + b)/k$, where v_I is the firm's per-innovation profit, b is the worker's private benefit, and k is the cost-of-effort parameter.

Growth. Firm size evolves by reinvesting total profits:

$$n_{t+1} = n_t + \phi \cdot \big[\Pi_O(n_t, \theta_t) + v_I \cdot I_{t-1}\big], \tag{18}$$

where $\phi > 0$ is the reinvestment rate and $v_I \cdot I_{t-1}$ is the innovation profit accruing in stage t from innovations generated in stage $t-1$. The initial endowment I_0 supplies stage-1 innovation profit.

Present value. The firm maximizes by backward induction:

$$J_t = \Pi_O(n_t, \theta_t) + v_I \cdot I_{t-1} + \beta \cdot J_{t+1}. \tag{19}$$

Since $v_I \cdot I_{t-1}$ is predetermined, the firm's choice of θ_t affects J_t through Π_O directly and through the innovations I_t generated this period (which enter J_{t+1} as $v_I \cdot I_t$).

Parameter restrictions. $k \geq v_I + b$ (effort cost exceeds combined returns, ensuring $e^* \leq 1$), $\alpha > 2$ (coordination costs are superlinear in n, required for the hump shape), $\zeta \in (0, 1)$ (diminishing marginal innovation contribution per worker), and $\delta, F, \Delta_A, \phi > 0$.

A.2 Solving by backward induction

Stage 3 (terminal). The Stage-3 problem is simpler than Stages 1 and 2 because there is no future to discount: no Stage 4 exists, so the innovations I_3 that the firm could generate at Stage 3 are worthless, their profit would only accrue at the (non-existent) Stage 4 and is therefore lost. The Stage-3 cash flow is $\Pi_O(n_3, \theta_3) + v_I \cdot I_2$, where $v_I \cdot I_2$ is the profit from innovations generated last period (predetermined from the Stage-3 perspective) and $\Pi_O(n_3, \theta_3)$ is the operating profit the firm chooses θ_3 to optimize. Since $v_I \cdot I_2$ does not depend on θ_3, the first-order condition reduces to $\partial\Pi_O/\partial\theta_3 = 0$:

$$n_3\Delta_A - F + 2\delta n_3^\alpha(1-\theta_3) = 0. \tag{20}$$

When $n_3\Delta_A \geq F$, the alignment gain exceeds the per-firm overhead and the solution saturates: $\theta_3^* = 1$, with value $J_3^* = n_3(\pi_0 + \Delta_A) - F + v_I \cdot I_2$. When $n_3\Delta_A < F$ (small firms where overhead dominates), the solution is $\theta_3^* = \left[1 - (F - n_3\Delta_A)/(2\delta n_3^\alpha)\right]_+$: interior when $F - n_3\Delta_A < 2\delta n_3^\alpha$, and at the fully decentralized corner $\theta_3^* = 0$ for very small firms, where $F - n_3\Delta_A \geq 2\delta n_3^\alpha$ (equivalently $n_3 \leq n_{D,M}^*$ in the notation of Section A.5). The calibrated parameters place all empirically relevant firm sizes in the saturated regime; the interior case is included for completeness.

Interior stages (stages 1 and 2): first-order condition. Both interior stages share the same structure. The firm maximizes

$$J_t = \Pi_O(n_t, \theta_t) + v_I \cdot I_{t-1} + \beta \cdot J_{t+1}^*(n_{t+1}),$$

where for $t = 2$, I_1 is the innovation generated at Stage 1 by θ_1, and for $t = 1$, I_0 is the initial innovation endowment. Two timing relationships matter at each interior stage, and it is easy to conflate them:

- I_{t-1} is predetermined from the perspective of the Stage-t decision, it is part of the cash on hand, and flows into the next-period state through the growth equation:

$$n_{t+1} = n_t + \phi[\Pi_O(n_t, \theta_t) + v_I \cdot I_{t-1}].$$

  So I_{t-1}, not I_t, is what funds the transition from n_t to n_{t+1}, because the innovations generated at Stage t do not pay out as profit until Stage t+1.

- $I_t = n_t^\zeta(1 - \sigma\theta_t)(v_I + b)/k$ is the innovation generated at Stage t by the current choice

θ_t. It does not enter n_{t+1}; it enters J^*_{t+1} as the Stage-(t+1) profit $v_I \cdot I_t$, which is predetermined from the perspective of the Stage-(t+1) problem (since θ_t has already been chosen by then), but is fully responsive to θ_t from the perspective of the Stage-t problem we are solving here. One further step in the chain matters for $t = 1$: once the Stage-2 profit $v_I \cdot I_1$ arrives, it is itself reinvested through the growth equation and funds the transition $n_2 \to n_3$. A marginal unit of Stage-1 innovation cash is therefore worth *more* than one unit of Stage-2 profit, it is worth the Stage-2 growth multiplier defined below. For $t = 2$ this step is absent, because Stage-3 profits are terminal and nothing is reinvested after them.

To keep track of this last effect, define the marginal value of innovation cash arriving at Stage t+1,

$$\mu_{t+1} \equiv \frac{\partial J^*_{t+1}}{\partial(v_I I_t)}, \qquad \mu_3 = 1, \qquad \mu_2 = \Lambda_2(n_3),$$

where the terminal value $\mu_3 = 1$ holds because Stage-3 cash is consumed, and $\mu_2 = \Lambda_2$ holds because Stage-2 cash on hand enters the Stage-2 objective one-for-one *and* is reinvested into n_3 at rate ϕ (the envelope calculation establishing this is the same as for operating profit, given below). Differentiating J_t with respect to θ_t therefore picks up three channels: θ_t's direct effect on Π_O, its effect on n_{t+1} via Π_O in the growth equation, and its effect on J^*_{t+1} via the innovation I_t, the last channel weighted by the marginal value μ_{t+1} of the innovation cash when it arrives:

$$\frac{\partial J_t}{\partial \theta_t} = \frac{\partial \Pi_O}{\partial \theta_t} + \beta \frac{dJ^*_{t+1}}{dn_{t+1}} \cdot \frac{dn_{t+1}}{d\theta_t} + \beta\, \mu_{t+1}\, v_I \frac{\partial I_t}{\partial \theta_t} = 0. \tag{21}$$

Because both n_t and $v_I I_{t-1}$ are predetermined with respect to θ_t, we have $dn_{t+1}/d\theta_t = \phi \cdot \partial \Pi_O / \partial \theta_t$. Define the growth multiplier:

$$\Lambda_t(n_{t+1}) \equiv 1 + \beta\phi \cdot \frac{dJ^*_{t+1}}{dn_{t+1}}. \tag{22}$$

The growth multiplier captures the fact that a unit increase in Π_O at Stage t raises J_t by more than one: directly through the Π_O term, and indirectly through the $\phi \cdot \Pi_O$ contribution to future size n_{t+1}, which then raises J^*_{t+1} at rate dJ^*_{t+1}/dn_{t+1} and is discounted back by one stage, using the discount factor β. The envelope theorem applied to the next stage's problem gives an explicit expression for this rate.[12]

The first-order condition (21) can therefore be written compactly as:

$$\Lambda_t(n_{t+1}) \cdot \frac{\partial \Pi_O}{\partial \theta_t} + \beta\, \mu_{t+1}\, v_I \frac{\partial I_t}{\partial \theta_t} = 0, \qquad \mu_3 = 1, \quad \mu_2 = \Lambda_2(n_3). \tag{23}$$

The replication channel picks up the multiplier Λ_t because operating profit is reinvested into n_{t+1}; the innovation channel does not affect n_{t+1}, so it shows up only through the explicit $\beta\mu_{t+1}v_I \partial I_t/\partial\theta_t$ term. At the penultimate stage ($t = 2$) the weight on innovation is simply β, since $\mu_3 = 1$. At Stage 1 the weight is $\beta\Lambda_2$: Stage-1 innovations pay out at Stage 2, where the cash is reinvested once more, so a unit of Stage-1 innovation cash is amplified by the Stage-2 growth multiplier before being discounted back. Since $\Lambda_t(n_{t+1})$ depends on θ_t through $n_{t+1}(\theta_t)$, equation (23) is an implicit equation that does not yield a closed-form solution for θ^*_t in general. Comparative statics can still be obtained via the implicit function

[12] For the boundary case $t = 2$, where the next stage is the terminal Stage 3, the Stage-3 value function is $J^*_3(n_3) = \Pi_O(n_3, \theta^*_3(n_3)) + v_I \cdot I_2$, where I_2 does not depend on n_3 (it is determined by Stage-2 choices). The total derivative with respect to n_3 has two pieces: a direct effect from Π_O holding θ^*_3 fixed, and an indirect effect through $\theta^*_3(n_3)$:

$$\frac{dJ^*_3}{dn_3} = \frac{\partial \Pi_O(n_3, \theta^*_3)}{\partial n_3} + \underbrace{\frac{\partial \Pi_O(n_3, \theta^*_3)}{\partial \theta_3} \cdot \frac{d\theta^*_3}{dn_3}}_{=0 \text{ (see text)}}.$$

The second term vanishes in all three Stage-3 regimes: at an interior optimum because θ^*_3 satisfies the Stage-3 first-order condition $\partial\Pi_O/\partial\theta_3 = 0$; at the saturated corner $\theta^*_3 = 1$ because θ^*_3 is then locally constant in n_3, so $d\theta^*_3/dn_3 = 0$ (the corner derivative $\partial\Pi_O/\partial\theta_3$ is strictly positive there, but it is multiplied by zero); and likewise at the fully decentralized corner $\theta^*_3 = 0$, where θ^*_3 is again locally constant in n_3 and the (there negative) corner derivative is multiplied by zero. (The full Stage-3 objective is $\Pi_O + v_I I_2$, but $v_I I_2$ is constant in θ_3, so the first-order condition for Π_O alone is what matters.) Evaluating the surviving direct effect using $\Pi_O(n_3, \theta_3) = n_3(\pi_0 + \Delta_A\theta_3) - \delta n_3^\alpha(1 - \theta_3)^2 - F\theta_3$ gives:

$$\frac{dJ^*_3}{dn_3} = \pi_0 + \Delta_A\theta^*_3 - \alpha\delta n_3^{\alpha-1}(1 - \theta^*_3)^2.$$

For $t = 1$, the analogous envelope step applies to the Stage-2 value function J^*_2, with dJ^*_2/dn_2 playing the role of dJ^*_3/dn_3.

theorem (Section A.3); an explicit closed form is available for large n once the multipliers stabilize at parameter values (Section A.4).

Closed form for large firm size. Generally at the interior Stages 1 and 2, there are no closed-form solutions for θ_t^*; the optimal centralization is characterized only by the first-order condition (23). However, for firms that anticipate being very large in the terminal stage (Stage 3), closed-form solutions are available. The intuition behind this result runs in two steps. First, firms that are very large at Stage 3 are certain to fully centralize then ($\theta_3^* = 1$), because the per-worker alignment gain Δ_A scales with size while the overhead cost F does not, so for n_3 large enough the efficiency gains from centralization dominate any remaining decentralization motive, especially since Stage-3 innovations are worthless and there is no future to protect. Second, a firm that is certain to fully centralize at Stage 3 is also a firm for which the marginal return to size at Stage 3 is a constant: when $\theta_3^* = 1$, coordination costs vanish at the margin, and every additional worker contributes the same per-worker surplus $\pi_0 + \Delta_A$ regardless of n_3. A constant marginal return to size at Stage 3 in turn means the Stage-2 growth multiplier Λ_2 collapses to a parameter rather than varying with n_3:[13]

$$\Lambda = 1 + \beta\phi(\pi_0 + \Delta_A).$$

At the penultimate stage this is the only multiplier that appears ($\mu_3 = 1$), so the Stage-2 first-order condition (23) is linear in θ_2 and admits the closed-form solution stated in Proposition 2. One stage further upstream the multipliers do *not* all collapse to Λ. If the firm also saturates at Stage 2 ($\theta_2^* = 1$; because the Stage-2 first-order condition retains

[13]Two pieces combine to give this parameter. First, the Stage-3 first-order condition (20) is $n_3\Delta_A - F + 2\delta n_3^\alpha(1-\theta_3) = 0$. When $n_3\Delta_A \geq F$, the residual coordination-cost term $2\delta n_3^\alpha(1-\theta_3)$ would need to be negative to satisfy this condition, which is impossible since both $\delta > 0$ and $(1-\theta_3) \geq 0$; the boundary $\theta_3^* = 1$ is therefore the solution. Second, plugging $\theta_3^* = 1$ into the envelope expression

$$\frac{dJ_3^*}{dn_3} = \pi_0 + \Delta_A\theta_3^* - \alpha\delta n_3^{\alpha-1}(1-\theta_3^*)^2$$

kills the coordination-cost term (since $(1-1)^2 = 0$) and leaves $\pi_0 + \Delta_A$, a constant in n_3. Substituting into $\Lambda_2(n_3) = 1 + \beta\phi \cdot dJ_3^*/dn_3$ then yields the displayed parameter, which depends only on primitives.

the innovation term, in services this requires $n_2 \geq n^*_{C,S}$, the full centralization threshold of Section A.4, which is strictly stronger than the terminal-stage condition $n_2\Delta_A \geq F$, the latter delivers $\theta^* = 1$ only at Stage 3, where innovations are worthless, or in manufacturing), the envelope derivative of the Stage-2 value function is

$$\frac{dJ_2^*}{dn_2} = \underbrace{\Lambda(\pi_0 + \Delta_A)}_{\text{Stage-2 profit, reinvested}} + \underbrace{\beta(\pi_0 + \Delta_A)}_{\text{direct effect on } n_3},$$

(derived from the explicit expansion of J_2^* below), so the Stage-1 growth multiplier is the strictly larger constant

$$\Lambda_1 \;=\; 1 + \beta\phi(\pi_0 + \Delta_A)(\Lambda + \beta) \;>\; \Lambda, \tag{24}$$

reflecting the fact that a marginal unit of Stage-1 operating profit compounds through *two* rounds of reinvestment rather than one. The Stage-1 first-order condition is then $\Lambda_1 \partial\Pi_O/\partial\theta_1 + \beta\Lambda v_I \partial I_1/\partial\theta_1 = 0$ (using $\mu_2 = \Lambda$), which is again linear in θ_1 and yields the same functional form as the Stage-2 closed form with the innovation weight β/Λ replaced by $\beta\Lambda/\Lambda_1$. At the calibrated parameters $\Lambda = 1.0503$ and $\Lambda_1 = 1.0780$, so $\Lambda/\Lambda_1 = 0.974$ versus $1/\Lambda = 0.952$, a 2.3% reweighting of the innovation force that shifts interior θ_1^* by less than 0.008 across the empirical size bins (Section A.4 quantifies this against the exact numerical backward induction). The calibration in Appendix B interprets the cross-sectional policy as the penultimate-stage problem, for which the closed form with Λ is exact; I show there that this saturated regime is not just a theoretically convenient but also empirically realistic case.

To see the closed form arise directly, consider Stage 2 as the base case. Expand J_2 using the Stage-3 value function $J_3^* = n_3(\pi_0 + \Delta_A) - F + v_I I_2$ and the growth equation

$n_3 = n_2 + \phi(\Pi_{O,2} + v_I I_1)$:

$$J_2 = \Pi_{O,2} + v_I I_1 + \beta\big[(n_2 + \phi(\Pi_{O,2} + v_I I_1))(\pi_0 + \Delta_A) - F + v_I I_2\big]$$
$$= \Lambda \cdot \Pi_{O,2} + \Lambda \cdot v_I I_1 + \beta v_I I_2 + H(n_2),$$

where $H(n_2) = \beta(\pi_0 + \Delta_A)n_2 - \beta F$ collects terms constant in θ_2. The replication and predetermined terms both pick up Λ because they enter cash flow at Stage 2 and are reinvested into n_3. The innovation term $\beta v_I I_2$ does *not* pick up Λ, because I_2 generates Stage-3 profit directly rather than by funding the transition $n_2 \to n_3$, and Stage-3 profit is terminal. Since I_1 is predetermined, only $\Lambda \cdot \Pi_{O,2}$ and $\beta v_I I_2$ depend on θ_2, and the first-order condition is:

$$\Lambda \frac{\partial \Pi_O}{\partial \theta_2} - \frac{\sigma \beta v_I (v_I + b) n_2^{\zeta}}{k} = 0,$$

where the second term used $\partial I_2 / \partial \theta_2 = -\sigma n_2^{\zeta}(v_I + b)/k$, which equals $-n_2^{\zeta}(v_I + b)/k$ for services ($\sigma = 1$) and zero for manufacturing ($\sigma = 0$).

The Stage-1 analogue follows from the same expansion taken one step further. The Stage-2 expansion above, evaluated at the saturated optimum $\theta_2^* = 1$, gives the Stage-2 value function as a function of its two state variables:

$$J_2^*(n_2, I_1) = \Lambda \cdot \Pi_O(n_2, 1) + \Lambda \cdot v_I I_1 + \beta v_I I_2(n_2, 1) + \beta(\pi_0 + \Delta_A)n_2 - \beta F,$$

where $\Pi_O(n_2, 1) = n_2(\pi_0 + \Delta_A) - F$ is linear in n_2 and $I_2(n_2, 1) = n_2^{\zeta}(1 - \sigma)(v_I + b)/k$ (zero for services). Reading off the coefficients: $\partial J_2^* / \partial(v_I I_1) = \Lambda$, confirming $\mu_2 = \Lambda$, and for services $dJ_2^*/dn_2 = \Lambda(\pi_0 + \Delta_A) + \beta(\pi_0 + \Delta_A) = (\pi_0 + \Delta_A)(\Lambda + \beta)$, confirming (24). Substituting into the Stage-1 objective $J_1 = \Pi_{O,1} + v_I I_0 + \beta J_2^*(n_2, I_1)$ with $n_2 = n_1 + \phi(\Pi_{O,1} + v_I I_0)$ and collecting the θ_1-dependent terms yields

$$J_1 = \Lambda_1 \cdot \Pi_{O,1} + \beta\Lambda \cdot v_I I_1 + (\text{terms constant in } \theta_1),$$

and hence the Stage-1 first-order condition

$$\Lambda_1 \frac{\partial \Pi_O}{\partial \theta_1} - \frac{\sigma\beta\Lambda\, v_I(v_I + b)n_1^{\zeta}}{k} = 0.$$

This has the same linear structure as the Stage-2 condition, with the innovation weight $\beta\Lambda/\Lambda_1$ in place of β/Λ; the resulting closed form and its quantitative proximity to the Stage-2 form are developed in Section A.4. For manufacturing ($\sigma = 0$) the innovation term vanishes at both stages, the multipliers Λ_1, Λ cancel from the respective first-order conditions, and the Stage-1 and Stage-2 policies coincide exactly.

A.3 Comparative statics via the implicit function theorem

This subsection establishes the regularity condition (used in Proposition 2) and the comparative statics of θ^* in the interior regime (Proposition 2(a)–(b) and Proposition 3(a)–(c) on interior θ^*).

The interior-stage first-order condition (23) implicitly defines θ_t^* as a function of the state n_t and the parameters. Define

$$F(\theta_t; n_t, \beta, \Delta_A, \sigma) \equiv \Lambda_t(n_{t+1}(\theta_t)) \cdot \frac{\partial \Pi_O}{\partial \theta_t} + \beta v_I \frac{\partial I_t}{\partial \theta_t},$$

so that the first-order condition is $F(\theta_t^*; \cdot) = 0$. By the implicit function theorem, $d\theta_t^*/dx = -(\partial F/\partial x)/(\partial F/\partial \theta_t)$ for any parameter x, and the sign of $d\theta_t^*/dx$ equals the sign of $\partial F/\partial x$ provided $\partial F/\partial \theta_t < 0$.

Scope. As written, F is the exact first-order condition for the penultimate stage, where $\mu_3 = 1$ (Section A.2); the derivations below are stated for that stage. The Stage-1 condition multiplies the innovation term by $\mu_2 = \Lambda_2 > 0$ (a constant under the saturated chain; in general $\Lambda_2(n_3)$ is endogenous, but at calibrated parameters it varies only within [1.0492, 1.0503]) and replaces Λ_t by the Stage-1 multiplier. Because μ_2 enters as a positive multiplicative factor on the innovation term, every comparative-static sign derived below is unchanged at Stage 1; only magnitudes shift, and by little (the innovation weight changes by 2.3% at calibrated parameters, Section A.4). The second-order analysis is also conservative at Stage 1 in the empirically relevant case: under the saturated chain ($\theta_2^* = 1$), the Stage-2 value function J_2^* is linear in n_2 for services and concave in n_2 for manufacturing (its only nonlinear term is $\beta v_I I_2(n_2, 1) \propto n_2^{\zeta}$ with $\zeta < 1$), so the destabilizing reinvestment-feedback term analyzed in Steps 1–7 below is zero or negative at Stage 1, and the Stage-1 second-order condition holds whenever the static coordination convexity is positive.

The trade-off underlying the second-order condition. The model contains three forces shaping the centralization choice. *Alignment gain* $n\Delta_A\theta$ rewards centralization because aligned worker activity is more productive per worker. *Overhead cost* $-F\theta$ penalizes centralization because the centralized infrastructure costs a fixed sum. *Coordination friction* $-\delta n^{\alpha}(1-\theta)^2$ rewards centralization because the friction of un-aligned activity at scale is steep and minimized at $\theta = 1$. Set against these is the *initiative drag*: centralization kills worker initiative and hence innovation, with the lost innovation profit discounted back to today and growing with firm size (since more employees' initiative is curtailed in a larger firm).

At the margin, the first-order condition (23) balances the growth-amplified centralization payoff (alignment + coordination relief + overhead) against the initiative drag. But the first-order condition tells us only where the balance sits; the second-order condition asks whether the trade-off is well-behaved enough that this balance is a unique interior maximum. For that, what matters is not the magnitude of each force but the *curvature* each contributes to the operating profit as a function of θ. Alignment ($n\Delta_A\theta$) and overhead ($-F\theta$) are linear in θ and contribute zero curvature, they shift the slope but not the bend. *Only the coordination friction, being quadratic in θ, contributes any static curvature to the operating-profit landscape.* The second-order condition therefore stands or falls on whether this single static restoring force, coordination cost convexity in θ, is strong enough to dominate the dynamic feedback through which the initiative drag perturbs the choice via reinvestment.

The remainder of this section makes this contest precise in five steps.

Step 1: Exact second-order condition. Differentiating F with respect to θ_t (using the product rule on $\Lambda_t(n_{t+1}(\theta_t)) \cdot \partial\Pi_O/\partial\theta_t$, since both factors depend on θ_t) yields three terms:

$$\frac{\partial F}{\partial \theta_t} = \underbrace{\Lambda_t \cdot \frac{\partial^2 \Pi_O}{\partial \theta_t^2}}_{\text{(i) coordination convexity}} + \underbrace{\frac{d\Lambda_t}{dn_{t+1}} \cdot \frac{dn_{t+1}}{d\theta_t} \cdot \frac{\partial \Pi_O}{\partial \theta_t}}_{\text{(ii) reinvestment feedback}} + \underbrace{\beta v_I \frac{\partial^2 I_t}{\partial \theta_t^2}}_{\text{(iii) innovation curvature}} .$$

Term (i) is $-\Lambda_t \cdot \kappa_\theta(n_t)$, where $\kappa_\theta(n) \equiv 2\delta n^\alpha$ is the **coordination-cost convexity in** θ: the second derivative of the coordination cost $\delta n^\alpha(1-\theta)^2$ with respect to θ. This is the only static curvature in the model. It grows in n as n^α, larger firms have steeper coordination cost penalties, which sharpens the operating-profit peak in θ.

Term (iii) is zero because innovation $I_t = n_t^\zeta(1-\sigma\theta_t)(v_I + b)/k$ is linear in θ_t. Term (ii) is the **reinvestment feedback**: θ_t raises current profits, which raise n_{t+1} through reinvestment, which then affects the future value function J^*_{t+1} and hence Λ_t. Two factors of ϕ enter through composition: one because $dn_{t+1}/d\theta_t = \phi \cdot \partial\Pi_O/\partial\theta_t$ (the growth equation transmits at rate ϕ), and one because $\Lambda_t = 1 + \beta\phi \cdot dJ^*_{t+1}/dn_{t+1}$ contains an explicit ϕ that survives differentiation as $d\Lambda_t/dn_{t+1} = \beta\phi \cdot d^2 J^*_{t+1}/dn^2_{t+1}$. Combining:

$$\frac{\partial F}{\partial \theta_t} = -\Lambda_t \cdot \kappa_\theta(n_t) + \beta\phi^2 \cdot \frac{d^2 J^*_{t+1}}{dn^2_{t+1}} \cdot \left(\frac{\partial \Pi_O}{\partial \theta_t}\right)^2 . \tag{25}$$

The second-order condition $\partial F/\partial\theta_t < 0$ therefore reduces to a contest:

$$\Lambda_t \cdot \kappa_\theta(n_t) \;>\; \beta\phi^2 \cdot \frac{d^2 J^*_{t+1}}{dn^2_{t+1}} \cdot \left(\frac{\partial \Pi_O}{\partial \theta_t}\right)^2 \tag{26}$$

between the growth-amplified coordination-cost convexity in θ (left-hand side) and the reinvestment feedback from the future value function (right-hand side). The left-hand side is unambiguously positive. The sign of the right-hand side depends on the curvature of J^*_{t+1} in n_{t+1}.

Saturated regime ($\theta^*_{t+1} = 1$, i.e., $n_{t+1}\Delta_A \geq F$). If the next-stage problem saturates, $J^*_{t+1} = n_{t+1}(\pi_0 + \Delta_A) - F + v_I I_t$ is linear in n_{t+1}, so $d^2 J^*_{t+1}/dn^2_{t+1} = 0$. The reinvestment feedback vanishes and the second-order condition simplifies dramatically: $\partial \mathcal{F}/\partial\theta_t = -\Lambda_t \kappa_\theta(n_t) < 0$ directly. **When the saturated case applies, the entire interior-regime apparatus below is unnecessary; the implicit function theorem applies unconditionally.** The saturation chain established in Section A.2 and verified numerically in Appendix B places all empirically relevant firm sizes in this regime, so this is the typical case in practice.

Interior regime ($\theta^*_{t+1} < 1$, i.e., $n_{t+1}\Delta_A < F$). If the next stage is in the interior regime, substituting the interior solution $(1 - \theta^*_{t+1}) = (F - n_{t+1}\Delta_A)/(2\delta n^\alpha_{t+1})$ into the Stage-(t+1) value function yields (after dropping terms linear in n_{t+1}, which contribute zero to the second derivative):

$$J^*_{t+1}(n_{t+1}) = \text{linear in } n_{t+1} + \frac{(F - n_{t+1}\Delta_A)^2}{4\delta n^\alpha_{t+1}} + \beta J^*_{t+2}.$$

Computing the second derivative:

$$\frac{d^2}{dn^2_{t+1}}\left[\frac{(F - n_{t+1}\Delta_A)^2}{4\delta n^\alpha_{t+1}}\right] = \frac{1}{4\delta n^{\alpha+2}_{t+1}}\left[2\Delta_A^2 n^2_{t+1} + 4\alpha\Delta_A(F - n_{t+1}\Delta_A)n_{t+1} + \alpha(\alpha+1)(F - n_{t+1}\Delta_A)^2\right], \tag{27}$$

which is positive since each bracketed term is positive in the interior regime (where $F - n_{t+1}\Delta_A > 0$). The recursive term βJ^*_{t+2} contributes additional non-negative curvature, but $d^2 J^*_{t+1}/dn^2_{t+1} > 0$ already follows from the operating-profit term alone; moreover, in the three-stage model under boundary condition (B2) the recursion terminates at the linear terminal value J^*_3 (whose curvature is zero), so for the only interior optimization on the empirical trajectory the relevant curvature comes entirely from the displayed operating-profit term, with no recursive contribution. Either way the future value function is *convex* in size in the interior regime. The reinvestment feedback term in (26) is therefore strictly positive, and the second-order condition could in principle fail. Steps 2–5 derive a sufficient condition for it to hold.

Step 2: First-order-condition substitution. The endogenous quantity $\partial\Pi_O/\partial\theta_t$ in (26) can be eliminated at the optimum using the first-order condition. Rearranging (23):

$$\Lambda_t \cdot \frac{\partial\Pi_O}{\partial\theta_t} = -\beta v_I \frac{\partial I_t}{\partial\theta_t} \equiv \mathbf{ID}(n_t).$$

Define $\mathbf{ID}(n) \equiv -\beta v_I \cdot \partial I/\partial\theta$ as the **initiative drag** at firm size n: the discounted marginal innovation loss from centralization. Using $\partial I/\partial\theta = -\sigma n^\zeta (v_I + b)/k$:

$$\mathbf{ID}(n) = \frac{\sigma\beta v_I (v_I + b) n^\zeta}{k}.$$

The initiative drag is zero in manufacturing ($\sigma = 0$, where centralization does not affect R&D) and grows as n^ζ in services ($\sigma = 1$). At the interior optimum, the first-order condition equates the growth-amplified centralization payoff $\Lambda_t \cdot \partial\Pi_O/\partial\theta_t$ with the initiative drag $\mathbf{ID}(n_t)$, this is the model's central tension, expressed as an equality at the optimum.

Substituting $(\partial\Pi_O/\partial\theta_t)^2 = \mathbf{ID}(n_t)^2/\Lambda_t^2$ into (26):

$$\Lambda_t^3 \cdot \kappa_\theta(n_t) \ > \ \beta\phi^2 \cdot \frac{d^2 J_{t+1}^*}{dn_{t+1}^2} \cdot \mathbf{ID}(n_t)^2. \tag{28}$$

This step is exact, no slack has been introduced. The condition compares growth-amplified coordination convexity to value-function convexity in n times squared initiative drag. Note that the initiative drag now appears *squared*, because the reinvestment feedback involved the marginal payoff $(\partial\Pi_O/\partial\theta_t)^2$ appearing twice (once for each transmission through ϕ), and the first-order-condition substitution turned each appearance into a copy of $\mathbf{ID}$.

Step 3: Conservative bound on the value-function curvature. The interior-regime convexity (27) is bounded above using the interior-regime constraints $\Delta_A n_{t+1} \leq F$ and

$F - n_{t+1}\Delta_A \leq F$. Each bracketed term in (27) satisfies $2\Delta_A^2 n_{t+1}^2 \leq 2F^2$, $4\alpha\Delta_A(F - n_{t+1}\Delta_A)n_{t+1} \leq 4\alpha F^2$, and $\alpha(\alpha+1)(F - n_{t+1}\Delta_A)^2 \leq \alpha(\alpha+1)F^2$. Summing:

$$\frac{d^2 J_{t+1}^*}{dn_{t+1}^2} \leq \frac{(\alpha^2 + 5\alpha + 2)F^2}{4\delta n_{t+1}^{\alpha+2}} = \frac{(\alpha^2 + 5\alpha + 2)F^2}{2n_{t+1}^2 \cdot \kappa_\theta(n_{t+1})}. \tag{29}$$

The second equality uses $4\delta n_{t+1}^{\alpha+2} = 2n_{t+1}^2 \cdot 2\delta n_{t+1}^{\alpha} = 2n_{t+1}^2 \cdot \kappa_\theta(n_{t+1})$, expressing the bound in κ_θ-units. Under (B2) the recursion terminates at the linear terminal value J_3^*, so this bound applies with no recursive contribution; in any longer-horizon variant the recursive contribution from βJ_{t+2}^* is expected to satisfy the same form of bound by induction (a sketch: the substitution uses the static interior policy at each stage, which is exact only when the following stage is terminal, as in the three-stage model; none of the results in this paper rely on the longer-horizon extension).

The bound introduces slack in two ways. First, the substitution $F - n_{t+1}\Delta_A \leq F$ is tight only when the firm is deep in the interior regime ($n_{t+1}\Delta_A \ll F$); for firms approaching saturation, the actual quantity $F - n_{t+1}\Delta_A$ is small but the bound replaces it with the much larger F. Second, the recursive induction step accumulates further slack at each level. Both effects *overstate* the reinvestment-feedback force on the right-hand side of (26), which means the resulting sufficient condition *understates* how easily the second-order condition is satisfied. Why $\kappa_\theta(n_{t+1})$ ends up in the denominator: a steeper next-stage coordination convexity pins down θ_{t+1}^* more tightly, which keeps J_{t+1}^* from being too convex in n_{t+1}, so larger next-stage convexity damps the reinvestment feedback.

Substituting the bound into (28):

$$\Lambda_t^3 \cdot \kappa_\theta(n_t) \; > \; \frac{\beta\phi^2(\alpha^2 + 5\alpha + 2)F^2 \mathbf{ID}(n_t)^2}{2n_{t+1}^2 \cdot \kappa_\theta(n_{t+1})}.$$

Step 4: Conservative bound on the growth multiplier. The growth multiplier $\Lambda_t = 1 + \beta\phi \cdot dJ_{t+1}^*/dn_{t+1}$ satisfies $\Lambda_t \geq 1$ whenever the marginal value of size is non-negative, $dJ^*/dn \geq 0$. This holds exactly in the saturated regime, where $dJ^*/dn = \pi_0 + \Delta_A >$

0; and numerical backward induction confirms $dJ^*/dn > 0$ (so $\Lambda_t \approx 1.05$) all along the calibrated interior trajectory, so the interior-regime use of $\Lambda_t \geq 1$ here is verified rather than assumed (consistent with this being a robustness exercise relative to the saturated regime). Setting $\Lambda_t^3 \to 1$ on the left-hand side gives a strictly more conservative condition. At the calibrated parameters $\Lambda_t \approx 1.05$, so $\Lambda_t^3 \approx 1.16$, this step discards a roughly 16% multiplicative buffer that does in fact help satisfy the second-order condition. The reason to discard it: the resulting condition contains no endogenous quantities and is directly verifiable from primitives.

After this step:

$$\kappa_\theta(n_t) \cdot \kappa_\theta(n_{t+1}) \cdot n_{t+1}^2 > \frac{\beta\phi^2(\alpha^2 + 5\alpha + 2)}{2} \cdot F^2 \cdot \mathbf{ID}(n_t)^2.$$

Step 5: Dimensionless form (pointwise). The condition becomes interpretable when rescaled by the model's natural units. Define $\tilde{n} \equiv n\Delta_A/F$, *firm size relative to the saturation threshold* $\bar{n} \equiv F/\Delta_A$ (equal to 1 at the threshold, less than 1 in the interior regime). Define $\eta \equiv \phi\Delta_A$, the *reinvestment elasticity* (ϕ converts profit to headcount, Δ_A converts headcount to per-worker surplus, their product is dimensionless). Multiplying both sides of the previous inequality by Δ_A^2/F^2 and dividing by $\kappa_\theta(n_t)\kappa_\theta(n_{t+1})$:

$$\tilde{n}_{t+1}^2 > \frac{\beta\eta^2(\alpha^2 + 5\alpha + 2)}{2} \cdot \frac{\mathbf{ID}(n_t)^2}{\kappa_\theta(n_t) \cdot \kappa_\theta(n_{t+1})}. \tag{30}$$

This is a *pointwise* sufficient condition: it must hold at each interior-regime stage along the firm's equilibrium trajectory. The right-hand side still contains the endogenous quantities n_t and n_{t+1}.

Step 6: Uniform bound on the trajectory. A primitive-only condition is obtained by replacing each endogenous quantity in (30) by its worst-case value over the interior-regime trajectory. The interior regime is defined by $n_t\Delta_A < F$, equivalently $n_t < \bar{n}$. Let $\underline{n}$ denote the lower endpoint of the interior portion of the firm's trajectory: under the boundary conditions of Proposition 1, $\underline{n} = n_D^*$ (the firm enters the interior regime when it first leaves the corner $\theta^* = 0$); without (B1), $\underline{n} = n_1$ (the firm may be born interior). Along the interior portion, $\underline{n} \le n_t \le n_{t+1} < \bar{n}$. Replace each endogenous quantity by the worst-case value that maximizes the right-hand side:

- The initiative drag $\mathbf{ID}(n_t) = \sigma\beta v_I(v_I+b)n_t^\zeta/k$ is increasing in n_t, so $\mathbf{ID}(n_t) < \mathbf{ID}(\bar{n}) = \sigma\beta v_I(v_I+b)\bar{n}^\zeta/k$. Using $\bar{n}^\zeta = (F/\Delta_A)^\zeta$:

$$\mathbf{ID}(n_t) \;<\; \mathbf{ID}(\bar{n}) \;=\; \frac{\sigma\beta v_I(v_I+b)}{k} \cdot \left(\frac{F}{\Delta_A}\right)^\zeta .$$

  This is the *maximum initiative drag* the firm experiences over the interior regime.

- The coordination-cost convexities $\kappa_\theta(n_t) = 2\delta n_t^\alpha$ and $\kappa_\theta(n_{t+1}) = 2\delta n_{t+1}^\alpha$ are increasing in n, so $\kappa_\theta(n_t), \kappa_\theta(n_{t+1}) \ge \kappa_\theta(\underline{n}) = 2\delta\underline{n}^\alpha$. This is the *minimum coordination convexity* over the trajectory.

- The left-hand side $\tilde{n}_{t+1}^2$ is increasing in n_{t+1}, so $\tilde{n}_{t+1} \ge \underline{\tilde{n}} = \underline{n}\Delta_A/F$, and the smallest value is $\underline{\tilde{n}}^2$.

Substituting these worst-case bounds into (30):

$$\underline{\tilde{n}}^2 \;>\; \frac{\beta\eta^2(\alpha^2+5\alpha+2)}{2} \cdot \frac{\mathbf{ID}(\bar{n})^2}{\kappa_\theta(\underline{n})^2}. \tag{31}$$

This is a uniform sufficient condition: if it holds for the relevant lower endpoint $\underline{n}$, then (30) holds at every stage of the interior-regime trajectory.

Step 7: Two threshold forms. Substituting the explicit primitive expressions $\underline{\tilde{n}} = \underline{n}\Delta_A/F$, $\mathbf{ID}(\bar{n}) = \sigma\beta v_I(v_I+b)(F/\Delta_A)^\zeta/k$, $\kappa_\theta(\underline{n}) = 2\delta\underline{n}^\alpha$, and $\eta = \phi\Delta_A$ into (31):

$$\left(\frac{\underline{n}\Delta_A}{F}\right)^2 \;>\; \frac{\beta(\phi\Delta_A)^2(\alpha^2+5\alpha+2)}{2} \cdot \frac{\sigma^2\beta^2 v_I^2(v_I+b)^2F^{2\zeta}/(k^2\Delta_A^{2\zeta})}{(2\delta\underline{n}^\alpha)^2}.$$

Simplifying and cancelling Δ_A^2:

$$\frac{\underline{n}^2}{F^2} > \frac{\sigma^2\beta^3\phi^2(\alpha^2+5\alpha+2)v_I^2(v_I+b)^2F^{2\zeta}}{8\delta^2k^2\underline{n}^{2\alpha}\Delta_A^{2\zeta}}.$$

Rearranging to isolate $\underline{n}$ on the left-hand side:

$$\underline{n}^{2\alpha+2} > \frac{\sigma^2\beta^3\phi^2(\alpha^2+5\alpha+2)v_I^2(v_I+b)^2F^{2(\zeta+1)}}{8\delta^2k^2\Delta_A^{2\zeta}}.$$

Define the *regularity threshold* explicitly in primitives:

$$n^{**} \equiv \left[\frac{\sigma^2\beta^3\phi^2(\alpha^2+5\alpha+2)v_I^2(v_I+b)^2\cdot F^{2(\zeta+1)}}{8\delta^2k^2\cdot\Delta_A^{2\zeta}}\right]^{1/(2\alpha+2)}. \tag{32}$$

Taking the $(2\alpha+2)$-th root, the regularity condition takes the form $\underline{n} > n^{**}$. Specializing the lower endpoint $\underline{n}$ to its two cases gives two distinct regularity conditions.

Boundary-restricted regularity condition (under (B1)). Under boundary condition (B1) of Proposition 1, the firm is born at the corner $\theta_1^* = 0$ and does not begin interior optimization until it has grown past the full decentralization threshold n_D^* (derived in Appendix A.5). The interior portion of the trajectory therefore lives on $[n_D^*, \bar{n})$, so $\underline{n} = n_D^*$ and the regularity condition is:

$$\boxed{n_D^* > n^{**}} \qquad \text{(boundary-restricted regularity condition, under (B1)).} \tag{33}$$

This is the regularity condition relevant for the empirical exercise in this paper, where the firm is a scaling venture and (B1) is satisfied at the relevant initial size.

General regularity condition. Without (B1), the firm could in principle be born interior. The interior portion of the trajectory then lives on $[n_1, \bar{n})$, so $\underline{n} = n_1$ and the regularity condition is:

$$\boxed{n_1 > n^{**}} \qquad \text{(general regularity condition).} \tag{34}$$

Which condition applies in which context. The boundary-restricted condition is weaker whenever (B1) holds, since $n_D^* > n_1$ implies $n_D^* > n^{**}$ whenever $n_1 > n^{**}$, but not vice versa. The boundary-restricted condition can hold even when $n_1 < n^{**}$, as long as the firm has grown past n^{**} by the time it enters the interior regime. The empirical exercise in this paper invokes (B1), so the boundary-restricted condition (33) is the relevant check for the interior comparative statics in Proposition 2(a)–(b) and Proposition 3. The general condition (34) is reported for completeness: it certifies the comparative statics in any model variant where the firm is allowed to be born interior, such as new establishments of multi-unit firms that inherit the parent's high θ (Result (d) of Proposition 4). At the calibrated services parameters, $n^{**} \approx 15$ and $n_{D,S}^* \approx 108$, so the boundary-restricted condition holds with substantial margin (ratio approximately 7).

The regularity threshold n^{**} itself depends only on the model's primitives. Its structure is the same in both conditions; only the variable being compared to it changes with the assumed boundary condition.

How the regularity threshold scales. Each primitive in n^{**} enters with a specific economic role. The threshold rises (regularity becomes harder) with primitives that strengthen the initiative drag or expand the interior regime:

- *Innovation channel strength* $\sigma^2 v_I^2 (v_I + b)^2$: a stronger innovation force means greater initiative drag, so the firm must be larger to anchor its centralization choice against the dynamic reinvestment feedback. In manufacturing ($\sigma = 0$), $n^{**} = 0$ and the second-order condition holds unconditionally: with no innovation channel, there is no initiative drag for the reinvestment feedback to amplify, and the static coordination convexity carries the optimum on its own.

- *Patience* β^3: more patient firms weigh future innovation losses more heavily. One power of β enters through the initiative drag's discount factor; two more powers enter through the dynamic feedback's compounded discounting, since the feedback runs through both the next-stage value function and its derivative.

- *Reinvestment rate* ϕ^2: faster reinvestment amplifies the dynamic feedback. Two powers of ϕ enter because the centralization choice affects the second-order condition through a double composition, once through the growth equation ($\theta_t \to n_{t+1}$) and once through the growth multiplier's dependence on the next-stage value function ($n_{t+1} \to \Lambda_t$).

- *Overhead cost* $F^{2(\zeta+1)}$: higher overhead pushes the saturation threshold $\bar{n} = F/\Delta_A$ upward, enlarging the interior regime over which the uniform bound must hold and raising the worst-case initiative drag $\mathbf{ID}(\bar{n})$.

The threshold falls (regularity becomes easier) with primitives that strengthen the coordination convexity or weaken the innovation channel:

- *Coordination cost magnitude* $\delta^{-2/(2\alpha+2)}$: steeper coordination costs raise the coordination convexity $\kappa_\theta(\underline{n}) = 2\delta\underline{n}^\alpha$ uniformly, anchoring the centralization choice more firmly against perturbations from reinvestment feedback.

- *Cost of effort* $k^{-2/(2\alpha+2)}$: a higher cost of effort makes innovation more expensive to pursue at the margin, which lowers the initiative drag $\mathbf{ID}(n)$ directly through its $1/k$ scaling.

- *Alignment gain* $\Delta_A^{-2\zeta/(2\alpha+2)}$: higher alignment lowers the saturation threshold $\bar{n} = F/\Delta_A$, shrinking the interior regime over which the uniform bound must hold.

The exponent governing the threshold's sensitivity to all primitives is $1/(2\alpha+2)$. As α grows, this exponent shrinks: coordination costs that scale rapidly in n overwhelm everything else, so the firm needs only a modest size to enter the regime where the second-order condition holds, and n^{**} becomes insensitive to the other primitives.

At the calibrated parameters $n^{**} \approx 15$, and the boundary-restricted condition $n^*_{D,S} > n^{**}$ holds with $n^*_{D,S} \approx 108$ (a ratio of about 7). Under (B1) the firm sits at the corner $\theta_1^* = 0$ until it grows past n^*_D, so no interior optimization (and no second-order condition) is at issue below n_D^*; the boundary-restricted condition then certifies the remainder of the trajectory. The general condition $n_1 > n^{**}$ fails marginally at the smallest empirical bin ($n_1 = 7 < 15$), but this is immaterial for the empirical exercise, which invokes (B1).

Three takeaways. *(a)* The regularity condition says the interior critical point is a profit-maximizing maximum rather than a corner collapse to $\theta = 0$; it is easier to meet when the initiative drag is small or the coordination convexity large, and since $\mathbf{ID}(n) \sim n^\zeta$ while $\kappa_\theta(n) \sim n^\alpha$ with $\alpha > 2 > \zeta$, coordination convexity overtakes initiative drag as firms scale, so the condition becomes progressively easier with size. *(b)* The condition is sufficient, not necessary, and conservative in four places: the bound $F - n_{t+1}\Delta_A \leq F$ (Step 3), the recursive treatment of βJ_{t+2}^* (Step 3), $\Lambda_t^3 \to 1$ (Step 4, a $\approx 16\%$ buffer), and the worst-case uniform substitution (Step 6). Failing (33) or (34) therefore does not mean the second-order condition fails; the pointwise condition (30) is a much tighter, trajectory-verifiable check. *(c)* The interior-regime analysis is a robustness exercise: in the empirically relevant saturated regime J_{t+1}^* is linear in n_{t+1}, the reinvestment feedback is exactly zero, and the second-order condition holds unconditionally, so the interior condition is not required for any empirical result.

In both regimes $\partial F/\partial \theta_t < 0$, and the comparative statics below are derived without specializing to either.

Comparative statics. By the implicit function theorem, $d\theta_t^*/dx = -(\partial F/\partial x)/(\partial F/\partial \theta_t)$. Since $\partial F/\partial \theta_t < 0$, the sign of $d\theta_t^*/dx$ equals the sign of $\partial F/\partial x$. I derive the sign of $\partial F/\partial x$ for each comparative static below.

Result (a): $\partial \theta_t^*/\partial n_t > 0$. Increasing n_t affects F through both $\partial \Pi_O/\partial \theta_t$ and $\partial I_t/\partial \theta_t$. The first piece is $\partial \Pi_O/\partial \theta_t = n_t \Delta_A - F + 2\delta n_t^\alpha (1 - \theta_t)$, so $\partial^2 \Pi_O/\partial \theta_t \partial n_t = \Delta_A + 2\alpha\delta n_t^{\alpha-1}(1 - \theta_t) > 0$: larger n_t raises both the marginal alignment gain and the marginal coordination cost reduction, both pushing toward more centralization. The second piece is $\partial I_t/\partial \theta_t = -\sigma n_t^\zeta (v_I + b)/k$, so $\partial^2 I_t/\partial \theta_t \partial n_t = -\sigma\zeta n_t^{\zeta-1}(v_I + b)/k$, which is negative for services ($\sigma = 1$)

and zero for manufacturing ($\sigma = 0$). For services, larger n_t amplifies the (negative) marginal innovation effect of centralization, which works against centralization. But because $\alpha > 2 > \zeta$, the Π_O effect grows faster in n_t than the innovation effect. The growth multiplier Λ_t also depends on n_t via $n_{t+1}(\theta_t)$, but this channel is itself nonnegative at an interior optimum on a growing path (where $\partial\Pi_O/\partial\theta_t > 0$, $d\Lambda_t/dn_{t+1} \geq 0$, and $\partial n_{t+1}/\partial n_t = 1 + \phi\,\partial\Pi_O/\partial n_t \geq 0$; the last requires only $\partial\Pi_O/\partial n_t \geq -1/\phi$, a very weak restriction that holds at all calibrated sizes, where $\partial\Pi_O/\partial n_t > 0$). Dropping the two nonnegative reinforcing channels, a sufficient condition for $\partial F/\partial n_t > 0$ is $\Lambda_t\Delta_A \geq \beta\sigma\zeta n_t^{\zeta-1} v_I(v_I + b)/k$, which (using $\Lambda_t \geq 1$) holds for all

$$n_t \;\geq\; n^{(a)} \;\equiv\; \left[\frac{\beta\sigma\zeta\, v_I(v_I + b)}{k\,\Delta_A}\right]^{\frac{1}{1-\zeta}}.$$

At the calibrated parameters $n^{(a)} \approx 51$, which lies strictly below the lower edge of the interior regime $n^*_{D,S} \approx 108$ (Appendix B.11); the condition therefore holds at every point of the interior regime, and θ^*_t increases with n_t along the entire trajectory. In the saturated regime the result is exact and immediate from the closed form (14): the numerator carries the per-worker innovation force $n^{\zeta-1}$ and the per-worker overhead F/n, both decreasing in n, while the denominator $2\delta n^{\alpha-1}$ is increasing (using $\alpha - 1 > 0 > \zeta - 1$), so $1 - \theta^*$ falls monotonically and θ^* rises.

Proposition 3(a): $\partial\theta^*_t/\partial\beta < 0$. The parameter β enters F in two places: directly multiplying the innovation term, and through $\Lambda_t(n_{t+1}) = 1 + \beta\phi \cdot dJ^*_{t+1}/dn_{t+1}$. The direct effect is $\partial(\beta v_I \cdot \partial I_t/\partial\theta_t)/\partial\beta = v_I \cdot \partial I_t/\partial\theta_t < 0$ for services: a more patient firm weighs the innovation loss from centralization more heavily, pushing toward decentralization. The indirect effect through Λ_t is $\phi \cdot dJ^*_{t+1}/dn_{t+1} > 0$ multiplied by $\partial\Pi_O/\partial\theta_t$ (positive at an interior optimum), so it pushes toward centralization. The two effects oppose, but the direct innovation effect dominates: Λ_t is bounded above (by $1+\beta\phi(\pi_0+\Delta_A)$ in the saturated regime), so the indirect effect is bounded, while the direct effect scales with n_t^ζ and is large. Hence $\partial F/\partial\beta < 0$ in the relevant range (a dominance argument for the interior regime; verified numerically along the calibrated interior trajectory in the replication notebook), and θ_t^* decreases with β. In the saturated regime this is exact: β enters the closed form (14) only through the numerator factor $\beta/\Lambda = \beta/[1 + \beta\phi(\pi_0 + \Delta_A)]$, whose derivative $1/\Lambda^2 > 0$ raises the numerator and hence $1 - \theta^*$, so θ^* falls in β unambiguously.

Proposition 3(b): $\theta_{t,S}^* < \theta_{t,M}^*$. The separability parameter σ enters F only through the innovation derivative: $\partial I_t/\partial\theta_t = -\sigma n_t^\zeta(v_I + b)/k$. With $\sigma = 1$ (services), this term is strictly negative; with $\sigma = 0$ (manufacturing), it is zero. The other term $\Lambda_t \cdot \partial\Pi_O/\partial\theta_t$ is identical across sectors. Hence at any candidate θ_t, $F(\theta_t; \sigma = 1) < F(\theta_t; \sigma = 0)$: the services firm's first-order condition is satisfied at a smaller θ_t than the manufacturing firm's, so $\theta_{t,S}^* < \theta_{t,M}^*$ whenever at least one sector's optimum is interior. At sizes where both policies sit at the same corner ($n \leq n_{D,M}^*$, where both equal 0, or $n \geq n_{C,S}^*$, where both equal 1), the two coincide and the inequality holds weakly.

Proposition 3(c): $\partial\theta_t^*/\partial\Delta_A > 0$. Δ_A enters F in two reinforcing places. First, directly in $\partial\Pi_O/\partial\theta_t$: differentiating $\partial\Pi_O/\partial\theta_t = n_t\Delta_A - F + 2\delta n_t^\alpha(1-\theta_t)$ with respect to Δ_A gives $n_t > 0$, raising the marginal alignment gain. Second, through Λ_t: $\partial\Lambda_t/\partial\Delta_A > 0$ (a higher alignment gain raises the marginal value of size at the next stage), amplifying the growth multiplier on Π_O. Both channels make $\partial F/\partial\Delta_A > 0$, so θ_t^* increases with Δ_A: higher worker–firm misalignment makes centralization more attractive.

A.4 Closed-form solution for large firm size and the Stage-1 correction

This subsection derives the closed-form expression for θ^* stated in Proposition 2 (equation 14), quantifies the Stage-1 correction to it, and establishes the age-irrelevance result (Proposition 2(b)).

When the firm operates in the saturated regime (Section A.2), the growth multiplier Λ

is a parameter and the penultimate-stage first-order condition collapses to:

$$\Lambda[n\Delta_A - F + 2\delta n^\alpha(1-\theta)] = \frac{\beta v_I(v_I + b)n^\zeta}{k}. \tag{35}$$

This is a rearrangement of equation (23) at $\Lambda_t = \Lambda$ and $\mu_{t+1} = \mu_3 = 1$, using $\partial\Pi_O/\partial\theta = n\Delta_A - F + 2\delta n^\alpha(1-\theta)$ and $\partial I/\partial\theta = -n^\zeta(v_I + b)/k$ for services ($\sigma = 1$). The equation is linear in θ and can be solved explicitly.

Services closed form. Isolating $(1-\theta)$ in (35):

$$\begin{aligned} 2\Lambda\delta n^\alpha(1-\theta) &= \frac{\beta v_I(v_I + b)n^\zeta}{k} - \Lambda(n\Delta_A - F), \\ 1 - \theta &= \frac{\beta v_I(v_I + b)n^\zeta/(k\Lambda) - n\Delta_A + F}{2\delta n^\alpha}. \end{aligned}$$

Dividing numerator and denominator by n and clamping to $[0, 1]$ yields equation (14) of Proposition 2:

$$1 - \theta_S^*(n) = \left[\frac{\beta v_I(v_I + b)n^{\zeta-1}/(k\Lambda) - \Delta_A + F/n}{2\delta n^{\alpha-1}}\right]_0^1.$$

Manufacturing closed form. For manufacturing ($\sigma = 0$), the innovation term in (35) vanishes because $\partial I/\partial\theta = 0$: centralizing the firm's operations has no effect on R&D, so there is no innovation cost to decentralization. The first-order condition reduces to $\Lambda[n\Delta_A - F + 2\delta n^\alpha(1-\theta)] = 0$, and since $\Lambda > 0$:

$$1 - \theta_M^*(n) = \left[\frac{F/n - \Delta_A}{2\delta n^{\alpha-1}}\right]_0^1.$$

When $n\Delta_A \geq F$, the right-hand side is non-positive and the firm fully centralizes: $\theta_M^* = 1$. When $n\Delta_A < F$, the solution is interior.

Stage-1 correction. Equation (35) is the *penultimate-stage* condition. At Stage 1 the innovation cash arrives at Stage 2 and is reinvested once more, so the exact Stage-1 condition under the saturated chain (terminal-stage saturation together with Stage-2 saturation $\theta_2^* = 1$, which in services requires $n_2 \geq n_{C,S}^*$; along calibrated trajectories from interior sizes this premise fails and the resulting form is an approximation, quantified below) carries the multipliers derived in Section A.2: $\Lambda_1[n\Delta_A - F + 2\delta n^\alpha(1-\theta)] = \beta\Lambda\, v_I(v_I+b)n^\zeta/k$ with $\Lambda_1 = 1+\beta\phi(\pi_0+\Delta_A)(\Lambda+\beta)$ from (24). Solving as above gives the same clamped form with the innovation weight $\beta\Lambda/(k\Lambda_1)$ replacing $\beta/(k\Lambda)$:

$$1 - \theta_{S,1}^*(n) = \left[\frac{\beta(\Lambda/\Lambda_1)\, v_I(v_I+b)\, n^{\zeta-1}/k - \Delta_A + F/n}{2\delta n^{\alpha-1}}\right]_0^1.$$

Quantitatively, the correction is small but not literally zero. At the calibrated parameters (Appendix B), $\Lambda = 1.0503$ and $\Lambda_1 = 1.0780$, so the Stage-1 weight is $\Lambda/\Lambda_1 = 0.974$ against the penultimate-stage weight $1/\Lambda = 0.952$: the Stage-1 innovation force is about 2.3% *stronger*, so the Stage-1 policy is very slightly less centralized at any interior size. Verified against the exact numerical backward induction of the full three-stage problem (replication notebook), the penultimate-stage closed form (14) reproduces the exact Stage-2 policy with maximum error below 3×10^{-6} across the empirical size bins. Using it for Stage 1 overstates θ_1^* by at most 0.008 across the bins; on a dense size grid the deviation peaks at approximately 0.017 just above the regime boundary (the exact Stage-1 policy leaves the corner at $n \approx 108.4$, slightly later than the closed form's $n_D^* \approx 107.5$) and exceeds 0.01 on roughly $n \in [108, 143]$ before decaying. The corrected Stage-1 form above reduces the error to below 7×10^{-4} everywhere, and below 10^{-4} at every empirical size bin (maximum 6×10^{-5}, at the 175 bin) and for all $n \gtrsim 160$; the residual near the boundary reflects the fact that the constant-Λ_1 premise (Stage-2 saturation) is furthest from holding there, where the continuation policy θ_2^* is interior. Because Λ/Λ_1 is a positive constant in $(0,1)$, all comparative-static signs of the closed form are preserved at Stage 1; in particular, the β-monotonicity of Proposition 3(a) carries over, since the Stage-1 weight $\beta\Lambda/\Lambda_1$ is, like β/Λ, strictly increasing in β over the admissible range (verified analytically from (24) and numerically in the replication notebook).

For manufacturing the correction is moot: the innovation term is absent, the multipliers cancel, and the Stage-1 and Stage-2 policies coincide exactly. Throughout the calibration I use the penultimate-stage closed form, which is the exact policy for the stage being matched to the cross-sectional data and accurate to within 0.008 for the remaining interior stage.

Age irrelevance. The first-order condition (35), and likewise its Stage-1 analogue above, depends on n and the structural parameters $(v_I, b, k, \Delta_A, \delta, \alpha, \zeta, F, \beta, \phi, \pi_0)$, but not on firm age. Two firms of different ages and the same size n therefore choose the same θ^*. $\square$

Myopic limit (vanishing patience). As $\beta \to 0$: the innovation force $\beta v_I(v_I+b)n^{\zeta-1}/(k\Lambda) \to 0$ and both multipliers collapse, $\Lambda \to 1$ and $\Lambda_1 \to 1$. The services first-order condition becomes $n\Delta_A - F + 2\delta n^\alpha(1-\theta) = 0$ at every interior stage, identical to the manufacturing first-order condition. The sectoral difference vanishes: $\theta_S^* \to \theta_M^*$. For $n \geq F/\Delta_A$ both sectors give $\theta^* = 1$. For $n < F/\Delta_A$, the overhead force $F/n > \Delta_A$ keeps the firm partially decentralized even at $\beta = 0$, not to preserve innovation (which the myopic firm does not value) but because the per-worker overhead exceeds the alignment gain. $\square$

Full centralization threshold. The closed form clamps at both bounds. For very small n, the per-worker overhead F/n and (in services) the per-worker innovation force $n^{\zeta-1}$ are both large, the numerator $N(n)$ exceeds the denominator $D(n) = 2\delta n^{\alpha-1}$ (which is small for small n since $\alpha > 2$), and $1-\theta^*$ clamps at 1, giving $\theta^* = 0$. For very large n, both F/n and $n^{\zeta-1}$ vanish, $N(n)$ becomes negative (driven by $-\Delta_A$), and $1-\theta^*$ clamps at 0, giving $\theta^* = 1$. The full centralization threshold n_C^* is the smallest n at which $\theta^* = 1$, equivalently where $N(n_C^*) = 0$: for services, $\beta v_I(v_I+b)(n_C^*)^{\zeta-1}/(k\Lambda) + F/n_C^* = \Delta_A$; for manufacturing, $n_{C,M}^* = F/\Delta_A$ since the innovation force is absent. Both left-side terms in the services equation are positive and decreasing in n, so the equation has a unique solution, and the services threshold is larger than the manufacturing threshold: $n_{C,S}^* > n_{C,M}^*$. Services firms therefore remain partially decentralized over a wider range of sizes than manufacturing firms. □

Performance pay. Under Aghion–Tirole (1997), worker initiative $e^* = (1-\theta)(v_I + b)/k$ is decreasing in θ. To maintain effort on execution tasks as decentralization falls, the firm substitutes toward performance pay, which is complementary to centralization. Since $\theta^*(n)$ is increasing in n, performance pay adoption rises with firm size, consistent with the empirical pattern in Figure 3. □

A.5 Full decentralization threshold and venture comparative statics

This subsection defines the full decentralization threshold n_D^* used in the statement of Proposition 1 (boundary condition B1) and derives the threshold comparative statics in Proposition 3(a)–(c).

For firm sizes below the full decentralization threshold n_D^*, the optimal centralization is $\theta^* = 0$: the firm is fully decentralized and the closed-form numerator clamps at the upper bound $N(n) \geq D(n)$. The threshold n_D^* marks the largest size at which the firm remains fully decentralized; for n marginally above n_D^*, the firm begins centralizing.

At the threshold, the closed-form clamp transitions from active to inactive. Define:

$$G(n) \equiv N(n) - D(n) = \frac{\sigma\beta v_I(v_I + b)n^{\zeta-1}}{k\Lambda} - \Delta_A + \frac{F}{n} - 2\delta n^{\alpha-1}.$$

The threshold n_D^* is the unique positive root of $G(n_D^*) = 0$ (uniqueness follows because G is strictly decreasing: $\partial G/\partial n < 0$, since each of the four terms in G either decreases in n or has zero derivative, and at least one strictly decreases). For services ($\sigma = 1$):

$$\frac{\beta v_I(v_I + b)(n_{D,S}^*)^{\zeta-1}}{k\Lambda} + \frac{F}{n_{D,S}^*} - \Delta_A \;=\; 2\delta(n_{D,S}^*)^{\alpha-1}.$$

For manufacturing ($\sigma = 0$):

$$\frac{F}{n^*_{D,M}} - \Delta_A \;=\; 2\delta(n^*_{D,M})^{\alpha-1}.$$

The services innovation force is positive and the manufacturing innovation force is absent, so at any given n the services left-hand side is larger than the manufacturing left-hand side. Since both sides are decreasing in n in the relevant range (the left side at rate $\zeta - 1 < 0$ for innovation and -1 for overhead, the right side at rate $\alpha - 1 > 1$), the services threshold lies to the right: $n^*_{D,S} > n^*_{D,M}$. Services firms remain fully decentralized over a wider size range than manufacturing firms, the innovation channel gives the small services firm an additional reason to stay decentralized that the manufacturing firm does not have.

Comparative statics on the decentralization threshold via the implicit function theorem. By the implicit function theorem applied to $G(n^*_D; x) = 0$:

$$\frac{dn^*_D}{dx} = -\frac{\partial G/\partial x}{\partial G/\partial n}\bigg|_{n=n^*_D}.$$

The denominator $\partial G/\partial n < 0$ (established above), so the sign of dn^*_D/dx equals the sign of $\partial G/\partial x$. Each primitive of interest enters through a single channel or a small number of reinforcing channels.

Patience (β): β enters only through the innovation-force term $\beta v_I(v_I + b)n^{\zeta-1}/(k\Lambda)$, both directly and through $\Lambda = 1 + \beta\phi(\pi_0 + \Delta_A)$. The direct effect is $\partial[\beta/\Lambda]/\partial\beta = 1/\Lambda^2 > 0$, so the innovation force is increasing in β. Therefore $\partial G/\partial\beta > 0$, and:

$$\frac{dn^*_D}{d\beta} > 0.$$

More patient firms stay fully decentralized to larger sizes. The economic meaning is that patient ownership preserves the high-initiative phase: by valuing future innovation more, patient firms find decentralization more attractive at any given size, postponing the onset of centralization.

Separability (σ): σ enters multiplicatively in the innovation-force term and not elsewhere. $\partial G/\partial\sigma = \beta v_I(v_I + b)n^{\zeta-1}/(k\Lambda) > 0$ in the interior. Therefore $dn_D^*/d\sigma > 0$, and services firms ($\sigma = 1$) have larger decentralization thresholds than manufacturing firms ($\sigma = 0$): $n_{D,S}^* > n_{D,M}^*$. Smaller services ventures stay decentralized longer because centralizing their operations would simultaneously curtail their innovation.

Incentive misalignment (Δ_A): Δ_A enters G in two reinforcing places. Directly: $\partial G/\partial\Delta_A = -1-\beta\phi\cdot$[innovation force divided by Λ] < 0 (the second piece comes from $\partial\Lambda/\partial\Delta_A = \beta\phi > 0$ raising Λ and thereby reducing the innovation force). Therefore $dn_D^*/d\Delta_A < 0$. Firms with greater incentive misalignment exit full decentralization at smaller sizes: when worker preferences diverge sharply from firm objectives, even small firms find centralization attractive.

Effort cost (k) and innovation value (v_I): k enters the innovation force as $1/k$, so $\partial G/\partial k < 0$ and $dn_D^*/dk < 0$. Higher effort costs make innovation more expensive to pursue at the margin, shrinking the decentralization regime. Conversely, higher v_I raises the innovation force, so $\partial G/\partial v_I > 0$ and $dn_D^*/dv_I > 0$. Ventures with greater innovation rents stay decentralized longer.

Overhead cost (F): F enters G directly through $F/n > 0$. $\partial G/\partial F = 1/n_D^* > 0$, so $dn_D^*/dF > 0$. Higher overhead extends the decentralization regime because the fixed cost of centralization is harder for small firms to amortize.

Coordination cost magnitude (δ): δ enters G through $-2\delta n^{\alpha-1}$. $\partial G/\partial\delta = -2(n_D^*)^{\alpha-1} < 0$, so $dn_D^*/d\delta < 0$. Steeper coordination costs make centralization attractive at smaller sizes, shortening the decentralization regime.

Summary of venture comparative statics. For firms in the venture range ($n < n_D^*$), the comparative statics operate at the level of the threshold rather than the level of θ^* (which is pinned at zero). The signs are:

- Threshold rises with: patience β, separability σ, innovation value v_I, overhead F.
- Threshold falls with: misalignment Δ_A, effort cost k, coordination magnitude δ.

The signs of these effects on n_D^* are exactly the signs of the corresponding effects on θ^* in the interior regime, but *reversed*: a primitive that raises θ^* at interior sizes lowers n_D^*, and vice versa. This is intuitive: a primitive that makes centralization more attractive at any size both raises θ^* where the firm is interior and brings forward the size at which the firm starts to centralize.

A.6 Boundary conditions for the lifecycle trajectory

This subsection proves Proposition 1 by characterizing the parameter region in which the model's lifecycle trajectory (decentralized $\rightarrow$ interior $\rightarrow$ centralized) is the unique equilibrium path.

The model's signature prediction is the empirical lifecycle trajectory: firms start fully decentralized, transition through an interior regime as they grow, and end fully centralized. For this trajectory to be the model's equilibrium prediction, three conditions on primitives and initial conditions must hold. These are not regularity conditions (which concern whether the implicit function theorem delivers sharp comparative statics) but *boundary conditions* on the theory: the parametric region where the model's empirical claims apply.

Boundary condition (B1): The firm is born fully decentralized. The firm's initial size n_1 satisfies $n_1 \leq n_D^*$, where n_D^* is the full decentralization threshold derived in Section A.5. Equivalently:

$$\textbf{(B1)} \qquad \frac{\sigma\beta v_I(v_I + b)n_1^{\zeta-1}}{k\Lambda} + \frac{F}{n_1} - \Delta_A \;\geq\; 2\delta n_1^{\alpha-1}.$$

At $n_1 \leq n_D^*$, the closed-form numerator clamps at the upper bound, $\theta_1^* = 0$, and the firm executes the high-initiative phase. Failing (B1) means the firm is already in the interior or saturated regime at birth and skips the high-initiative phase entirely.

Boundary condition (B2): The firm eventually saturates. By the terminal stage, the firm has grown enough that $n_3\Delta_A \geq F$, equivalently $n_3 \geq \bar{n} = F/\Delta_A$. The cumulative growth across the three stages must lift the firm from n_1 (in the decentralized regime) to n_3 (in the saturated regime):

$$\textbf{(B2)} \qquad n_3 \;\geq\; \bar{n} = F/\Delta_A.$$

The cumulative growth is determined by the growth equation $n_{t+1} = n_t + \phi[\Pi_O(n_t, \theta_t^*) + v_I I_{t-1}]$, evaluated along the optimal θ_t^* path. Whether (B2) holds depends on the calibrated values of ϕ, the operating profits earned at each stage, and the innovation contribution. Failing (B2) means the firm remains in the interior or decentralized regime through the terminal stage.

Boundary condition (B3): Interior passage. The intermediate stage falls strictly inside the interior regime:

$$\textbf{(B3)} \qquad n_D^* \;<\; n_2 \;<\; n_C^*,$$

where n_C^* is the full centralization threshold of Section A.4. Conditions (B1) and (B2) pin down the regimes at the endpoints of the trajectory but, with only three discrete stages, do not by themselves guarantee that the middle stage is interior: a slow-growing firm could still sit at the decentralized corner at Stage 2 ($n_2 \leq n_D^*$, jumping from $\theta^* = 0$ directly toward the terminal regime), while an explosively growing firm could overshoot the interior regime entirely ($n_2 \geq n_C^*$, jumping from $\theta^* = 0$ to $\theta^* = 1$). In a model with many short stages, (B3) would follow from (B1)–(B2) by continuity of the growth path (a path that starts below n_D^* and ends above $\bar{n}$ must cross the interval (n_D^*, n_C^*)) but with three discrete stages it must be imposed as a separate restriction on the growth rate. Economically, (B3) bounds per-stage growth from below (the firm must clear n_D^* by Stage 2) and from above (it must not clear n_C^* before Stage 3).

The trajectory implication. Under (B1)–(B3), the firm's centralization path is non-decreasing in n and strictly increasing once interior:

$$\theta_1^* = 0 \xrightarrow{\text{growth}} \theta_2^* \in (0,1) \xrightarrow{\text{growth}} \theta_3^* = 1.$$

By (B1), the firm starts at the lower clamp: $\theta_1^* = 0$. By (B3), the Stage-2 state n_2 lies strictly inside (n_D^*, n_C^*), where the policy is interior, so $\theta_2^* \in (0,1)$; and by Proposition 2(a), θ^* is strictly increasing in n on this interval. By (B2), the terminal state satisfies $n_3 \geq \bar{n}$, so the terminal-stage policy saturates: $\theta_3^* = 1$. Since firm size grows monotonically along the path ($\Pi_O > 0$ and $v_I I \geq 0$ both contribute positively to growth) and the policy is monotone in size, the sequence $(\theta_1^*, \theta_2^*, \theta_3^*) = (0, \text{interior}, 1)$ is non-decreasing and visits each regime exactly once, without backtracking. The trajectory "decentralized → interior → centralized" is therefore the unique equilibrium path. □

Verification at calibrated parameters. At the central-case calibrated parameters ($v = 1.844, b = 0.406, k = 2.758, \Delta_A = 0.00653, \delta = 9.06 \times 10^{-5}, \alpha = 2.043, \zeta = 0.200, F = 1.439, \beta = 0.5, \phi = 0.1, \pi_0 = 1.0$) and $\sigma = 1$ for services, the verification must distinguish two questions: whether the *thresholds* are well-ordered and bracket a wide interior regime (a property of the policy function), and whether a given firm's *growth path* traverses them within three stages (a property of the trajectory).

- *Thresholds and (B1).* The full decentralization threshold is $n_{D,S}^* \approx 108$. The empirical sample includes firm-size bins starting at 1–4 workers, all well below $n_{D,S}^*$, so (B1) is satisfied for all firms entering the sample as small ventures. The thresholds are well-ordered and well-separated: $n^{**} \approx 15 < n^{(a)} \approx 51 < n_{D,S}^* \approx 108 < \bar{n} \approx 220 < n_{C,S}^* \approx 617$.

- *Traversal, (B2)–(B3).* Along the calibrated equilibrium path, per-stage growth for a small services firm is modest: a firm born at $n_1 = 5$ grows by roughly 10–15%

per stage ($n_3 \approx 6.3$), and even a firm starting at the edge of the decentralized regime ($n_1 = n^*_{D,S} \approx 108$) reaches only $n_3 \approx 131$. Traversing from $n^*_D \approx 108$ to $\bar{n} \approx 220$ within the two remaining stage transitions requires average per-stage growth of $(\bar{n}/n^*_D)^{1/2} - 1 \approx 43\%$, which at the model's per-worker surplus corresponds to an effective reinvestment rate of $\phi \approx 0.42$, roughly three to four times the average rate of $\phi \approx 0.10$–0.15 estimated from the growth of single-unit continuing establishments (Section B.13). At the *average* calibrated reinvestment rate, then, (B2) and (B3) do not both hold for a firm born small: the average surviving venture remains in the decentralized regime throughout its three stages, which is itself consistent with the data, since establishments in the smallest size bins display $\theta \approx 0$ at every age. The full three-regime traversal applies to the high-growth ventures in the right tail of the growth-rate distribution, whose effective reinvestment rates are several times the average, precisely the scale-ups (VC-backed and franchise-bound ventures) that the trajectory narrative of Section 2 describes.

- *The cross-section versus the time series.* The calibration of Appendix B matches the *cross-sectional* policy function $\theta^*(n)$ across firms of heterogeneous sizes; it does not require any single firm to traverse all size bins. Large incumbents in the 350+ bins satisfy the saturation inequality $n\Delta_A \geq F$ as entities, the policy prediction $\theta^* = 1$ applies at their observed sizes, regardless of how long their own histories took to get there. The boundary conditions (B1)–(B3) are statements about the trajectory of an individual scaling venture, and their joint satisfaction requires the high-growth parameterization just described.

Alternative trajectories under failed boundary conditions. The boundary conditions delineate the parameter region where the model produces the empirically observed pattern. Outside this region, the model produces qualitatively different trajectories:

- If (B1) fails (firm born above n^*_D): the firm starts in the interior regime with $\theta^*_1 > 0$, skipping the high-initiative phase. This describes new establishments of multi-unit firms, which inherit their parent's high-θ (Result (d) of Proposition 4), or buyout-funded entrepreneurial firms that begin with substantial scale.

- If (B2) fails (firm never reaches $\bar{n}$): the firm remains in the decentralized or interior regime throughout its lifecycle. This describes lifestyle businesses, stagnant ventures, and, at the calibrated average reinvestment rate, the typical surviving small establishment, whose growth dynamics do not generate enough operating profit to traverse the interior regime within three stages.

- If (B3) fails from below ($n_2 \leq n_D^*$ despite $n_3 \geq \bar{n}$): the firm is still at the decentralized corner at the intermediate stage and compresses the organizational transition into a single jump between Stages 2 and 3, an abrupt "professionalization event" rather than a gradual transition. If (B3) fails from above ($n_2 \geq n_C^*$): the firm centralizes fully already at Stage 2, a blitzscaling path that skips the interior regime.

- If (B1)–(B3) all hold: the empirically observed trajectory obtains, with the firm passing through all three regimes in order.

The boundary conditions therefore provide a structural account of which firms the model's empirical claims apply to: scaling ventures, born small, with growth dynamics strong enough to traverse the interior regime within their lifecycle, but not so explosive as to skip it.

Relationship to the regularity condition. The two sets of conditions are distinct and neither implies the other: (B1)–(B3) place the firm on the empirically observed trajectory, while the regularity condition of Appendix A.3 certifies that the interior comparative statics are well-founded. They connect only through (B1), under which the firm's interior portion begins at n_D^*, making $n_D^* > n^{**}$ the relevant (and weaker) regularity check. Both threshold orderings hold at calibrated parameters; the traversal component of (B2)–(B3) additionally requires the high-growth parameterization quantified above.

A.7 Proof of Proposition 4

The following results use only the properties $\theta^*(n)$ increasing (Result (a)) and $\theta^*_S < \theta^*_M$ (Result (b)), which are established in Section A.3.

Result (a): Hump-shaped services innovation. **Phase 1 (small n, $\theta^* = 0$):** $I(n) = n^\zeta(v_I + b)/k$, strictly increasing since $\zeta > 0$.

Phase 2 (large n, θ^* interior): Differentiate $I(n) = n^\zeta(1 - \theta^*(n))(v_I + b)/k$ by the product rule:

$$\frac{dI}{dn} = \underbrace{\zeta n^{\zeta-1}(1-\theta^*)(v_I + b)/k}_{\text{scale effect} > 0} - \underbrace{n^\zeta(v_I + b)/k \cdot d\theta^*/dn}_{\text{centralization effect} > 0}.$$

The scale effect weakens with n ($\zeta n^{\zeta-1}$ is decreasing since $\zeta - 1 < 0$, and $(1-\theta^*)$ is decreasing by Proposition 2(a)), while the centralization effect is positive and bounded away from zero once θ^* becomes interior (n^ζ is increasing and $d\theta^*/dn > 0$). This decomposition shows the two forces pull in opposite directions and supplies the intuition for an interior turning point, but the existence of the hump does not rest on the relative monotonicity of the two terms. The rigorous argument is by continuity: $I(n)$ is continuous, strictly increasing on the decentralized segment (Phase 1), and tends to zero as $n \to \infty$ (next paragraph). A continuous function that is increasing on an initial interval and tends to zero must attain a maximum and ultimately decline; the services innovation profile therefore rises to a peak near the decentralization threshold n^*_D and falls toward zero for larger n. (Strict monotonicity of the decline at every interior size is not needed for the result and is not claimed in general; at the calibrated parameters it does hold ($I(n)$ peaks exactly at n^*_D and decreases strictly on (n^*_D, n^*_C), reaching $I = 0$ at n^*_C) as verified in the replication notebook.)

Decline below reference. For $n \geq n^*_C$ the closed form clamps at $\theta^* = 1$, so $I(n) = 0$ exactly. More generally, as $n \to \infty$, $\theta^* \to 1$: in the closed form for $1 - \theta^*$ the denominator $2\delta n^{\alpha-1}$ diverges (since $\alpha > 1$) while the numerator stays bounded (the innovation force $\propto n^{\zeta-1}$ and the overhead F/n both vanish, leaving $-\Delta_A$). Hence $I(n) \to 0 < I(n_{\text{ref}})$:

innovation eventually falls below the small-firm reference level. $\square$

Result (b): Monotonic manufacturing innovation. With $\sigma = 0$, $I(n) = n^{\zeta}(v_I + b)/k$, independent of θ. Since $\zeta > 0$, $dI/dn > 0$. $\square$

Result (c): Growth despite innovation collapse. For large n where $\theta^* \approx 1$:

$$\Pi_O(n, 1) = n(\pi_0 + \Delta_A) - F,$$

which is linear in n with slope $\pi_0 + \Delta_A > 0$. The coordination cost vanishes ($(1-1)^2 = 0$) and per-worker surplus rises from π_0 to $\pi_0 + \Delta_A$. $\square$

Result (d): Flat establishment growth. An establishment inherits $\theta_{\text{parent}} \approx 1$ from the parent. Initiative $e^* = (1 - \theta_{\text{parent}})(v_I + b)/k \approx 0$, so $I \approx 0$ regardless of establishment size. A de novo startup at the same size would choose $\theta^*(n_e) \ll 1$, preserving initiative. The establishment, constrained to the parent's organizational form, forgoes innovation-driven growth. $\square$

Irreversibility. $\Pi_O(n, \theta^*(n)) + v_I \cdot I_{t-1} > 0$ for all relevant n (since $\Pi_O > 0$ at calibrated parameters and $v_I \cdot I_{t-1} \geq 0$), so $n_{t+1} > n_t$. Since $\theta^*(n)$ is increasing (Proposition 2(a)), θ^* is non-decreasing along the path. Since $I(n)$ is decreasing post-peak (Result (a)), innovation is non-increasing. $\square$

A.8 Project-selection microfoundation for the alignment gap and private benefit

This closing subsection supplies a microfoundation for the reduced-form primitives v_I, b, and Δ_A used from Section A.1 onward. It does not prove a proposition; it makes the project distribution explicit so that the restriction $\Delta_A \leq b$, used as an identifying bound in the *sector-specific calibration (Appendix B.14), becomes visible.*

The project structure. Each worker faces a set of available projects $\mathcal{J}$, where project $j \in \mathcal{J}$ is characterized by a pair (v_j, b_j): the firm value $v_j \geq 0$ from a successful innovation on project j, and the worker's private benefit $b_j \geq 0$ from pursuing it. The worker pursues exactly one project. Under decentralization ($\theta = 0$) the worker chooses it; under centralization ($\theta = 1$) the firm chooses it; under intermediate governance $\theta \in (0, 1)$ the realized project shifts continuously between the two, in the spirit of Aghion and Tirole (1997)'s real authority. The worker, internalizing both v_j (via job retention) and b_j (private), chooses $j^W \in \arg\max_j [v_j + b_j]$; the firm, capturing only v_j, prefers $j^F \in \arg\max_j v_j$. Write $v^W \equiv v_{j^W}$, $v^F \equiv v_{j^F}$, and $b \equiv b_{j^W}$, the private benefit on the project actually pursued under decentralization. The alignment gap is $\Delta_A \equiv v^F - v^W \geq 0$.

The restriction $\Delta_A \leq b$. The worker's choice reveals $v^W + b \geq v_j + b_j$ for all j; applied to $j = j^F$, this gives $v^F - v^W \leq b - b_{j^F}$, so

$$\Delta_A \;=\; v^F - v^W \;\leq\; b - b_{j^F} \;\leq\; b, \tag{36}$$

since $b_{j^F} \geq 0$. The weak bound $\Delta_A \leq b$ always holds by revealed preference; the strict version $\Delta_A < b$ is what a non-degenerate alignment problem requires. The bound is tight when $b_{j^F} = 0$ (the firm-preferred project carries no private benefit) and $\Delta_A = 0$ when $v_{j^W} = v_{j^F}$ (the worker's and firm's preferred projects coincide in firm value, so the principal–agent problem vanishes and centralization buys no alignment gain). The generic case is $0 < \Delta_A < b$: the worker pursues $j^W \neq j^F$, the firm captures only v^W under decentralization, and centralization buys a strictly positive alignment gain. The case $\Delta_A > b$ is ruled out by revealed preference, since the worker would then prefer j^F.

A worked example. Let $\mathcal{J} = \{A, B\}$ with $(v_A, b_A) = (V, 0)$ and $(v_B, b_B) = (V - x, b)$, where $V, x, b > 0$: project A is commercially valuable with no private benefit, project B slightly less valuable but with private benefit b. The firm prefers A, so $v^F = V$. The worker prefers B iff $V - x + b > V$, i.e. $b > x$; when this holds, $v^W = V - x$, the realized private benefit is b, and $\Delta_A = x < b$, consistent with (36). The condition $b > x$ for the worker to choose B is exactly $b > \Delta_A$.

Innovation value and per-worker surplus. The per-innovation payoff and the operating surplus $v(\theta) = v^W + \Delta_A\theta$ are *distinct cash flows* sharing a common project-value structure: raising θ redirects effort from the worker-preferred to the firm-preferred project, lifting both. To avoid double-counting, the reduced form books the θ-dependent part $\Delta_A\theta$ *once*, inside operating profit $(\pi_0 + \Delta_A\theta)$, and holds the separate innovation stream $v_I \cdot I$ at the constant level $v_I = v^W$ (its value at $\theta = 0$). Fixing $v_I = v^W$ rather than $v^W + \Delta_A\theta$ is the conservative choice: centralization is never credited with raising the value of each innovation, so the innovation channel acts only *against* centralization, through the lost *number* of innovations. The schedule $v_I(\theta) = v^W + \Delta_A\theta$ satisfies $v_I(\theta) \leq v^F$ with equality at $\theta = 1$. In the operating-profit expression, π_0 is the per-worker firm value at $\theta = 0$ net of wages ($\pi_0 = v^W - w$); since a stage spans roughly seven years (Section 3.5), π_0 is the seven-year net operating surplus per worker, the natural scale on which the firm-value parameters $(v_I, \Delta_A, b, F, \delta)$ are measured, with $\pi_0 = 1$ the calibration normalization.

Implications for calibration. First, $\Delta_A \leq b$ is a valid upper bound: imposing $\Delta_A < (b/v)\, v_I$ enforces it under the normalization ($v_I = 1$, $b/v = 0.22$ gives $\Delta_A < 0.22$). In the central-case calibration the data place $\Delta_A \approx 0.007$, far below the bound, so it does not bind; in the sector-specific extension (Appendix B.14) the bound binds for some seed/specification combinations. Second, the data favoring $\Delta_A \ll b$ rather than $\Delta_A \approx b$ imply that the alignment problem is real but quantitatively small, so centralization in this calibration is primarily about coordination-cost relief rather than alignment redirection. Third, π_0 and v_I

are firm values in the same units, and scale invariance means only their ratio matters; fixing $\pi_0 = 1$ identifies v_I in units of one stage of one worker's operating surplus.

B Calibration

This appendix describes the quantitative exercise used to assess whether the model's single mechanism, replication killing initiative, can simultaneously generate the qualitative patterns observed in the data on services and manufacturing centralization and innovation. The exercise is best characterized as shape-matching, not structural estimation: it tests whether one set of structural parameters can simultaneously produce the correct qualitative profiles across four panels (services innovation, services centralization, manufacturing innovation, manufacturing centralization).

Roadmap of the calibration appendix. The appendix proceeds in three stages. *First, the central-case cross-section calibration* (Sections B.1 through "Interpreting the fit"): model equations, data targets, parameters, matching strategy, estimation algorithm, replication pseudocode, an analysis of the under-identification of b and the literature-anchored resolution (§B.7), the central-case estimates and fit, sensitivity to the b/v normalization (§B.10), verification of the boundary and regularity conditions at the calibrated values (§B.11), and an interpretive coda on what the four panel correlations measure. *Second, two extensions of the central case* (§B.13–§B.14): adding lifecycle growth-rate moments to free the reinvestment rate ϕ, and allowing sector-specific $(\pi_0, \phi, \zeta, \Delta_A)$ to differentiate services and manufacturing. *Third, a synthesis* (§B.16): a complete taxonomy of the restrictions imposed on the parameter space (mathematical, microfoundation, plausibility, normalization, externally calibrated), and a characterization of the residual under-identification that all three calibration variants share.

B.1 Model equations used in the calibration

For a firm of size n, the saturated-regime closed forms of Proposition 2 give optimal centralization $\theta^*(n)$ and innovation $I(n)$. The calibration uses these closed forms throughout, treating the growth multiplier Λ as the parameter

$$\Lambda \;=\; 1 + \beta\phi(\pi_0 + \Delta_A), \tag{37}$$

which is the large-n limit derived in Section A.4. These are the *penultimate-stage* policies, for which the closed form is exact under terminal-stage saturation; the Stage-1 policy differs by the small correction factor Λ/Λ_1 on the innovation force, which shifts θ^* by at most 0.008 across the empirical size bins at the calibrated parameters (Section A.4), an order of magnitude below the cross-bin variation the calibration fits. The empirical validity of this approximation is verified at the end of this appendix.

Services.

$$\theta^*_S(n) = \left[1 - \frac{\beta v(v+b)\, n^{\zeta-1}/(k\Lambda) - \Delta_A + F/n}{2\delta n^{\alpha-1}}\right]_0^1, \tag{38}$$

$$I_S(n) = \frac{n^\zeta(1-\theta^*_S(n))(v+b)}{k}, \tag{39}$$

where $[\cdot]_0^1$ denotes clamping to $[0, 1]$.

Manufacturing.

$$\theta^*_M(n) = \left[1 - \frac{F/n - \Delta_A}{2\delta n^{\alpha-1}}\right]_0^1, \tag{40}$$

$$I_M(n) = \frac{n^\zeta(v+b)}{k}. \tag{41}$$

Manufacturing R&D is independent of θ because $\sigma = 0$: centralizing operations has no effect on innovation.

In the calibration, the parameter v corresponds to v_I in the body of the paper; the subscript is dropped here for brevity.

B.2 Data targets

The empirical targets come from Kueng et al. (2017), who estimate size-bin regressions on a representative Canadian firm-level panel. The data consist of regression coefficients for seven size bins (5–9, 15–19, 50–99, 150–199, 300–399, 500–749, 1000+ employees), measured relative to an omitted 1–4 employee category, in four panels:

1. **Services innovation:** LPM coefficients for process innovation. Hump-shaped, peaking at 50–99 employees and declining sharply thereafter.
2. **Services centralization:** Coefficients on a standardized centralization index. Rising with size through the 500–749 bin; anomalous drop at 1000+.
3. **Manufacturing innovation:** LPM coefficients for product innovation. Monotonically increasing.
4. **Manufacturing centralization:** Coefficients on the same centralization index. Rising with size; anomalous drop at 1000+.

Each size bin is represented by its midpoint: $n \in \{7, 17, 75, 175, 350, 625, 1500\}$. The omitted category midpoint is $n_{\text{ref}} = 2.5$. The 1000+ bin is excluded from the centralization panels due to its anomalous drop (likely reflecting compositional effects in very large firms), leaving 6 centralization bins and 7 innovation bins per sector, for a total of 26 moments. Each moment is a size-bin regression coefficient, equivalently a conditional bin-level average relative to the omitted 1–4 employee category, not an individual firm observation.

A note on the choice of innovation measure: the targets use *process* innovation for services and *product* innovation for manufacturing. This asymmetry mirrors the model's mechanism. In services, replication standardizes the delivery process itself, so process innovation is the margin on which centralization most directly kills initiative; in manufacturing, the separable R&D function's headline output is new products, while production processes are the object of replication. As documented in Section 2, the decline in services innovation is parallel across all three innovation types (product, process, and incremental) in the WES data, so the qualitative target patterns, the services hump and the manufacturing monotone increase, do not depend on which series is used; the chosen series are the sharpest representatives of each sector's pattern.

Empirical moments. The exact values used in the calibration are:

Bin midpoint n	7	17	75	175	350	625	1500
Services innovation	0.03	0.17	0.20	−0.20	−0.25	−0.35	−0.20
Services centralization	−0.15	0.35	0.10	1.00	1.80	2.90	(
Mfg. innovation	−0.05	0.00	0.05	0.08	0.10	0.12	0.15
Mfg. centralization	0.00	0.00	0.00	1.10	1.00	1.70	)

B.3 Parameters

The model has 11 structural parameters. Three are calibrated to external values; eight are estimated.

Calibrated parameters.

- $\beta = 0.5$: discount factor over a stage of approximately 7 years, corresponding to an annual discount rate of $(1/0.5)^{1/7} - 1 \approx 10\%$.
- $\pi_0 = 1.0$: normalization of baseline per-worker surplus.
- $\phi = 0.1$: reinvestment rate, scaling operating profits into headcount growth.

Estimated parameters. v (firm innovation payoff), b (worker private benefit), k (effort cost), Δ_A (alignment gain), δ (coordination cost scale), α (coordination cost curvature), ζ (innovation returns to scale), F (overhead cost of centralization).

B.4 Matching strategy

Model predictions. For a candidate parameter vector $(v, b, k, \Delta_A, \delta, \alpha, \zeta, F)$, the model produces $\theta^*_\sigma(n)$ and $I_\sigma(n)$ at each size bin n and for each sector $\sigma \in \{S, M\}$. Innovation predictions are differenced relative to the reference category: $\Delta I_\sigma(n) \equiv I_\sigma(n) - I_\sigma(n_{\text{ref}})$. Centralization predictions are differenced relative to the smallest bin: $\Delta\theta_\sigma(n) \equiv \theta^*_\sigma(n) - \theta^*_\sigma(n_1)$, where $n_1 = 7$.

Objective function. The objective maximizes the weighted average Pearson correlation between model and data across the four panels:

$$\min_{v,b,k,\Delta_A,\delta,\alpha,\zeta,F} \; -\big(0.25\,\rho_{I,S} + 0.25\,\rho_{\theta,S} + 0.15\,\rho_{I,M} + 0.15\,\rho_{\theta,M}\big) - 0.10\,\bar{d}_\theta + \text{penalties}, \tag{42}$$

where:

- $\rho_{I,\sigma}$ is the Pearson correlation between data innovation coefficients and $\Delta I_\sigma(n)$ predictions (7 bins each).
- $\rho_{\theta,\sigma}$ is the Pearson correlation for centralization (6 bins, excluding 1000+).
- $\bar{d}_\theta = \frac{1}{7}\sum_i |\theta^*_S(n_i) - \theta^*_M(n_i)|$ rewards sectoral divergence, ensuring the model produces visibly different services and manufacturing centralization patterns.
- Penalties: a value of 0.3 is added to the objective for each adjacent pair where $\theta^*(n_i) > \theta^*(n_{i+1}) + 0.05$ (services or manufacturing), discouraging non-monotonic centralization profiles. This is a soft enforcement of the monotonicity in Proposition 2(a).

Services panels receive higher weight (0.25 vs. 0.15) because the services innovation hump is the model's most distinctive and informative prediction; manufacturing innovation is monotonic by construction.

Constraint. The effort cost must exceed the combined return to ensure interior worker effort $e^* = (v + b)/k < 1$: $k \geq v + b$. Parameter vectors violating this constraint receive a large objective penalty (10^6).

Scale factors for visualization. The data and model are in different units (LPM coefficients vs. model units). For displaying the fit graphically, each panel p receives an OLS scale factor

$$s_p^* = \frac{\sum_i d_{p,i} \cdot m_{p,i}}{\sum_i m_{p,i}^2}, \tag{43}$$

where $d_{p,i}$ is the data coefficient and $m_{p,i}$ is the model prediction for bin i. The scale factors are not free parameters; they are determined by OLS given the model predictions. They are reported for transparency but enter the objective only through the correlation, which is scale-invariant.

B.5 Estimation algorithm

The objective is minimized using differential evolution (Storn and Price, 1997), a global optimization method well suited to non-convex landscapes with multiple local optima. The algorithm maintains a population of candidate parameter vectors and iteratively improves them through mutation and crossover.

Parameter bounds.

Parameter	Lower bound	Upper bound
v	0.1	2.0
b	0.05	1.5
k	0.5	10.0
Δ_A	0.005	0.5
$\log \delta$	-12	-5
α	2.01	3.5
ζ	0.2	0.7
F	0.5	50

The bound $\alpha \geq 2.01$ enforces the theoretical requirement $\alpha > 2$ for coordination costs to grow faster than n^2 (Section A.3); δ is estimated in logs to facilitate search over a wide range of magnitudes; the bound $\zeta \leq 0.7$ ensures returns to scale in innovation are interior; the upper bound on F is set generously since F scales the overhead cost relative to baseline per-worker surplus $\pi_0 = 1$.

Algorithm settings.

- Population size: 40 candidate vectors.
- Maximum iterations: 1500.
- Convergence tolerance: 10^{-12} (spread of objective values across the population).
- Random seed: 42 (for reproducibility).

B.6 Replication procedure (pseudocode)

The following pseudocode describes the calibration procedure independently of any specific programming language. The actual Python implementation is available in the supplementary materials.

Inputs: Empirical moments $d_{p,i}$ for panels $p \in \{I{:}S, \theta{:}S, I{:}M, \theta{:}M\}$ at size bins $\{n_i\} = \{7, 17, 75, 175, 350, 625, 1500\}$; reference size $n_{\text{ref}} = 2.5$; calibrated $\beta = 0.5$, $\pi_0 = 1.0$, $\phi = 0.1$.

Step 1: Define a function $\textsc{ModelPredictions}(v, b, k, \Delta_A, \delta, \alpha, \zeta, F, \sigma, n)$ that returns the model's predicted (θ, I) at firm size n in sector σ:

Set $r \leftarrow v + b$ and $\Lambda \leftarrow 1 + \beta\phi(\pi_0 + \Delta_A)$.

If $\sigma = 1$ (services):

Compute innov_force $\leftarrow \beta \cdot v \cdot r \cdot n^{\zeta}/(k \cdot \Lambda)$.

Compute num $\leftarrow$ innov_force $-(n \cdot \Delta_A - F)$ and denom $\leftarrow 2 \cdot \delta \cdot n^{\alpha}$.

Set $\theta \leftarrow \text{clip}(1 - \text{num}/\text{denom},\ 0,\ 1)$ and $I \leftarrow n^{\zeta} \cdot (1 - \theta) \cdot r/k$.

Else if $\sigma = 0$ (manufacturing):

If $n \cdot \Delta_A \geq F$ **then** $\theta \leftarrow 1$.

Else $\theta \leftarrow \text{clip}(1 - (F - n \cdot \Delta_A)/(2 \cdot \delta \cdot n^{\alpha}),\ 0,\ 1)$.

Set $I \leftarrow n^{\zeta} \cdot r/k$ (innovation independent of θ).

Return (θ, I).

Step 2: Define a function $\textsc{Objective}(v, b, k, \Delta_A, \log\delta, \alpha, \zeta, F)$:

Set $\delta \leftarrow \exp(\log\delta)$.

If $k < v + b$, return 10^6 (constraint $e^* < 1$ violated).

For each sector $\sigma \in \{1, 0\}$:

Compute $(\theta_\sigma(n_i), I_\sigma(n_i))$ at each bin n_i via $\textsc{ModelPredictions}$.

Compute reference $I_\sigma(n_{\text{ref}})$ at $n_{\text{ref}} = 2.5$.

Form $\Delta I_\sigma(n_i) = I_\sigma(n_i) - I_\sigma(n_{\text{ref}})$ for $i = 1, \ldots, 7$.

Form $\Delta\theta_\sigma(n_i) = \theta_\sigma(n_i) - \theta_\sigma(n_1)$ for $i = 1, \ldots, 6$.

Compute four Pearson correlations vs. data: $\rho_{I,S}, \rho_{\theta,S}, \rho_{I,M}, \rho_{\theta,M}$.

Compute $\bar{d}_\theta = \frac{1}{7}\sum_i |\theta_S(n_i) - \theta_M(n_i)|$.

Set score $\leftarrow -(0.25\rho_{I,S} + 0.25\rho_{\theta,S} + 0.15\rho_{I,M} + 0.15\rho_{\theta,M}) - 0.10\,\bar{d}_\theta$.

For $i = 1, \ldots, 5$:

If $\theta_S(n_i) > \theta_S(n_{i+1}) + 0.05$, set score $\leftarrow$ score $+0.3$.

If $\theta_M(n_i) > \theta_M(n_{i+1}) + 0.05$, set score $\leftarrow$ score $+0.3$.

Return score.

Step 3: Run differential evolution on OBJECTIVE over the parameter bounds listed in the previous subsection, with population size 40, maximum iterations 1500, convergence tolerance 10^{-12}, and random seed 42 for reproducibility.

Step 4: Output the estimated parameter vector $(\hat{v}, \hat{b}, \hat{k}, \hat{\Delta}_A, \hat{\delta}, \hat{\alpha}, \hat{\zeta}, \hat{F})$ achieving the minimum objective value.

For users implementing this in a language without a built-in differential evolution routine, any global optimization method that explores the parameter bounds without requiring derivatives (e.g., simulated annealing, particle swarm, CMA-ES) will produce qualitatively similar results. Local optimizers (Nelder–Mead, BFGS) are not recommended because the objective is non-convex.

B.7 Under-identification of the private benefit and external anchoring

Before reporting estimates, I document a structural identification issue and the literature-anchored choice used to resolve it.

The under-identification of the worker's private benefit. Multistart calibration with multiple random seeds reveals that the worker's private benefit parameter b is poorly identified by the data: across six seeds achieving essentially the same objective value (-0.76305 to four decimal places), b takes values ranging from 0.60 to 1.30 (relative SD of approximately 21%), while other parameters move concomitantly; the exact range depends on optimizer settings and library version (in the replication environment, for example, the full-settings seed-42 run lands at $b \approx 1.4$ and lighter-settings seeds span roughly $[0.5, 1.1]$), while the invariance of the objective value and of the identified combinations does not. A moment-by-moment sensitivity analysis confirms the diagnosis: perturbing any single empirical moment by +0.05 causes b to move by anywhere from −87% to +215% across the 26 moments, with a relative standard deviation of 79%. By comparison, the other six free parameters move by at most 30% across the same perturbations (with α and ζ effectively pinned at their bounds). The data simply do not contain enough information to identify b separately.

Structurally, this is because the model's predictions depend on b only through the worker effort ratio $e^* = (v + b)/k$ and the worker return $r = v + b$, both of which are jointly determined with v and k. The model is identified up to the combinations

$$\bar{n} = F/\Delta_A, \qquad \delta/\Delta_A, \qquad v(v+b)/(k\Delta_A), \qquad \alpha, \qquad \zeta. \tag{44}$$

These five combinations are recovered consistently across multistart seeds; the eight individual parameters are not.

Resolution: fix the ratio of private benefit to innovation value at a literature-anchored value. To reduce under-identification, I fix the ratio b/v exogenously. The parameter b/v has a clean economic interpretation: it measures the worker's private benefit from initiative relative to the firm's payoff from innovation, equivalently the share of the worker–firm return $r = v + b$ accruing to the worker. The personnel economics literature provides three credible anchors for this ratio in frontline production and service contexts.

Central case, $b/v \approx 0.22$ (Lazear, 2000). Lazear (2000) documents the productivity effect of switching from hourly wages to piece rates at Safelite Glass Corporation. Output per worker rose by approximately 44% in total, which Lazear decomposes into two channels: approximately 22 percentage points from within-worker incentive response (workers exerting more effort under piece rates) and approximately 22 percentage points from selection effects (more able workers retained, less able workers leaving). For a calibration that takes the firm's workforce as fixed, only the within-worker incentive channel is informative: it identifies how much output workers withhold when their pay does not depend on output.

The mapping to b/v requires a stance on what a flat-wage worker internalizes, and I state it explicitly. In the model, total worker initiative is proportional to the return the worker internalizes, out of the total stake $r = v + b$ attached to her initiative: v is the firm-value component and b the private (intrinsic-interest, reputation, satisfaction) component. The maintained assumption is that a flat-wage worker in an ongoing employment relationship internalizes the firm-value component v (through job retention, monitoring, and career concerns, she works to produce the output her employer values) but does not enjoy the private component b, which attaches to discretionary, initiative-intensive activity that scripted flat-wage work does not deliver. Full incentivization (piece rates, or in the model, decentralization with residual-claimant initiative) adds the private component, raising the internalized return from v to $v+b$. The within-worker effort response to full incentives is then $(v+b)/v = 1.22$, equivalently $e^*_{\text{flat}}/e^*_{\text{piece}} = v/(v+b) = 1/1.22 \approx 0.82$, giving $b/v = (1-0.82)/0.82 \approx 0.22$: workers, when not directly incentivized, withhold approximately 18% of their potential effort.

Two remarks on this mapping. First, the polar opposite reading, a flat-wage worker internalizes *only* the private benefit b and none of v, would invert the ratio and imply $b/v = 0.22/0.045 \approx 4.5$, i.e., a private benefit several times the firm value. That reading sits poorly with the institutional setting (Safelite installers were monitored employees whose output was observable, not volunteers working for intrinsic enjoyment), and I do not adopt it; the internalization-of-v reading is also the conservative one, since it makes b small relative to v. Second, and more importantly, the calibration's substantive conclusions do not hinge on the stance: as established above and in Section B.10, the data identify only combinations in which b enters jointly with v (through $r = v + b$ and the innovation force $v(v+b)/k$), the remaining parameters re-equilibrate under any fixed b/v, and the derived thresholds, fit, and lifecycle predictions are essentially invariant across normalizations. The choice of b/v disciplines the reported parameter vector, not the model's predictions.

Lower bound, $b/v \approx 0.06$ (Ichniowski, Shaw, and Prennushi, 1997). Ichniowski et al.

(1997) estimate the productivity effects of bundled human-resource-management practices on uptime in steel finishing lines. Their most conservative econometric specification, with detailed technology controls and fixed effects, finds that the most innovative HRM system (incentive pay, problem-solving teams, employment security, flexible job assignment, extensive screening, training, and labor–management communication) raises line uptime by approximately 6.7 percentage points relative to traditional practices (hourly pay, narrow job definitions, close supervision).

Mapping this to b/v is less clean because the HRM bundle conflates incentive pay with several other channels. But interpreted as an upper bound on the residual agency wedge in well-monitored, technologically paced production processes, the 6.7-point gap implies $e^*_{\text{trad}} \approx 1/1.067 \approx 0.94$, giving $v/(v+b) \approx 0.94$ and $b/v \approx 0.06$. This is the soft floor: in production environments where monitoring is strong and complementary practices align worker behavior with firm objectives, the residual private-benefit margin is small.

Upper bound, $b/v \approx 0.33$ *(Bandiera, Barankay, and Rasul, 2005).* Bandiera et al. (2005) compare worker productivity under two distinct incentive regimes: relative-incentive pay (where individual pay depends on individual output relative to the field average) and pure piece rates (where pay depends only on individual output). Using within-worker variation as the same fruit pickers experience both regimes over the course of a season, they find that average productivity is at least 50% higher under piece rates than under relative incentives, attributing the gap to workers internalizing the negative externality their effort imposes on co-workers.

The 50% gap is between two distorted regimes, not between aligned and unaligned. But interpreted as an upper bound on the within-worker effort-response margin in settings with social/peer externalities, the gap implies $e^*_{\text{rel}} \approx 1/1.5 = 0.67$, so $v/(v+b) \approx 0.67$ and $b/v \approx 0.5$. This is plausibly an overstatement of b because the relative-pay regime is unusually distorted by the social-externality channel; a more conservative reading of the same evidence yields $b/v \approx 0.33$ if one assumes that a fraction of the gap reflects observable peer comparison rather than intrinsic private benefit. The upper bound is therefore in the range $b/v \in [0.33, 0.50]$, with 0.33 as the conservative reading.

Realistic range. Taken together, the three anchors place $b/v \in [0.06, 0.33]$ for frontline production and service workers, with 0.22 as the modal calibration consistent with Lazear's within-worker incentive channel. The calibration reports results for the central case $b/v = 0.22$ in the main table below, with sensitivity analyses across the full range in Section B.10.

B.8 Estimated parameters at central case (Lazear anchor)

The calibration with $b = 0.22v$ fixed, leaving seven free parameters $(v, k, \Delta_A, \delta, \alpha, \zeta, F)$, yields:

Parameter	Value	Interpretation
v	1.844	Firm's innovation payoff
b	0.406	Worker's private benefit ($= 0.22 \cdot v$)
k	2.758	Effort cost
Δ_A	0.00653	Alignment gain (small relative to $\pi_0 = 1$)
δ	9.06×10^{-5}	Coordination cost scale
α	2.043	Coordination cost curvature
ζ	0.200	Innovation returns to scale (at lower bound)
F	1.439	Overhead cost of centralization

These values are means across six multistart seeds, with one disclosed adjustment: the multistart mean of Δ_A is 0.00648, and the reported value $0.00653 = F/220.4$ (a departure of under one percent, well within Δ_A's multistart dispersion of 9.2%) is chosen so that the reported vector is internally consistent with the verification table of Section B.11, all of whose thresholds and bin-level predictions are computed at exactly the reported values. The relative standard deviations are reported in Section B.10 below. Derived quantities: $r = v + b = 2.25$, $\Lambda = 1 + \beta\phi(\pi_0 + \Delta_A) = 1.0503$, $\bar{n} = F/\Delta_A \approx 220$, $e^* = (v + b)/k \approx 0.81$ (well below the interiority bound of 1).

Pinned parameters. Two estimated parameters reach the lower bounds of their search range across all multistart seeds:

- α = 2.01–2.05, against the lower bound of 2.01. The theory requires $\alpha > 2$ for the regularity condition and for super-quadratic coordination costs in firm size. The calibration's preference for the lowest admissible value indicates that the data favor coordination costs that grow only slightly faster than n^2. The empirical hump-shaped services innovation pattern requires steeply rising coordination costs (so that centralization becomes attractive at intermediate sizes), but not steeper than n^2.

- $\zeta = 0.200$, exactly at the lower bound of 0.2. Lower values of ζ correspond to more severe diminishing returns to size in innovation. The calibration's preference for the lowest admissible value indicates the data favor steeply diminishing returns, consistent with the empirical observation that services innovation collapses past 150 employees. The lower bound at 0.2 is not theoretically required but reflects a deliberate restriction to keep returns-to-scale in innovation within a plausible range.

Both pinnings should be read as informative: they describe where the data want the parameters to go, and the model fit reported below is achieved under these binding constraints.

B.9 Fit

Panel	ρ	Scale s^*	Scaled RMSE	Moments
Services innovation	0.944	0.277	0.074	7
Services centralization	0.929	2.249	0.396	6
Manufacturing innovation	0.976	0.058	0.026	7
Manufacturing centralization	0.936	1.265	0.243	6
Mean correlation	0.946			26

The scaled RMSE is the root-mean-squared residual after applying the OLS scale factor s^* to the model predictions, i.e., RMSE $= \sqrt{(1/N)\sum_i (d_i - s_p^* m_i)^2}$. It is reported as a magnitude-aware complement to the (scale-invariant) correlation: $\rho \approx 0.94$ implies $\sqrt{1-\rho^2} \approx 0.34$, and the scaled RMSE in each panel is correspondingly about a third of the data standard deviation. The bottom row is the unweighted mean of the four panel correlations; the moment-weighted mean (weights $7/26, 6/26, 7/26, 6/26$) is 0.947, and the objective-function weighting $(0.25, 0.25, 0.15, 0.15)$ gives 0.944, the headline value is insensitive to the weighting convention. The table is computed at the reported central-case parameter vector, so that the fit, the parameter table, and the verification of Section B.11 all refer to a single vector. At unconstrained (free-b) estimates, the correlations are the same to within ± 0.004, but the innovation-panel scale factors vary along the under-identified ridge in proportion to $k/(v+b) = 1/e^*$; the previously reported $s^*_{I,S} = 0.23$ corresponds to a seed-42 free-b solution at the multistart settings, with worker effort $e^* \approx 0.98$ rather than 0.82, which rescales the model's innovation units without affecting their shape.

B.10 Sensitivity to the private-benefit normalization

To assess how the central-case calibration depends on the chosen value of b/v, I re-estimate the model at the lower bound ($b/v = 0.06$, Ichniowski–Shaw–Prennushi anchor), the central case ($b/v = 0.22$, Lazear anchor), and the upper bound ($b/v = 0.33$, Bandiera–Barankay–Rasul anchor). For each value, I (i) run multistart with six different random seeds to test optimizer-induced parameter dispersion, and (ii) perturb each of the 26 empirical moments by $+0.05$ and re-estimate, measuring how each parameter responds.

Fit is essentially identical across normalization choices. All three normalizations achieve a mean correlation of $\rho \approx 0.947$ across the four panels, with objective function values differing only in the fifth decimal place. The choice of b/v does not affect the model's ability to fit the empirical moment patterns; it affects only which structural-parameter representation of that fit is selected.

Parameter stability under multistart. Table 1 reports the relative standard deviation of each parameter across six multistart seeds (all six seeds are retained at every ratio; at the multistart settings, every seed converges to the same objective value to within 10^{-4}). Stability deteriorates as b/v increases, particularly for k and the level parameters Δ_A, δ, F:

Table 1: Multistart stability across six seeds (relative SD of estimates). The b row equals the v row by construction, since $b = (b/v)\,v$ is imposed.

Parameter	$b/v = 0.06$	$b/v = 0.22$	$b/v = 0.33$
v	4.8%	5.7%	9.8%
b	4.8%	5.7%	9.8%
k	5.4%	11.2%	27.1%
Δ_A	9.6%	9.2%	15.0%
δ	10.4%	25.6%	14.7%
α	0.2%	2.5%	0.3%
ζ	0.0%	0.0%	0.0%
F	9.1%	10.5%	21.5%

The lower bound case ($b/v = 0.06$) produces the tightest multistart estimates: most free parameters have relative SD below 10%. The upper bound case ($b/v = 0.33$) shows substantial parameter dispersion (10–27% for the level parameters). The central case ($b/v = 0.22$) sits between, with moderate dispersion for most parameters and somewhat larger spread for δ.

Moment-by-moment sensitivity. I perturb each of the 26 empirical moments by +0.05 in its raw coefficient units and re-estimate the calibration, then measure how each parameter changes. A parameter is well identified by the data if its estimate is stable across all moment perturbations; it is poorly identified if perturbing some moment moves it substantially. Table 2 reports the relative standard deviation of each parameter across the 26 perturbations.

Three patterns are clear:

(1) α and ζ are robustly identified across all three normalizations. Their relative SDs are 2.5% or less in every case. The data unambiguously want α at its lower bound (slightly above 2) and ζ exactly at 0.2.

Table 2: Moment-perturbation sensitivity (relative SD of estimate across 26 perturbations).

Parameter	$b/v = 0.06$	$b/v = 0.22$	$b/v = 0.33$
v	7.4%	6.7%	6.9%
b	7.4%	6.7%	6.9%
k	16.2%	14.1%	10.0%
Δ_A	11.1%	17.4%	20.3%
δ	28.9%	45.7%	37.9%
α	2.4%	2.1%	2.0%
ζ	0.1%	0.4%	0.3%
F	21.1%	30.7%	30.5%

(2) v and b are tightly co-identified once b/v is fixed. With $b = (b/v) \cdot v$ imposed exogenously, v moves by only 7% across perturbations under any b/v value, and b moves identically. By contrast, in the unconstrained calibration (Section B.7 above), b moves by 79% across the same perturbations. Fixing b/v resolves the under-identification of b.

(3) The level parameters Δ_A, δ, F are less stable across moment perturbations, especially at higher b/v. At $b/v = 0.06$, the lowest stability is 28.9% (for δ); at $b/v = 0.22$ this rises to 45.7%; at $b/v = 0.33$ it is 37.9%. The level parameters move together in ways that preserve the identified combinations (44); they are not jointly identified by the data, only their ratios are.

Identified combinations are stable. The derived structural thresholds and identified-combination ratios are stable across all three b/v normalizations. Table 3 reports the mean derived quantities:

Table 3: Derived quantities at the three b/v normalizations (mean across multistart seeds).

Quantity	$b/v = 0.06$	$b/v = 0.22$	$b/v = 0.33$
$\bar{n} = F/\Delta_A$ (saturation threshold)	226	222	237
$n^*_{D,S}$ (full decentralization)	108.6	109.8	108.9
n^{**} (regularity threshold)	14.8	14.9	15.0

The model's economically meaningful predictions are identified across the entire $b/v \in [0.06, 0.33]$ range: a firm of size $n \approx 109$ exits full decentralization, the model saturates at $n \approx 220$–240 workers, and the regularity threshold sits around 15 workers regardless of the b/v value chosen. The boundary conditions (B1) and (B2) and the regularity condition hold with similar large margins across all three normalizations. Sectoral predictions ($\theta^*_S(n_i)$ and $\theta^*_M(n_i)$ at each empirical bin) agree to within approximately 0.01 across the multistart optima at the three normalizations; part of the residual dispersion reflects optimizer noise at the multistart settings, and comparisons at the mean parameter vectors (which lie slightly off the identified ridge) can differ by more at a single manufacturing bin.

Implication for the central-case reporting. The choice of $b/v = 0.22$ is anchored on the Lazear (2000) Safelite estimate, the most-cited single estimate of within-worker incentive-channel productivity in personnel economics. The central-case calibration's headline results (fit correlations, threshold values, lifecycle predictions, comparative statics) are robust to the choice of b/v across the literature-anchored range. The structural parameters v, b, k, Δ_A, δ, F should be read as one consistent representation among many; the derived quantities $\bar{n}$, n^*_D, n^{**}, and the predicted θ^* and I trajectories are what the calibration actually identifies.

B.11 Verification of model conditions at calibrated values

The propositions in Section 2 are stated under boundary conditions (B1) and (B2), and the comparative statics in Propositions 2–3 require the regularity condition derived in Appendix A.3. This subsection verifies all three conditions at the calibrated parameters.

Computing the relevant thresholds. From the calibrated values, three structural thresholds can be computed:

Full decentralization threshold n^*_D (largest size at which $\theta^* = 0$). For services, $n^*_{D,S}$ is the unique positive root of $G(n) = 0$, where

$$G(n) = \frac{\beta v(v+b)n^{\zeta-1}}{k\Lambda} - \Delta_A + \frac{F}{n} - 2\delta n^{\alpha-1}.$$

For manufacturing, the innovation-force term vanishes and $n^*_{D,M}$ solves $F/n - \Delta_A = 2\delta n^{\alpha-1}$.

Full centralization threshold n_C^* (smallest size at which $\theta^* = 1$). Solves $N(n) = 0$ where $N(n)$ is the numerator of the closed-form $1 - \theta^*$.

Saturation threshold $\bar{n} = F/\Delta_A$, the size at which the operations-only first-order condition first hits $\theta = 1$.

Regularity threshold n^{**}, given explicitly in primitives by equation (32) of Appendix A.3.

Computed values at the calibrated parameters.

Threshold	Value (employees)
$n_{D,S}^*$ (full decentralization, services)	107.5
$n_{D,M}^*$ (full decentralization, mfg.)	67.7
$n_{C,S}^*$ (full centralization, services)	616.7
$n_{C,M}^*$ (full centralization, mfg.) $= \bar{n}$	220.4
$\bar{n}$ (saturation threshold)	220.4
$n^{(a)}$ (interior comparative-static threshold, services)	50.5
n^{**} (regularity threshold, services)	14.7
n^{**} (regularity threshold, mfg.)	0

The new row $n^{(a)} = [\beta\sigma\zeta v_I(v_I + b)/(k\Delta_A)]^{1/(1-\zeta)}$ is the explicit threshold above which the interior comparative static $\partial\theta^*/\partial n > 0$ of Proposition 2(a) is guaranteed (Appendix A.3). Since $n^{(a)} = 50.5 < n_{D,S}^* = 107.5$, the condition holds at every point of the interior regime: the "relevant range" qualifier in the interior comparative statics is verified, not assumed.

Boundary condition (B1): Born decentralized. The empirical sample's smallest size bin is $n_1 = 7$ employees. The services full decentralization threshold $n_{D,S}^* = 107.5$ exceeds this by a factor of 15.4. Manufacturing analogue: $n_{D,M}^* = 67.7$, exceeding n_1 by a factor of 9.7. (B1) holds with very large margin in both sectors. At the smallest empirical bins (5–9 and 15–19 employees), the calibrated model predicts $\theta_S^* = \theta_M^* = 0$ exactly, matching the empirical observation that small ventures are near-fully decentralized.

Boundary conditions (B2)–(B3): Saturation and interior passage. Two distinct statements must be kept apart here. The first is about the *cross-sectional policy*: the empirical sample's largest size bins (350, 625, 1500 employees) all exceed the saturation threshold $\bar{n} = 220.4$, the largest by a factor of 6.8, so the saturated-regime policy $\theta^* = 1$ applies to the large incumbents in the sample as entities, and the closed forms used by the calibration are valid across the empirical size range. The policy function crosses $n^*_{D,S} = 107.5$ between the 50–99 and 150–199 bins (consistent with the empirical centralization profile) and reaches full centralization at $n^*_{C,S} = 616.7$, comfortably below the 1000+ bin.

The second statement is about the *trajectory of an individual firm*, and here the verification is more demanding. Along the calibrated equilibrium path, per-stage growth is $\phi \cdot [\Pi_O + v_I I]/n \approx$ 10–15%: a services firm born at $n_1 = 5$ reaches only $n_3 \approx 6.3$, and a firm starting at the edge of the decentralized regime ($n_1 = n^*_{D,S}$) reaches $n_3 \approx 131 < \bar{n}$. Joint satisfaction of (B1)–(B3) (traversal from below $n^*_{D,S} = 107.5$ to above $\bar{n} = 220.4$ within two stage transitions) requires average per-stage growth of about 43%, an effective reinvestment rate of $\phi \approx 0.42$, three to four times the average rate estimated from single-unit continuing establishments ($\phi \approx$ 0.10–0.15, Appendix B.13). The full three-regime traversal therefore characterizes the high-growth scale-ups in the right tail of the growth distribution, not the average survivor, and the average survivor's predicted path (remaining decentralized at small scale) is itself what the data show in the small bins. The cross-section calibration is a static policy-function exercise across heterogeneous firms and does not require any single firm to traverse all bins; Appendix A.6 develops this distinction in full.

Regularity condition. The boundary-restricted regularity condition $n^*_D > n^{**}$ requires the firm to have grown past the regularity threshold by the time it enters the interior regime. In services, $n^*_{D,S} = 107.5$ exceeds $n^{**} = 14.7$ by a factor of approximately 7. In manufacturing, $\sigma = 0$ makes $n^{**} = 0$ identically, and the regularity condition holds trivially. The general regularity condition $n_1 > n^{**}$ marginally fails at the smallest empirical size bin

($n_1 = 7 < n^{**} = 14.7$), but this is irrelevant because the model puts the firm at the corner $\theta_1^* = 0$ at that size, where no interior optimization is being performed and the second-order condition is not at issue.

Predicted centralization and innovation at each size bin.

n	$\theta_S^*(n)$	$\theta_M^*(n)$	$I_S(n)$	$I_M(n)$
7	0.000	0.000	1.204	1.204
17	0.000	0.000	1.438	1.438
75	0.000	0.227	1.935	1.935
175	0.667	0.957	0.763	2.292
350	0.949	1.000	0.135	2.633
625	1.000	1.000	0.000	2.956
1500	1.000	1.000	0.000	3.522

The pattern is exactly the empirical one: services innovation rises with size in the decentralized regime, peaks just before the policy enters the interior regime around 75–175 employees, then collapses as centralization kills initiative; manufacturing innovation rises monotonically because operations centralization does not affect R&D. Services centralization rises sharply between 175 and 625 employees, reflecting the interior segment of the policy function. The thresholds bracket the empirical size range exactly as the boundary conditions require.

Summary of verification. At the calibrated parameters:

- (B1) holds: $n_1 = 7 \ll n_{D,S}^* = 107.5$.
- (B2) holds in the policy sense: the large bins satisfy $n \geq \bar{n} = 220.4$ (largest bin by a factor of 6.8), so the saturated closed forms apply across the empirical size range.
- (B2)–(B3) as trajectory conditions require an effective reinvestment rate of $\phi \approx 0.42$, three to four times the estimated average; they characterize high-growth scale-ups

rather than the average surviving establishment (Appendix A.6).

- Interior comparative statics hold throughout the interior regime: $n^{(a)} = 50.5 < n^*_{D,S} = 107.5$.
- Boundary-restricted regularity holds: $n^*_{D,S} = 107.5 \gg n^{**} = 14.7$.
- Manufacturing analogues hold by larger margins or trivially.

The propositions in Section 2 therefore apply at the calibrated parameters, and the implicit function theorem comparative statics in Propositions 2–3 are mathematically well-founded across the empirically relevant range of firm sizes.

B.12 Interpreting the fit

What the correlations test. The four correlations are not equally informative. Manufacturing innovation ($\rho = 0.976$) is almost mechanical: $I_M = n^\zeta r/k$ is monotonically increasing by construction, and the data are also monotonic. Any positive ζ would achieve a high correlation. Similarly, centralization in both sectors ($\rho \approx 0.93$) is partly structural: $\theta^*(n)$ is monotonically increasing by construction (Proposition 2(a)), and the data are also broadly monotonic. The genuinely informative test is services innovation ($\rho = 0.944$), where the model must produce the correct peak location (around 50–100 employees) and the correct rate of decline. This is the one panel where the model's shape is not predetermined by its functional form.

Degrees of freedom. The central-case calibration has 26 cross-section moments and 8 estimated parameters (with π_0, β, ϕ, b/v fixed externally). Correlation is invariant to shift and scale, so each 7-moment series has 5 effective shape degrees of freedom and each 6-moment series has 4, giving 18 effective shape d.o.f. against 8 parameters (ratio 2.2:1). The lifecycle-augmented calibration of Appendix B.13 adds 6 moments (3 stages $\times$ 2 sectors) and 1 parameter (ϕ), improving the ratio to 24 effective d.o.f. against 9 parameters (2.7:1). The sector-specific extension of Appendix B.14 adds 6 sector-specific parameters (4 estimated sectorally instead of jointly) for a total of 14 parameters against 32 moments (24 effective shape d.o.f., ratio 1.7:1). All three versions are best understood as shape-matching tests whether a single mechanism can qualitatively reproduce the observed moment patterns, not as precisely identified structural estimations.

Scale factors. The scale factor $s^*_{I,S} = 0.28$ for services innovation indicates that the model's raw innovation units are approximately 3.6 times the data's LPM coefficient units. This reflects the different units, not a mismatch: the data are linear probability model coefficients (bounded between -1 and 1), while the model's innovation is in arbitrary units. The scale factors are not free parameters, they are determined by OLS given the model predictions, and are reported for transparency.

Boundary and regularity conditions as emergent, not imposed. The objective function (42) does not explicitly enforce the boundary conditions or the regularity condition. The only constraints in the optimization are the parameter bounds, the worker-effort interiority constraint $k \geq v + b$, and a soft monotonicity penalty on $\theta^*(n)$. The fact that (B1), the threshold orderings $n^{**} < n^{(a)} < n^*_D < \bar{n} < n^*_C$, and the regularity condition all hold at the calibrated values with substantial margin (verified above) is a property of the data, not of the calibration procedure: the empirical cross-sectional pattern (small establishments decentralized, large firms centralized) places the thresholds exactly where the boundary conditions need them. The trajectory components of (B2)–(B3), by contrast, are genuinely restrictive, they single out the high-growth parameterization quantified above, and are not implied by the cross-section fit.

Empirical validity of the large-firm closed form. The closed-form expressions (38)–(41) are derived under the assumption that the terminal-stage solution saturates at $\theta^*_3 = 1$, so that the growth multiplier $\Lambda_2(n_3)$ collapses to the parameter $\Lambda = 1 + \beta\phi(\pi_0 + \Delta_A)$. At the calibrated parameters, this assumption is essentially without loss. Exact numerical backward induction of the full three-stage problem (using the corrected recursion of Section A.2, in which Stage-1 innovation cash is reinvested at Stage 2) shows (replication notebook): (i) the Stage-2 growth multiplier $\Lambda_2(n)$ varies in the narrow interval [1.0492, 1.0503] across the empirically relevant range of firm sizes (the minimum is attained at $n \approx n^*_{D,M}$, where the Stage-3 coordination drag on the marginal value of size peaks), against the large-n limit of 1.0503; (ii) the penultimate-stage closed form reproduces the exact Stage-2 policy with maximum θ^* error below 3×10^{-6} at every size bin; (iii) the Stage-1 multiplier is $\Lambda_1 \approx 1.078$, and using the penultimate-stage closed form for Stage 1 misstates θ^*_1 by at most 0.008 across the bins, with a dense-grid peak of approximately 0.017 just above n^*_D (Section A.4), while the corrected Stage-1 form of Section A.4 (innovation weight $\beta\Lambda/\Lambda_1$) reduces this to below 10^{-4} at every bin and below 7×10^{-4} everywhere. All of the bin-level errors are orders of magnitude smaller than the cross-bin variation that the calibration is fitting. The large-n closed form is therefore not just a theoretically convenient specialization; it is the exact penultimate-stage policy and an accurate approximation at Stage 1, and the comparative statics in Propositions 2–4 can be read off the closed form with confidence.

B.13 Lifecycle moments and the dynamic calibration

The cross-section calibration of Appendix B.7–B.11 matches 26 size-bin moments at a point in time. The model also makes predictions about firm growth over time through its growth equation $n_{t+1} = n_t + \phi[\Pi_O + v_I I_{t-1}]$, which the cross-section moments do not discipline. This subsection extends the calibration to incorporate lifecycle growth-rate moments from the same WES sample and uses them to identify the reinvestment rate ϕ, which the baseline calibration sets externally at $\phi = 0.1$.

Empirical lifecycle moments. Kueng et al. (2017) report age-effect regressions for single-unit continuing establishments in services and manufacturing separately. These regressions track the log-employment size effect by age (controlling for the size-bin coefficients used in the cross-section calibration), from age 1 to age 21. Reading values off the published figures at three stage midpoints (age 3, 10, and 17, corresponding to seven-year stages 1, 2, and 3 of the model), the empirical log-size effects relative to the age-0 baseline are approximately:

Sector	Stage 1 (age 3)	Stage 2 (age 10)	Stage 3 (age 17)
Services	0.10	0.32	0.58
Manufacturing	0.10	0.40	0.46

The corresponding stage-to-stage log-growth moments (computed by differencing) are:

Sector	$\Delta \log n$ $(0 \to 1)$	$\Delta \log n$ $(1 \to 2)$	$\Delta \log n$ $(2 \to 3)$
Services	0.10	0.22	0.26
Manufacturing	0.10	0.30	0.06

The two sectors share a similar early growth rate (≈ 0.10 log points in stage $0 \to 1$) but diverge thereafter: services growth accelerates through stage 3, while manufacturing growth peaks at stage 2 and collapses by stage 3. The empirical pattern reflects the population of single-unit continuers, a survival-conditioned subsample distinct from the cross-section, in which large multi-unit firms are dominant.

Construction of the model-side lifecycle moments. For a candidate parameter vector, simulate the firm trajectory starting from $n_0 = 5$ (chosen to match the typical age-0 single-unit continuer's small initial size, which is well below n_D^* and so lies in the decentralized regime). The model-side log-growth moments at the three stages are

$$g_t^\sigma \equiv \log\left(n_{t+1}^\sigma / n_t^\sigma\right), \qquad t \in \{0, 1, 2\}, \tag{45}$$

where $n_{t+1}^\sigma = n_t^\sigma + \phi\left[\Pi_O(n_t^\sigma, \theta_t^*(n_t^\sigma)) + v_I I_{t-1}^\sigma\right]$ and the initial innovation endowment is set to $I_{-1}^\sigma = 0$. Sector-specific evaluation uses the sector's σ in the closed forms for θ^* and I.

Augmented objective. The lifecycle moments are added to the cross-section objective as an additional sum of squared errors:

$$\mathcal{L} = - \sum_{p \in \{IS, \theta S, IM, \theta M\}} w_p \rho_p + \lambda_{LC} \sum_{\sigma \in \{S,M\}} \sum_{t=0}^{2} \left(g_t^{\sigma} - g_t^{\sigma,\text{data}}\right)^2, \tag{46}$$

where the cross-section component is unchanged and $\lambda_{LC} = 1$ places the lifecycle moments on roughly comparable scale to the cross-section correlations. The lifecycle component adds six moments (three stages $\times$ two sectors) to the 26 cross-section moments, for a total of 32 moments against 9 free parameters (the original 7 estimated plus ϕ, which is now free, plus the same fixed normalizations $\beta = 0.5$, $\pi_0 = 1$, $b/v = 0.22$).

Result: the reinvestment rate is interior and tightly identified. With ϕ freed and lifecycle moments included, the calibration produces an interior estimate $\phi = 0.150 \pm 0.001$ across multistart seeds at $b/v = 0.22$. The implied per-stage reinvestment elasticity is approximately 50% higher than the externally-set value of 0.1 in the baseline. Other parameters move modestly:

Parameter	Cross-section only (ϕ fixed)	Cross-section + lifecycle (ϕ free)
v	1.844	2.000 (at upper bound)
b	0.406	0.440
k	2.758	2.440
Δ_A	0.00653	0.0083
δ	9.06×10^{-5}	1.28×10^{-4}
α	2.043	2.012
ζ	0.200 (at bound)	0.200 (at bound)
F	1.44	1.83
ϕ	0.10 (fixed)	0.150 (estimated)

Multistart relative SDs improve across all parameters when the lifecycle moments are added: v from 5.7% to 0.14%, k from 11.2% to 0.13%, Δ_A from 9.2% to 1.4%, δ from 25.6%

to 9.3%, F from 10.5% to 3.9%. The lifecycle moments tighten identification substantially by linking the model's growth equation to observable growth rates.

Fit. The cross-section panel correlations at the lifecycle-augmented calibration are $\rho_{I,S} = 0.94$, $\rho_{\theta,S} = 0.93$, $\rho_{I,M} = 0.98$, $\rho_{\theta,M} = 0.94$ (essentially unchanged from the cross-section-only fit). The predicted services lifecycle log-growth $(0.14, 0.20, 0.19)$ tracks the empirical $(0.10, 0.22, 0.26)$ at the right order of magnitude, capturing the qualitative pattern of acceleration through stages 1–2 and leveling off in stage 3. The predicted manufacturing lifecycle $(0.14, 0.20, 0.19)$ matches the empirical early acceleration but fails to capture the late-stage decline (empirical 0.06 in stage 2–3 vs. model 0.19). The model lacks any mechanism for slowing firm growth in the saturated regime: once a firm centralizes, per-worker surplus rises from π_0 to $\pi_0 + \Delta_A$ and growth accelerates rather than decelerates. The manufacturing late-stage growth collapse is therefore a feature of the data the model is not designed to reproduce, plausibly reflecting demand-side saturation or capacity constraints outside the model's scope.

The shared-parameter limitation. A baseline calibration with shared $(\pi_0, \phi, \zeta, \Delta_A)$ across sectors cannot differentiate the sectoral lifecycle patterns: both sectors start at $\theta^* = 0$ at small sizes, so their predicted growth is identical until firms cross n_D^*. This shared-parameter calibration therefore matches the average of the two sectoral patterns, not either one individually. The sector-specific extension below relaxes this.

Implication for the trajectory conditions. The estimate $\phi = 0.150$ quantifies the traversal discussion of Appendix A.6: at this rate a firm born at $n_0 = 5$ grows by a cumulative factor of roughly 1.7 over three stages, well short of the $\bar{n}/n_D^*$ traversal that the full three-regime trajectory requires, which corresponds to an effective rate of $\phi \approx 0.4$. The lifecycle moments are estimated on single-unit *continuing* establishments, a survival-conditioned population dominated by slow growers; the model's trajectory proposition applies to the high-growth tail whose reinvestment rates are a multiple of this average. The lifecycle calibration disciplines the average growth technology; it does not (and should not) place the average establishment on the scale-up trajectory.

B.14 Sector-specific extension

The baseline calibration assumes all structural parameters except σ are shared across sectors. This subsection relaxes that assumption for the four parameters most likely to differ economically across services and manufacturing: per-worker operating surplus π_0, reinvestment rate ϕ, innovation returns-to-scale ζ, and alignment gain Δ_A. The shared parameters $(v_I, b, k, \delta, \alpha, F)$ are retained because they reflect technology or labor-market primitives that operate similarly across the two sectors at the level of the frontline production worker.

Setup. With sector-specific $(\pi_0^\sigma, \phi^\sigma, \zeta^\sigma, \Delta_A^\sigma)$ and shared $(v_I, b, k, \delta, \alpha, F)$, the model has $6 + 4 \times 2 = 14$ free parameters against 32 moments. The Lazear normalization $b/v_I = 0.22$ remains in force, and the discount factor is held at $\beta = 0.5$. The firm-value normalization is set on manufacturing: $\pi_0^M = 1$, so that π_0^S identifies the per-worker operating surplus in services as a multiple of the surplus in manufacturing. (Choosing manufacturing rather than services is innocuous; the choice fixes which sector's π_0 serves as the numeraire and does not affect predictions.)

Theoretical restrictions imposed. Two restrictions derived from the model's microfoundation are enforced as constraints:

(R1) *Project-microfoundation restriction:* $\Delta_A^\sigma < b$ in each sector, where $b = 0.22 \cdot v_I$ (Appendix A.8). This rules out parameter regions in which the worker would prefer the firm's project anyway.

(R2) *Worker-effort interiority:* $k \geq v_I + b$, ensuring $e^*(0) = (v_I + b)/k \leq 1$ in both sectors (since v_I and b are shared, this is a single shared constraint).

The lower bound on ζ^σ is relaxed to 0.001 to allow the data to express a preference for arbitrarily severe diminishing returns to scale in innovation; the upper bound is left at 1.0.

Estimated parameters. Across six multistart seeds at $b/v = 0.22$, the sector-specific calibration produces (means, with relative SD in parentheses):

Parameter	Estimate (rel SD across seeds)
Shared	
v_I	1.99 (5%)
b	0.44 (5%)
k	2.44 (5%)
δ	1.7×10^{-3} (32%)
α	2.011 (0.4%)
F	49.1 (34%)
Services-specific	
π_0^S	1.24 (1%)
ϕ^S	0.118 (20%)
ζ^S	0.091 (2%)
Δ_A^S	0.13 (28%)
Manufacturing-specific	
π_0^M	1.00 (1%)
ϕ^M	0.113 (18%)
ζ^M	0.001–0.013 (at lower bound)
Δ_A^M	0.12 (37%)

Five substantive findings emerge.

(1) Services has higher per-worker operating surplus than manufacturing. The ratio $\pi_0^S/\pi_0^M \approx 1.24$ is tight across all seeds and all three b/v values. After netting out wages, services workers contribute 24% more per-worker per-stage surplus than manufacturing workers.

This is consistent with the higher value-added per worker typical of services industries with substantial human-capital intensity (professional services, finance, healthcare).

(2) The reinvestment rate is essentially the same across sectors. $\phi^S \approx \phi^M \approx 0.12$, with the two sector-specific values within 5% of each other in every seed. The model does not detect a sectoral difference in the rate at which retained earnings translate into headcount growth. This justifies the simplifying assumption $\phi^S = \phi^M$ in alternative specifications.

(3) Manufacturing has more severe diminishing returns to innovation scale. ζ^M pins at the lower bound of 0.001 in every multistart seed, while $\zeta^S \approx 0.09$ settles at a clean interior value. The data prefer manufacturing's firm-level innovation to be essentially scale-invariant: doubling the workforce produces only a marginal increase in distinct firm-level innovations. The economic mechanism is consistent with the separability of R&D from production in manufacturing: innovations are produced by a separate R&D function whose output does not scale linearly with manufacturing headcount.

(4) Alignment gaps are similar in size across sectors and small relative to b. $\Delta_A^S \approx \Delta_A^M \approx 0.12$, while $b = 0.44$. The ratio $\Delta_A^\sigma / b \approx 0.27$ in each sector, well within the upper bound of 1 imposed by the project microfoundation. Centralization in both sectors is therefore primarily about coordination-cost relief, not alignment redirection.

(5) Sectoral centralization thresholds are very close. The derived $n^*_{D,S}$ and $n^*_{D,M}$ are both approximately 110, even though the underlying sector-specific parameters differ substantially. The full-decentralization threshold is robust to the choice of which parameters are sector-specific.

Limitations of the sector-specific extension. The sector-specific calibration improves identification of π_0^σ but does not eliminate the residual scale ambiguity in v_I. The data identify the inter-sector ratio π_0^S / π_0^M but not the absolute level of v_I relative to π_0^M to more than $\pm 10\%$, because the model retains its firm-value scale invariance even with both sectors estimated jointly. A related and sharper caveat applies to the decomposition of growth into π_0^σ and ϕ^σ separately: what the lifecycle moments pin down tightly is the per-stage growth capacity, the *product* $\phi^\sigma \pi_0^\sigma \approx 0.15$ in each sector, while the split of that product into a level and a reinvestment rate moves along a ridge that depends on the search bounds (independent re-estimation in the replication notebook recovers the same products with a different level/rate split). Findings (1) and (2) above should accordingly be read jointly, as statements about the identified products and their cross-sector similarity, rather than as separate point identifications of π_0^σ and ϕ^σ. The sectoral lifecycle patterns also continue to be only partially matched: services growth acceleration is well captured but manufacturing's late-stage growth collapse remains outside the model's scope.

B.15 Nascent stage and stage-specific reinvestment

The sector-specific calibration of Appendix B.14 captures the services lifecycle profile in stages 2 and 3 reasonably well but overstates stage-1 growth in both sectors (predicted 0.14/0.13 in services/manufacturing vs. empirical 0.10 in each) and understates the manufacturing stage-2 acceleration. This subsection introduces two additional parameters that improve the lifecycle fit while preserving the cross-section fit: an initial innovation endowment I_{-1}, and a stage-split reinvestment rate (ϕ_1, ϕ_{23}). Both extensions have substantive economic justifications discussed below.

The initial innovation endowment as nascent-entrepreneurship effort. The model's growth equation at $t = 1$ is $n_1 = n_0 + \phi_1[\Pi_O(n_0, \theta_0^*) + v_I I_{-1}]$, where I_{-1} represents innovation rents accumulated before the start of stage 1. The baseline sector-specific calibration sets $I_{-1} = 0$, interpreting age 0 as a literal zero-history starting point. This interpretation, however, ignores a real economic stage: *nascent entrepreneurship*, the period of founder activity preceding incorporation.

The KYH (2017) age-effect regressions, like most longitudinal establishment datasets, measure firm age from the year of incorporation. But meaningful innovation activity typically begins well before that point. A new firm at age 0 with 5 workers has often been in development for 1–3 years prior, during which co-founders, advisors, and unpaid early collaborators engage in opportunity discovery, customer development, prototype building, business model design, and securing initial pilot customers. These activities produce exactly the kind of firm-value innovations the model is about: they yield a working product, validated customer demand, and a defensible technical approach by the time of incorporation. For venture-capital-backed firms, the first check typically pays for precisely this pre-incorporation intellectual output, not for innovations yet to be produced.

The reduced-form parameter I_{-1} in the calibration absorbs this entire nascent stage of innovation activity. Setting $I_{-1} = 0$ is equivalent to assuming founders contributed nothing before incorporation, which is empirically implausible. Allowing I_{-1} to be estimated lets the data tell us how large the nascent-stage innovation endowment is.

To assess whether the estimated value is economically reasonable, note that nascent-stage activity is concentrated in three ways. First, it is performed by a small but highly focused team, typically 2–4 active co-founders during the nascent period, of whom only a fraction will appear in the post-incorporation establishment data. Second, it is allocated almost entirely to innovation rather than operations: there are no customers to serve, no operations to run, no replication to do. The founder's 100% innovation allocation contrasts sharply with the split allocation of workers in an established firm, where time is divided between operations and innovation. Third, founder teams self-select on entrepreneurial talent, so per-time productivity is plausibly higher than that of an average established-firm worker. Combined, these three factors imply nascent-stage per-time productivity that is 5–10× that of established small-firm workers, applied over 2–3 years before incorporation. The implied I_{-1} is then roughly equivalent to one stage's worth of innovation output from a 50-worker firm, substantial, but consistent with the entrepreneurship literature on founder effort.

Stage-specific reinvestment as financial frictions. The baseline calibration uses a single reinvestment rate ϕ throughout the firm's lifecycle. This is a strong assumption: it implies the firm can convert one dollar of profit into the same fractional growth in workers regardless of its age or financial maturity. Real firms face financial frictions whose intensity varies systematically with firm age. New firms typically have limited access to external finance, must rely on retained earnings and founder savings, and face higher costs of capital due to information asymmetries. As firms age, they build collateral, develop banking relationships, and gain access to broader capital markets, all of which reduce the cost of capital and effectively raise the marginal reinvestment rate. The model abstracts entirely from this financial-frictions dimension. Allowing a stage-split reinvestment rate (ϕ_1, ϕ_{23}) provides a reduced-form representation of this evolution, distinguishing nascent firms (stage 1) from financially-mature firms (stages 2–3).

Specification. Relative to the sector-specific extension, the two sector-specific reinvestment rates (ϕ^S, ϕ^M) are replaced by a shared stage-split pair (ϕ_1, ϕ_{23}) (the sector-specific calibration found $\phi^S \approx \phi^M$, so the sectoral split is dropped in favor of the stage split) and a shared initial endowment I_{-1} is added. In the parameter-counting convention of Appendix B.14 this gives $14 - 2 + 3 = 15$ parameters against 32 moments, of which 13 are estimated freely (b is tied to v_I by the Lazear anchor and $\pi_0^M = 1$ is the numeraire). The lifecycle simulation starts the firm at $n_0 = 5$ with the endowment I_{-1}, applies ϕ_1 to the first stage transition and ϕ_{23} to the remaining two, and is otherwise identical to Appendix B.13. The search bounds are those of the sector-specific extension, with $\phi_1, \phi_{23} \in [0.005, 0.4]$ and $I_{-1} \in [0, 5]$.

Estimated parameters. Multistart estimation at $b/v = 0.22$ converges to the same optimum from every seed:

Parameter	Estimate	Interpretation
ϕ_1	0.034	Stage-1 reinvestment rate
ϕ_{23}	0.153	Stages-2 and -3 reinvestment rate
ϕ_1/ϕ_{23}	0.22	Nascent firms reinvest at roughly one-fifth the mature-firm rate
I_{-1}	4.6	Nascent-stage innovation endowment (weakly identified; see text)
$\phi_1(\pi_0^S + v_I I_{-1}/n_0)$	0.112	Identified stage-1 growth combination
π_0^S	1.43	Services per-worker surplus (mfg normalized to 1)

Three features of these estimates deserve comment. First, the mature-firm rate $\phi_{23} \approx 0.15$ coincides with the single-ϕ estimate of the lifecycle-augmented calibration (Appendix B.13, $\phi = 0.150$): freeing the nascent stage leaves the mature-firm reinvestment technology where the simpler specification put it, which is reassuring for both. Second, the ratio $\phi_1/\phi_{23} \approx 0.22$ is the central new finding: nascent firms convert profit into headcount at roughly one-fifth the mature-firm rate. Economically, this is consistent with the financial-frictions interpretation: young firms convert a smaller fraction of marginal profit into headcount because some of that profit is absorbed by debt service, lack of credit access, or precautionary savings under uncertainty. Third, ϕ_1 and I_{-1} are *not* separately well identified, and the table makes this explicit: both operate on the single stage-1 growth moment, the estimate of I_{-1} is large and poorly pinned (it drifts over $[4.2, 4.6]$ across normalizations and reaches the upper search bound of 5 under some optimizer seeds) and what the data actually determine is the combination $\phi_1(\pi_0^S + v_I I_{-1}/n_0) \approx 0.112$, the stage-1 growth rate itself. (The combination, unlike its components, is stable to three digits across normalizations and seeds; see the robustness table below.) The estimated split between a low ϕ_1 and a large endowment should therefore be read as one representation of a slow first stage, not as a point identification of pre-incorporation innovation output. With that caveat, the magnitude is directionally consistent with the nascent-entrepreneurship literature's view of a substantial pre-incorporation stock: a 2–3-year nascent period with 2–4 founders working at several times established-firm-worker innovation productivity.

Fit. The lifecycle fit improves relative to the baseline sector-specific calibration, with the gain concentrated at stage 1 (both columns computed under the specifications defined in this appendix; replication notebook):

Sector / Stage	Data	Sector-specific (Appx. B.14)	With nascent stage
Services 1	0.10	0.14	0.11
Services 2	0.22	0.24	0.24
Services 3	0.26	0.22	0.23
Manufacturing 1	0.10	0.13	0.09
Manufacturing 2	0.30	0.18	0.19
Manufacturing 3	0.06	0.17	0.18
SSE (lifecycle)	,	0.032	0.028

The lifecycle SSE improves by about 12%, with the improvement coming almost entirely from stage 1: predicted first-stage growth falls from 0.14/0.13 (services/manufacturing) to 0.11/0.09, against data of 0.10/0.10, the slow nascent stage is exactly the feature the extension was built to capture. The services profile (0.11 → 0.24 → 0.23) tracks the data (0.10 → 0.22 → 0.26) in both direction and magnitude. The manufacturing stage 2 acceleration is only partially captured (0.19 vs. 0.30), and the stage 3 manufacturing collapse remains outside the model's scope: the data drop from 0.30 to 0.06, while the model predicts a mild slowdown from 0.19 to 0.18.

The cross-section fit is essentially unchanged. Pearson correlations between model predictions and empirical moments are $\rho_{I,S} = 0.93$, $\rho_{\theta,S} = 0.95$, $\rho_{I,M} > 0.99$, $\rho_{\theta,M} = 0.94$, all within ± 0.01 of the sector-specific calibration's values. The additional parameters identify off lifecycle moments alone, without compromising the cross-section dimension.

Robustness across normalizations. Estimating the nascent-stage extension at the three b/v values from Appendix B.10 (0.06, 0.22, 0.33) confirms the identification structure described above: the identified quantities are stable across normalizations, while the weakly identified components of the (ϕ_1, I_{-1}) pair wander along their ridge:

b/v	ϕ_{23}	ϕ_1/ϕ_{23}	I_{-1}	$\phi_1(\pi_0^S + v_I I_{-1}/n_0)$	π_0^S
0.06	0.153	0.24	4.2	0.112	1.42
0.22	0.153	0.22	4.6	0.112	1.43
0.33	0.153	0.23	4.4	0.112	1.43

The mature-firm rate ϕ_{23}, the stage-1 growth combination, and π_0^S are stable to two or three digits, and the lifecycle predictions are essentially identical across all three normalizations, with sector-by-stage predictions varying by at most 0.001 in absolute terms. The components ϕ_1 and I_{-1} are individually the least stable quantities across normalizations and optimizer seeds, as the ridge structure implies they should be. The nascent-stage extension's identified content is therefore robust to the choice of Lazear anchor for b/v.

Caveats. Both extensions are reduced-form patches. I_{-1} is a stand-in for the entire nascent-entrepreneurship stage that the model does not explicitly represent; a fully structural treatment would model the pre-incorporation period as an additional model stage with its own decision-theoretic structure and primitives. The stage-split (ϕ_1, ϕ_{23}) is a reduced-form representation of financial frictions that are absent from the model's frictionless reinvestment equation; a fully structural treatment would embed an explicit financing constraint. Both extensions are useful in the sense that they identify a meaningful dimension along which the data depart from the baseline model's predictions, but they do not constitute structural identifications of the underlying nascent-stage parameters or financial-friction parameters. The estimated magnitudes should be interpreted as reduced-form quantities whose values absorb whatever combination of nascent-stage innovation effort and financial frictions the data require to match the lifecycle moments.

A direct test of the nascent-entrepreneurship interpretation would be to estimate I_{-1} separately for venture types with different nascent stage lengths. Venture-capital-backed technology startups, with long discovery and prototyping periods, should require larger I_{-1} than franchised retail or restaurant ventures, where the pre-opening period is short and most innovation occurs at the franchisor level. This cross-venture-type variation in I_{-1} is a falsifiable prediction of the interpretation offered here and a useful direction for future work.

B.16 Bounds, restrictions, and the residual under-identification

The calibration imposes a hierarchy of restrictions on the parameter space, distinguishing fundamental mathematical requirements from plausibility bounds and economic-microfoundation restrictions. This subsection makes the hierarchy explicit and characterizes the residual under-identification.

Fundamental mathematical restrictions. Three restrictions are required for the model's expressions to be well-defined:

(M1) $\theta \in [0, 1]$ (centralization is a fraction). Enforced by the closed-form clamping.

(M2) $k \geq v_I + b$ (worker effort $e^* \leq 1$, the probability bound). Enforced as a hard constraint via a large objective penalty if violated.

(M3) $\alpha > 2$ (super-quadratic coordination costs in firm size). Required for the regularity condition (Appendix A.3). Enforced as $\alpha \geq 2.01$ in the search bounds.

Theoretical-microfoundation restrictions. Two restrictions are derived from the model's economic structure:

(T1) $\Delta_A \leq b$ (project microfoundation, Appendix A.8). Strictly required for the alignment problem to exist; otherwise the worker would already pursue the firm-preferred project. Imposed in the sector-specific extension but not in the central case (where the data place Δ_A far below this bound regardless).

(T2) $\zeta \in (0, 1)$ (diminishing-but-positive returns to innovation scale). Required for the innovation function $I = n^\zeta r/k$ to exhibit the desired economic content. Enforced as $\zeta \geq 0.001$ in the relaxed sector-specific search; $\zeta \geq 0.2$ in the original calibration as a plausibility floor.

Plausibility bounds. Two restrictions are researcher choices reflecting reasonable economic magnitudes:

(P1) $v_I \leq 2$ (innovation value at most two stages of operating surplus per worker). Imposed in the central-case search bounds. Wider bounds ($v_I \leq 10$) are explored in the sensitivity analysis and produce slightly improved fit (-0.771 vs. -0.769) with v_I pushing to the wider upper bound. The economic predictions are essentially invariant.

(P2) $\zeta \geq 0.2$ as a soft floor on innovation returns to scale, reflecting plausible ranges in firm-level R&D production functions (Jones and Williams, 1998). Relaxing this to $\zeta \geq 0.01$ produces an interior estimate $\zeta \approx 0.09$ in the cross-section-only calibration, with fit improving from -0.763 to -0.769. The qualitative predictions (location of the services innovation hump, centralization thresholds) are essentially invariant.

Normalizations. Two parameters are fixed as normalizations because the model has exactly one scale invariance (the firm-value scaling) and one externally-anchored quantity:

(N1) $\pi_0 = 1$: the per-worker operating surplus is the numeraire. This uses the model's one degree of scale invariance.

(N2) $b/v_I = 0.22$ (Lazear anchor) or alternative values 0.06 and 0.33: this fixes the worker's private-benefit ratio externally, resolving the under-identification of b documented in Appendix B.7. In the sector-specific extension, this ratio is held constant across sectors but v_I remains shared, so b is implicitly the same in both sectors.

Externally-calibrated parameters. Three parameters are set from external evidence rather than estimated:

(E1) $\beta = 0.5$: per-stage discount factor, consistent with $\beta_A = 0.10$ annual discount rate (close to long-run S&P 500 average) compounded over 7 years (Section 3.5).

(E2) $\phi = 0.1$ in the cross-section-only calibration. Freed and estimated at $\phi \approx 0.15$ when lifecycle moments are included (Appendix B.13). In the nascent-stage extension (Appendix B.15), ϕ is split into $\phi_1 \approx 0.03$ (stage 1) and $\phi_{23} \approx 0.15$ (stages 2–3), absorbing financial-frictions variation across the firm lifecycle while leaving the mature-firm rate at the lifecycle-augmented estimate.

(E3) $\sigma_S = 1$, $\sigma_M = 0$: separability of operations from innovation. Theoretically motivated by Section 3.3.

Reduced-form patches for extensions. The nascent-stage extension (Appendix B.15) introduces two parameters that the model does not derive structurally but that absorb economically interpretable forces operating outside the model:

(R1) $I_{-1} \geq 0$ (initial innovation endowment). Absorbs the firm-value innovation output of the nascent entrepreneurship stage that precedes incorporation. Weakly identified individually (it trades off one-for-one against ϕ_1, drifting across normalizations and seeds); the identified object is the stage-1 growth combination $\phi_1(\pi_0^S + v_I I_{-1}/n_0) \approx 0.11$ (Appendix B.15).

(R2) (ϕ_1, ϕ_{23}) with $\phi_1 < \phi_{23}$ (stage-split reinvestment rate). Absorbs financial-frictions variation in the rate at which retained earnings translate into headcount growth across the firm's lifecycle. Estimated at $\phi_1 \approx 0.03$, $\phi_{23} \approx 0.15$; the ratio $\phi_1/\phi_{23} \approx 0.2$ and the stage-1 growth combination $\phi_1(\pi_0^S + v_I I_{-1}/n_0) \approx 0.11$ are the identified quantities (Appendix B.15).

These are reduced-form in that they substitute a small number of free parameters for the structural mechanisms (a nascent-stage decision-theoretic model; a financing constraint) that would generate them in a fuller treatment. They are nonetheless identified from the lifecycle moments alone, the cross-section fit is essentially unchanged with or without them.

Residual under-identification. Even with all restrictions imposed, the calibration retains one residual under-identified direction. The cross-section moments identify five combinations of the eight free parameters:

$$\bar{n} = F/\Delta_A, \qquad \delta/\Delta_A, \qquad v_I(v_I + b)/(k\Delta_A), \qquad \alpha, \qquad \zeta, \tag{47}$$

plus the lifecycle moments add a sixth (the reinvestment rate ϕ) when these are included. The individual values of Δ_A, δ, F are identified only jointly through these ratios; multistart seeds explore the equivalent-fit ridge along which Δ_A, δ, F co-scale.

The under-identification is a feature, not a bug: it tells us that the model's economic predictions (centralization thresholds, θ^* trajectories, innovation hump location) are robust to which point on the equivalent-fit ridge is selected, while the individual structural parameter values depend on the bound choices. The calibration's identified economic content is therefore best summarized by the derived quantities $\bar{n}$, n_D^*, n^{**}, and the predicted θ^* and I trajectories, all of which are stable across normalizations and bound choices. The structural parameter levels should be read as one consistent representation among many.

What additional moments would resolve. Additional empirical moments that would tighten identification along the residual direction include: (i) absolute innovation rates (the level of I at a benchmark firm size, which would pin v_I in absolute units); (ii) the share of firm value attributable to innovation versus operating profits (which would pin $v_I \cdot I/\Pi_O$ at a benchmark size); (iii) within-firm R&D-to-revenue ratios in the calibration sample, which would constrain the ratio of innovation rents to operating cash flow. None of these are directly available in the WES sample at the level of granularity needed; they are flagged here as targets for future work that builds on this calibration with richer firm-level data.

C Extension: Explicit Replication with Multiple Establishments

The baseline model captures replication indirectly: centralization raises operating profits, which fund growth. This appendix introduces replication explicitly by allowing the firm to open additional establishments, each a copy of the standardized original. The extension is shown to preserve all the baseline's qualitative predictions (Propositions 2–4) and to add two new comparative statics, on the per-establishment replication profit and on the cost of opening new establishments.

C.1 Setup

The firm operates an original unit with n workers and chooses both centralization $\theta \in [0, 1]$ and the number of replicated establishments $M \geq 0$. Each replicated establishment executes the firm's standardized template and earns per-establishment profit $\pi_M \cdot \theta$, where $\pi_M > 0$ is the per-unit profit at full standardization. At $\theta = 0$ (no standardization), there is nothing to copy and each establishment earns zero. Opening establishments entails a convex cost $c_M M^2/2$, reflecting market saturation, cannibalization across units, and increasing difficulty of finding suitable locations or markets.

Total per-period profits are:

$$\Pi^R(n, \theta, M) = \underbrace{\Pi_O(n, \theta)}_{\text{original unit}} + \underbrace{M \cdot \pi_M \theta}_{\text{replicated units}} - \underbrace{\frac{c_M M^2}{2}}_{\text{expansion cost}}, \tag{48}$$

where $\Pi_O(n, \theta) = n[\pi_0 + \Delta_A \theta] - \delta n^\alpha (1 - \theta)^2 - F\theta$ is the baseline operating profit from the original unit. Innovation occurs only in the original unit, $I(n, \theta) = n^{\zeta}(1 - \sigma\theta)(v_I + b)/k$; replicated establishments execute the template but do not innovate. This is the organizational content of replication: copies reproduce the firm's *standardized* operations, not its *innovative* activities.

C.2 Optimal number of establishments

The firm jointly chooses θ and M. Conditional on θ, the first-order condition for M is straightforward because M enters the growth equation only through its contribution to total profits, and M does not depend on the future state. Differentiating the firm's value with respect to M at any interior stage gives $\Lambda^R \cdot (\pi_M\theta - c_M M) = 0$, where Λ^R is the growth multiplier in the replication-augmented model (defined below). Since $\Lambda^R > 0$:

$$M^*(\theta) = \frac{\pi_M \theta}{c_M}. \tag{49}$$

The result is intuitive: the firm opens more establishments when standardization is higher (θ up), when per-unit profit is higher (π_M up), and when expansion costs are lower (c_M down). The growth multiplier cancels because it scales both the marginal benefit and the marginal cost equally.

C.3 First-order condition for centralization

Substituting $M^*(\theta)$ into total profits yields the reduced-form per-period profit:

$$\Pi^R(n, \theta, M^*(\theta)) = \Pi_O(n, \theta) + \frac{\pi_M^2 \theta^2}{2c_M}. \tag{50}$$

The replication profit $\pi_M^2\theta^2/(2c_M)$ is the firm's maximized return from the extensive margin: it is quadratic (convex) in θ because both the number of establishments and the profit per establishment rise with standardization.

The interior-stage value function under replication satisfies

$$J_t^R = \Pi_O(n_t, \theta_t) + \frac{\pi_M^2 \theta_t^2}{2c_M} + v_I I_{t-1} + \beta J_{t+1}^{R*}(n_{t+1}),$$

where the growth equation now includes replication profits:

$$n_{t+1} = n_t + \phi \left[\Pi_O(n_t, \theta_t) + \frac{\pi_M^2 \theta_t^2}{2c_M} + v_I I_{t-1} \right].$$

Differentiating J_t^R with respect to θ_t and using $dn_{t+1}/d\theta_t = \phi[\partial \Pi_O/\partial \theta_t + \pi_M^2 \theta_t / c_M]$ (since both operating profit and replication profit fund growth) yields, after collecting the direct and indirect effects of θ_t on $\Pi_O + \pi_M^2 \theta^2/(2c_M)$:

$$\Lambda_t^R(n_{t+1}) \cdot \left[\frac{\partial \Pi_O}{\partial \theta_t} + \frac{\pi_M^2 \theta_t}{c_M} \right] + \beta v_I \frac{\partial I_t}{\partial \theta_t} = 0, \tag{51}$$

where the replication-augmented growth multiplier is

$$\Lambda_t^R(n_{t+1}) \equiv 1 + \beta\phi \cdot \frac{dJ_{t+1}^{R*}}{dn_{t+1}}.$$

Equation (51) is the analog of the baseline first-order condition (23). The only structural change is that the marginal-profit bracket multiplied by Λ_t^R now contains an additional positive term $\pi_M^2 \theta_t / c_M$, capturing the marginal replication profit from a unit of standardization. When $\pi_M = 0$, equation (51) reduces exactly to (23). As in the baseline, the condition as written is exact for the penultimate stage; at Stage 1 the innovation term carries the additional factor $\mu_2 = \Lambda_2^R > 0$ (Section A.2), which rescales the innovation weight by a positive constant and changes none of the comparative-static signs derived below.

C.4 Comparative statics via the implicit function theorem

Define $F^R(\theta; n, \beta, \Delta_A, \sigma, \pi_M, c_M)$ from the left-hand side of (51):

$$F^R(\theta; \cdot) \equiv \Lambda^R(n_{t+1}(\theta)) \cdot \left[\frac{\partial \Pi_O}{\partial \theta} + \frac{\pi_M^2 \theta}{c_M} \right] + \beta v_I \frac{\partial I}{\partial \theta},$$

so that the first-order condition is $F^R(\theta_R^*; \cdot) = 0$. By the implicit function theorem, $d\theta_R^*/dx = -(\partial F^R/\partial x)/(\partial F^R/\partial \theta)$, and the sign of each comparative static equals the sign of $\partial F^R/\partial x$ provided $\partial F^R/\partial \theta < 0$.

Second-order condition (concavity). Differentiating F^R with respect to θ:

$$\frac{\partial F^R}{\partial \theta} = \Lambda^R \cdot \left[\frac{\partial^2 \Pi_O}{\partial \theta^2} + \frac{\pi_M^2}{c_M} \right] + \frac{d\Lambda^R}{dn_{t+1}} \cdot \frac{dn_{t+1}}{d\theta} \cdot \left[\frac{\partial \Pi_O}{\partial \theta} + \frac{\pi_M^2 \theta}{c_M} \right] + \beta v_I \frac{\partial^2 I}{\partial \theta^2}.$$

The first bracket is $-2\delta n^\alpha + \pi_M^2/c_M$, since $\partial^2 \Pi_O/\partial \theta^2 = -2\delta n^\alpha$ (concavity of operating profit) and the replication-profit curvature is $+\pi_M^2/c_M$ (convex). For this to be negative we require:

$$2\delta n^\alpha > \frac{\pi_M^2}{c_M}. \tag{52}$$

This is a substantive condition, it says that the concavity of coordination costs in θ must dominate the convexity of replication profits in θ. Since δn^α grows as n^α with $\alpha > 2$ while the replication-curvature term is constant in n, condition (52) holds for all n above a threshold. For very small firms it can fail, in which case the objective is convex in θ and the firm chooses a corner: typically $\theta^* = 1$ ("franchise from day one," standardize fully to maximize copies). I focus on the interior case where (52) holds. The second term in $\partial F^R/\partial \theta$ is the reinvestment-feedback term: in the saturated regime (the case analyzed here, and the empirically relevant one) it is exactly zero, because J_{t+1}^{R*} is linear in n_{t+1} (per-worker surplus is constant at $\theta = 1$ and the replication profit $\pi_M^2/(2c_M)$ is constant in n), so $d\Lambda^R/dn_{t+1} = 0$. (In the interior regime this term carries the same positive sign and structure as the baseline reinvestment feedback, and the analogue of the Steps 2–7 argument of Section A.3 would be required; it is not repeated here.) The third term is zero because I is linear in θ. Hence $\partial F^R/\partial\theta < 0$ under (52), the second-order condition holds, and the implicit function theorem applies.

Baseline propositions are preserved. The parameters $(n, \beta, \Delta_A, \sigma)$ enter F^R in exactly the same way they enter F in the baseline, plus through Λ^R. In the saturated regime $\Lambda^R = \Lambda = 1 + \beta\phi(\pi_0 + \Delta_A)$ *exactly*: at $\theta = 1$ the replication profit $\pi_M^2/(2c_M)$ and the establishment count $M^* = \pi_M/c_M$ are both constant in n, so replication adds nothing to the marginal value of size dJ^{R*}/dn. (By the envelope theorem, replication does not enter dJ^{R*}/dn in the interior regime either, since $\pi_M^2\theta^2/(2c_M)$ has no direct dependence on n.) The signs of the baseline channels therefore carry over unchanged.

Result (a): $\partial\theta_R^*/\partial n > 0$. The replication term $\pi_M^2\theta/c_M$ does not depend on n, so $\partial^2[\pi_M^2\theta/c_M]/\partial\theta\partial n = 0$. The other channels are identical to the baseline: $\partial^2\Pi_O/\partial\theta\partial n = \Delta_A + 2\alpha\delta n^{\alpha-1}(1-\theta) > 0$ raises F^R, and $\partial^2 I/\partial\theta\partial n = -\sigma\zeta n^{\zeta-1}(v_I + b)/k \leq 0$ (strictly negative in services, zero in manufacturing) lowers it. Because $\alpha > 2 > \zeta$, the operating-profit channel grows faster in n than the innovation channel, so $\partial F^R/\partial n > 0$ in the relevant range, and θ_R^* increases with n.

Result (b): $\partial\theta_R^*/\partial\beta < 0$. As in the baseline, β enters F^R directly through the innovation term ($v_I \cdot \partial I/\partial\theta < 0$ for services), pushing toward decentralization, and indirectly through Λ^R, pushing toward centralization. The direct effect dominates because $\Lambda^R = 1+\beta\phi(\pi_0+\Delta_A)$ is bounded in the saturated regime (a finite marginal value of size) while the innovation effect scales with n^ζ. Hence $\partial F^R/\partial\beta < 0$, and θ_R^* decreases with β.

Result (c): $\theta_{R,S}^* < \theta_{R,M}^*$. As in the baseline, σ enters F^R only through $\partial I/\partial\theta = -\sigma n^\zeta(v_I + b)/k$. The replication-augmented bracket $[\partial\Pi_O/\partial\theta + \pi_M^2\theta/c_M]$ is identical across sectors. Hence at any θ, $F^R(\theta; \sigma = 1) < F^R(\theta; \sigma = 0)$, so the services firm's first-order condition is satisfied at a smaller θ than the manufacturing firm's.

Result (d): $\partial\theta_R^*/\partial\Delta_A > 0$. Δ_A enters F^R directly through $\partial\Pi_O/\partial\theta$ (positive: higher alignment gain raises the marginal return to centralization), and indirectly through Λ^R (positive: higher Δ_A raises the marginal value of size). Both make $\partial F^R/\partial\Delta_A > 0$, so θ_R^* rises with Δ_A.

New comparative statics. Two new parameters enter the replication-augmented model.

Result (e): $\partial\theta_R^*/\partial\pi_M > 0$. π_M enters F^R directly through $\partial(\pi_M^2\theta/c_M)/\partial\pi_M = 2\pi_M\theta/c_M > 0$ (at interior $\theta > 0$), which raises F^R and pushes toward more centralization. In the saturated regime Λ^R does not depend on π_M (the replication profit at $\theta = 1$ is constant in n), so the effect operates entirely through this direct channel; equivalently, in the closed form (53) the term $\pi_M^2/(c_M n)$ enters the numerator, raising θ_R^* throughout the interior regime. Hence $\partial F^R/\partial\pi_M > 0$, and firms with more profitable replication centralize more.

Result (f): $\partial\theta_R^*/\partial c_M < 0$. c_M enters F^R through $\partial(\pi_M^2\theta/c_M)/\partial c_M = -\pi_M^2\theta/c_M^2 < 0$, lowering the marginal return to centralization. As with π_M, in the saturated regime Λ^R is independent of c_M, so the effect operates through this direct channel; in the closed form (53) a higher c_M shrinks the replication terms $\pi_M^2/(c_M n)$, lowering θ_R^* throughout the interior regime. Hence $\partial F^R/\partial c_M < 0$, and firms facing higher expansion costs centralize less.

The optimal number of establishments inherits these properties through $M^* = \pi_M\theta_R^*/c_M$: M^* rises with n and π_M, falls with β and c_M, and is lower in services than manufacturing.

Innovation and growth propositions carry over. The proofs of Proposition 4 in Section A.7 use only the properties $\theta^*(n)$ increasing in n (Result (a) above) and $\theta_S^* < \theta_M^*$ (Result (c) above), both of which hold in the replication-augmented model. The hump-shaped pattern of services innovation, the monotonic increase of manufacturing innovation, the post-collapse growth result, and the flat establishment growth of replicated units therefore all carry over without modification. Replication amplifies centralization but does not alter the qualitative organizational lifecycle described by the baseline.

C.5 Closed-form solution in the saturated regime

When the firm operates in the saturated regime described in Section A.2 (terminal-stage centralization $\theta_3^* = 1$, so Λ^R becomes a parameter), the first-order condition (51) is linear in θ and admits a closed-form solution. Substituting $\partial \Pi_O / \partial \theta = n\Delta_A - F + 2\delta n^\alpha (1-\theta)$, using $\theta = 1 - (1-\theta)$ in the replication term, and rearranging:

$$\Lambda^R \left[n\Delta_A - F + \frac{\pi_M^2}{c_M} + \left(2\delta n^\alpha - \frac{\pi_M^2}{c_M} \right)(1-\theta) \right] = \frac{\beta v_I (v_I + b) n^\zeta}{k}.$$

Solving for $(1-\theta)$:

$$\theta_R^*(n) = \left[1 - \frac{\beta v_I (v_I + b)\, n^{\zeta-1}/(k\Lambda^R) \ - \ \Delta_A \ + \ F/n \ - \ \pi_M^2/(c_M n)}{2\delta n^{\alpha-1} - \pi_M^2/(c_M n)} \right]_0^1. \tag{53}$$

Compared to the baseline closed form (14), the replication terms $\pi_M^2/(c_M n)$ enter both the numerator (reducing it, pushing θ^* upward) and the denominator (also reducing it, amplifying the response of θ^* to any parameter change). The condition $2\delta n^\alpha > \pi_M^2 / c_M$ from (52) guarantees the denominator is positive.

C.6 Convergence to the baseline

Both replication terms $\pi_M^2/(c_M n)$ vanish as $n \to \infty$: for large firms, the per-worker replication gain becomes negligible and the extension converges to the baseline. Replication matters most for mid-sized firms in the transition region where the centralization decision is actively contested. This is consistent with the observation that the replication decision is most salient for firms in the scaling phase, not for very small startups (which have nothing to replicate) or very large firms (which have already standardized for other reasons).

At $\pi_M = 0$, the extension collapses exactly to the baseline. The baseline is therefore the conservative case: it captures the minimum organizational cost of scaling, driven purely by the intensive margin of operational efficiency. Any positive replication value strengthens the case for centralization and accelerates the innovation collapse.

C.7 Interpretation: why the baseline suffices

The extension reveals that replication is a *fourth primitive* governing organizational design, alongside ownership patience (β), separability (σ), and incentive alignment (Δ_A). The replication parameters π_M and c_M determine how much additional centralization pressure the extensive margin creates.

However, the fact that the baseline model (with $\pi_M = 0$) already matches the data well (average correlation 0.95 across four panels) suggests that the intensive margin alone captures the dominant organizational force. The replication channel amplifies centralization but is not required to explain the observed patterns. This is consistent with the data source: the Canadian Workplace and Employee Survey includes both single-establishment and multi-establishment firms, and the centralization–innovation patterns are present in both. The organizational transformation described by the baseline model, standardization killing initiative, occurs whether or not the firm is literally opening new units.