BIS Working Papers
No 1367

The AI investment race

by Phurichai Rungcharoenkitkul

Monetary and Economic Department

July 2026

JEL classification: G01, G32, L13, O33

Keywords: artificial intelligence, investment, contest theory, circular financing, boom-bust cycle, financial fragility, network

BIS Working Papers are written by members of the Monetary and Economic Department of the Bank for International Settlements, and from time to time by other economists, and are published by the Bank. The papers are on subjects of topical interest and are technical in character. The views expressed in this publication are those of the authors and do not necessarily reflect the views of the BIS or its member central banks.

This publication is available on the BIS website (www.bis.org).

ISSN 1020-0959 (print)
ISSN 1682-7678 (online)

The AI Investment Race*

Phurichai Rungcharoenkitkul†

7 July 2026

Abstract

The AI build-out ranks among the largest technology-driven investment booms in US history. Its scale, reliance on debt and circular equity ties raise questions about the boom's sustainability and financial stability. We study a dynamic contest in which firms competing for a few dominant positions over-commit resources. The over-investment leaves the sector exposed to revenue disappointment that could turn boom into bust. The larger the boom, the deeper the eventual bust. The race to commit early through debt and circular financing also makes a bust more likely. Calibrated to balance sheet and deal data, the model points to over-investment of around 1.5 times the efficient level, rising to around three times where demand is less elastic. A network analysis shows that stress in one firm could cascade to others through chains of financial exposures.

JEL: G01, G32, L13, O33

Keywords: artificial intelligence, investment, contest theory, circular financing, financial fragility, boom-bust cycle, network

1 Introduction

Technological breakthroughs are typically accompanied by investment booms and buoyant macroeconomic activity. Exuberance about the promise of new technologies intensifies competition among firms eager to capture a share of the revenues. The race to get ahead can result in excessive investment that makes the boom unsustainable and prone to a disruptive

*I thank Denis Gorea, Benoit Mojon, Fernando Perez-Cruz, Daniel Rees, Andreas Schrimpf, Vatsala Shreeti, Frank Smets, Fabrizio Zampolli and BIS seminar participants for helpful comments. Errors are mine alone. Views expressed do not necessarily reflect those of the Bank for International Settlements.

†Bank for International Settlements. Email: phurichai.rungcharoenkitkul@bis.org

end. This fragility is further aggravated by the leverage that accompanies the rapid ramp-up of investment. This boom-bust pattern recurs across history, from the US canal mania in the 1830s and the British railway mania in the 1840s, to the roaring 20s and the dotcom boom in the late 90s. These episodes all ended in sharp corrections (Figure 1, panel (a)), with wider economic fallout.

The current AI build-out boom shares many of these characteristics. Total capital expenditure (capex) by the major hyperscalers has grown rapidly over the past few years and is set to exceed $700 billion in 2026 alone. Over the coming years, aggregate investment in AI infrastructures is widely expected to run into the trillions.[1] The potential demand for AI services is clearly vast and could justify a substantial expansion in computational power. Yet relative to its pre-boom trough, the current build-out is on track to outgrow every previous episode only three years in (Figure 1, panel (a)). To the extent that contest motives also shape investment decisions this time around, the tendency toward excessive commitment raises questions about the sustainability of the current boom.

The financing of the current AI boom also displays the financial vulnerabilities that characterised earlier boom-bust episodes. Although many hyperscalers hold large cash balances, the rapid capex expansion has led some to resort increasingly to borrowing, not only through bond issuance but also through special purpose vehicles and non-bank lenders. Some firms have also engaged in so-called circular financing, in which a firm engaging in the build-out takes an equity position in an AI lab in exchange for the lab's commitment to purchase future compute. Such arrangements have enabled greater deployment of capital, but they also create interconnectedness, shown in panel (b) of Figure 1. Rising leverage and more complex financing structures have raised questions about the financial stability risks of the current AI boom.

In this paper, we examine these real and financial forces driving the AI boom in a simple setup. We develop a dynamic contest model, in which coalitions of AI firms compete with each other in a winner-take-most environment.[2] Aggregate investment determines total AI output, but most of the reward goes to a small number of contest winners rather than being shared proportionately. As a result, each coalition invests heavily to outcompete its rivals and improve its chance of capturing the reward. This contest motive creates externalities as each coalition does not internalise the effect of its investment on others' returns. The sector as a whole becomes prone to excessive investment relative to the socially optimal level.[3]

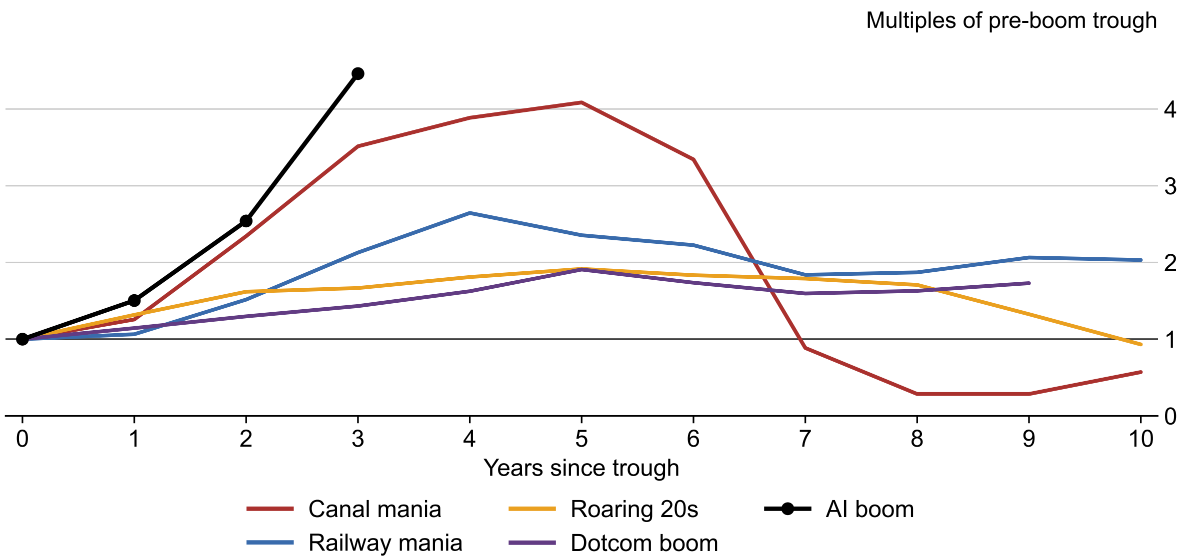
[1] See Huang (2025) and Van Nieuwerburgh (2026).

[2] Each coalition consists of a hyperscaler and a lab. This abstraction allows us to focus on the build-out competition that drives aggregate investment. We later relax this abstraction.

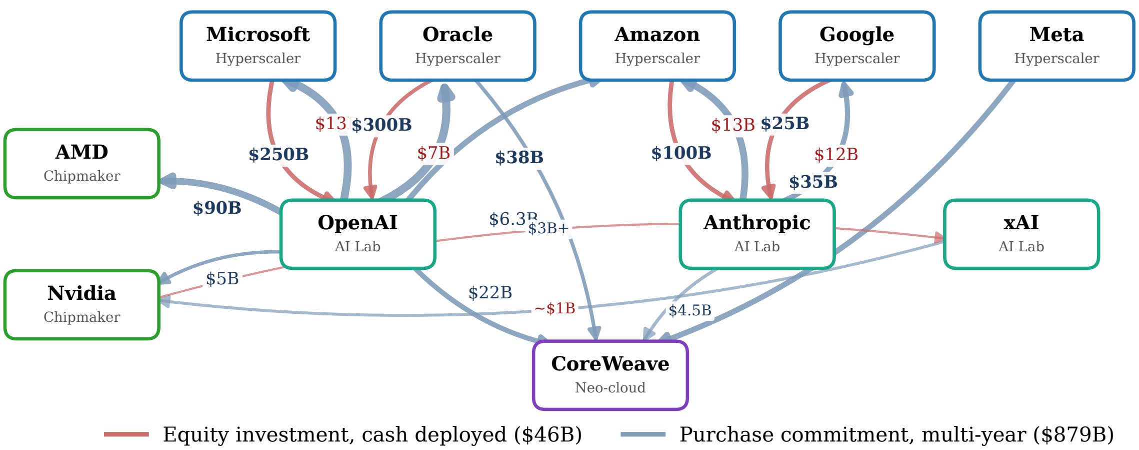
[3] These excesses raise questions about the sustainability of AI spending, the reversal of which could disrupt global growth momentum. See BIS (2026) on the macroeconomic implications of AI exuberance.

Figure 1: AI boom and its financing

(a) AI investment boom and historical episodes



(b) AI's circular financing structure



Note: Panel (a) shows investment during five innovation-led booms as multiples of each episode's pre-boom trough: canals (US, from 1835), railways (GB, from 1843), the 1920s (US, from 1921), dotcom (US, from 1995) and AI hyperscaler capex (from 2023, with 2026 from announced guidance). Panel (b) shows the network of equity stakes and compute-purchase commitments linking hyperscalers, chipmakers, AI labs and the leading neocloud, with thicker edges for larger deals. *Sources:* SEC filings (10-K, 10-Q, 8-K), company announcements, earnings-call guidance, and financial press.

We then study how this contest interacts with financial fragility. Coalitions fund the build-out not only from their own resources but also through leverage and circular stakes in their partner firms. The setup yields the paper's central result: an AI boom creates fragility that undermines itself. The more capacity the sector builds, the higher the productivity bar it must clear to sustain the boom, so a larger boom is both more likely to disappoint and more damaging when it does. Unlike the usual financial accelerator which amplifies an exogenous shock to balance sheets, the vulnerability here is endogenously generated by the boom itself. The financing structure is also central to the mechanism, with debt and circular stakes both propelling the boom and exacerbating the losses during busts. The debt-financed capital is specialised, so a synchronised retrenchment forces it into a thin resale market and the recovery rate falls the more the sector as a whole has borrowed, deepening the bust. Firms foresee this fire-sale when they choose their leverage and pull back correspondingly, yet the contest motive still leaves them over-building. The equity stakes behind circular financing add to that fragility rather than diversifying it away. Real and financial exuberance are thus bound together: the competition that over-builds the boom is also what selects the fragile financing that turns it into a bust.

To gauge the magnitudes involved, we calibrate the model to AI firms' balance sheets and to the public deal records. The concentration of the sector gives unusual, if incomplete, visibility into firms' financial positions, lending discipline to the calibration. The baseline results are plausible and material. The contest externality alone delivers over-investment of around 1.5 times the social optimum, and several times higher where demand for AI services is less elastic. Anticipating the fire sale lowers borrowing somewhat, but the contest motive still leaves over-investment around 1.4 times. The financing structure then creates vulnerabilities that can reduce net economic surplus (total output minus cost of capital) by more than half relative to the boom. Once the build-out is sufficiently large, which we estimate at around \$3 tn, net economic surplus turns negative in expectation. Industry projections of \$3–4 tn overall capex over the coming years thus place the sector in the region of significant downside risks.

Finally, we relax the coalition abstraction and consider the network structure with bilateral financial linkages between hyperscalers, AI labs, chipmakers and neoclouds. A new vulnerability emerges, as a shock to one firm could propagate and become systemic. Circular financing is again central. Because a hyperscaler holds equity in the labs it backs, the failure of one lab could hit the hyperscaler's balance sheet, forcing a retrenchment from another lab, distributing the stress further. The reach of such a cascade widens as the build-out grows and as funding concentrates on the backers that bridge several labs.

Related literature. The paper draws on and contributes to three literatures. The first is the theory of contests and innovation races, in which firms competing for a prize over-invest relative to the social optimum (Tullock, 1980; Loury, 1979; Konrad, 2009), with the patent-race tradition (Reinganum, 1989; Gilbert and Newbery, 1982; Aghion and Howitt, 1992; Akcigit and Kerr, 2018) the closest analytical antecedent. Our over-investment operates on a margin distinct from the classic business-stealing result of Mankiw and Whinston (1986). There, free entry is excessive on the extensive margin: too many firms enter the market. Here the set of competitors is fixed and the distortion is intensive, as each incumbent over-builds capacity because part of the return to a marginal dollar is share taken from rivals rather than output created. We embed this contest in concave aggregate production with bounded sectoral demand, which disciplines the size of the wedge, and we add a dynamic timing dimension. The novelty of our work relative to this strand is that the over-investment is self-undermining. The productivity a build-out must reach to repay itself rises with its scale, so the contest does not merely inflate investment but co-determines the boom and the bust.

The second literature is macro-finance and the amplification of shocks through balance sheets. We build on the leverage and collateral cycles of Brunnermeier and Pedersen (2009), Geanakoplos (2010) and Adrian and Boyarchenko (2015), the net-worth feedback of Kiyotaki and Moore (1997), Bernanke and Gertler (1989) and Gertler and Kiyotaki (2010), the fire-sale and intangible-capital channels of Lorenzoni (2008), Bianchi (2011) and Crouzet and Eberly (2021), the complexity-driven feedback of Caballero and Simsek (2013), and the financial-network models (Allen and Gale, 2000; Elliott et al., 2014; Acemoglu et al., 2015; Allen, Babus, and Carletti, 2012; Cabrales, Gottardi, and Vega-Redondo, 2017). Our mechanism departs from this work in two ways. First, the bust is endogenous to the boom rather than the consequence of an exogenous shock to net worth or collateral, because the productivity threshold for averting collapse rises with the contest-driven investment itself. Second, the shock propagates through a network of equity cross-holdings, so that a single firm's failure reaches others in the network, a channel that bilateral and own-balance-sheet models do not generate.

The third literature is the economics of the AI boom itself. Closest to our focus, Caballero (2026) interprets the build-out as a speculative-growth episode, in which coordinated optimism about the technology sustains a boom, which is fragile to uncertain payoff. Wachter and Wachter (2026) combine a bottom-up reconstruction of hyperscaler capital expenditure with a theory of rare productivity booms, inferring the observed investment as a revealed bet on a large advance in AI productivity and tracing its implications for growth and asset prices. We share their diagnosis of the boom's fragility and advance a complementary hypothesis grounded on contest externalities and their interaction with financial fragility. A contest for dominance that drives the build-out also generates fire-sale destruction in the bust, creating an endogenous cycle. One advantage our approach affords is the ability to quantify over-investment and fragility, through observable financing structure and build-out, which we exploit to calibrate the model. The rest of the literature examines the AI boom from varied perspectives. Acemoglu (2024) questions the aggregate productivity gains of AI and warns of over-investment, Korinek and Vipra (2025) studies the concentration that scaling laws imply, and Eisfeldt et al. (2023) examines AI's asset-pricing implications. On the descriptive side, Van Nieuwerburgh (2026) discusses the contemporaneous macro picture that we formalise in a model. Aldasoro et al. (2026) and Eren et al. (2026) document the debt-issuance, off-balance-sheet and bank-exposure dynamics that our financing block takes as its empirical counterpart. Finally, the capital-structure-and-product-market interaction (Brander and Lewis, 1986; Bolton and Scharfstein, 1990) and the moral-hazard foundation for debt capacity (Holmström and Tirole, 1997) provide the financial-frictions building blocks.

Outline. Section 2 sets out the dynamic contest model. Section 3 derives the analytical results under the symmetric specification. Section 4 calibrates the symmetric model to current AI coalitions and reports the headline magnitudes. Section 5 extends the model to the bipartite financing network and shows how a single lab's failure cascades across firms. Section 6 concludes.

2 Model

The model has two building blocks: (i) a physical environment governing investment and the contest for sector revenue, and (ii) a financing structure describing how investment is funded. In this section, we describe each in turn, specify the optimisation problems faced by coalitions and a benevolent planner, and state the equilibrium concept.

2.1 Environment

There are n identical coalitions, each consisting of a hyperscaler and an AI laboratory, treated as a joint decision-maker. This abstraction isolates the contest and financing externalities from the effects of a detailed ownership network and asymmetric leverage, which we return to in Section 5.

There are three periods. In period 1, each coalition k chooses an investment level $I_{1,k}$, financed by a combination of equity and debt. Because the contest rewards early scale, this period-1 commitment is strategically decisive and, once sunk, is exposed to the later productivity draw. In period 2, the AI productivity $\rho \in [0, 2]$ is drawn from a symmetric beta distribution, Beta(2,2) with mean of 1. $\rho > 1$ captures faster AI adoption and breakthrough capabilities, while $\rho < 1$ represents disappointments. After observing the productivity, the coalitions decide how much to invest in the second period, $I_{2,k}$. In period 3, the contest resolves, debts in periods 1 and 2 are repaid at gross rate R_d[4], and the aggregate AI payoff is determined and distributed to winning coalitions' equity holders.

Coalition k's investments over the two periods dictate its effective "compute":

$$Q_k(\rho) \;=\; \big(I_{1,k} + \theta\, I_{2,k}(\rho)\big)^{\nu}\,(1 + \psi\gamma_k) \tag{1}$$

$\theta \in (0, 1)$ discounts second-period capacity, which is deployed later and has less time to accumulate training time before the contest resolves. Capacity deployed early thus enjoys an *early-deploy premium.* $\nu \in (0, 1]$ reflects supply friction (GPU rationing, power and cooling bottlenecks). $(1 + \psi\gamma_k)$ represents the strategic synergy within the coalition. $\gamma_k \in [0, 1]$ is the share of 'circular financing', taking the form of the hyperscaler's equity stake in the lab in exchange for the lab's compute-purchase commitments. The hyperscaler wants to lock in demand for its build-out, but cannot enforce the lab's promise to purchase compute. An equity stake aligns the two sides' interests, making the commitment credible. More circular financing thus yields a productivity benefit from closer partnership at rate ψ, microfounded as the deadweight loss avoided by resolving this bilateral hold-up (see Appendix B.1). Equation (1) characterises Q_k under partnership continuation; if the partnership is disrupted (a condition derived in Proposition 4), partnership- and debt-financed capital are partially destroyed through asset specificity, with surviving fractions governed by parameters introduced in Section 2.2.

The aggregate AI sector revenue depends on total effective compute from all coalitions subject to a concave sector demand function:

$$\Theta(\rho) \;=\; \rho\, A\, (J(\rho))^{\beta}, \qquad J(\rho) \;=\; \sum_{k=1}^{n} Q_k(\rho). \tag{2}$$

The exponent $\beta < 1$ captures diminishing returns at the sector level, arising from demand saturation for AI services, complementary input bottlenecks and substitution with alternative technologies.

The contest awards $K \leq n$ winning slots through K independent Tullock lotteries with

[4] For simplicity, we ignore maturities and assume the same gross interest rate for all debts. This makes period-1 debt cheaper, favouring early commitment. The headline over-investment result is unaffected, however, and the effect on the bust probability is small under our calibration, as shown in Appendix C.

replacement. In each lottery coalition k wins with probability $I_{1,k}^{\varepsilon} Q_k(\rho)/\sum_j I_{1,j}^{\varepsilon} Q_j(\rho)$. The same coalition can win multiple slots, and each slot pays $\Theta(\rho)/K$. The expected number of slots coalition k wins is therefore

$$\pi_k(\rho) = K \frac{I_{1,k}^{\varepsilon} Q_k(\rho)}{\sum_j I_{1,j}^{\varepsilon} Q_j(\rho)} \tag{3}$$

and its expected gross contest payoff is $\pi_k(\rho)\Theta(\rho)/K = I_{1,k}^{\varepsilon} Q_k(\rho)\Theta(\rho)/\sum_j I_{1,j}^{\varepsilon} Q_j(\rho)$. The parameter K cancels in expected payoffs and only represents the number of winners that share contest revenue in any given draw. The exponent $\varepsilon \geq 0$ is the *contest head-start* elasticity: committed period-1 capacity captures contest share that the same capacity deployed later cannot replicate. Reaching the frontier first locks in users, builds the usage-and-data advantage that keeps a model ahead, and draws in the developers and researchers who compound the lead, so a latecomer with equal compute arrives to a market already taken. Appendix B.2 provides one microfoundation through a discrete-choice demand system. Setting $\varepsilon = 0$ nests the simple Tullock contest $\pi_k = KQ_k/J$, while $\varepsilon > 0$ introduces the head-start advantage. The contest is a lottery-with-replacement variant of the top-K contest of Clark and Riis (1996).

2.2 Financing structure

Each coalition holds cash $W_k > 0$ which is divided into internal equity $E_k = (1 - \gamma_k)W_k$ and circular financing $C_k = \gamma_k W_k$. Internal equity is deployed directly at opportunity cost r_f and is not subject to withdrawal. Circular financing is bundled with a compute-purchase commitment, carries the same nominal cost of capital, provides the strategic benefit ψ, and is fragile in the sense that partnership-financed capital is partially impaired when ρ falls below the bust threshold. The extent of circular financing $\gamma_k \in [0, 1]$ is a coalition choice variable. Coalition k may borrow $B_{t,k}$ in period $t \in \{1, 2\}$, with debts repaid at the end of period 3 from the realised contest payoff at gross cost R_d.

Of the circular budget C_k, the coalition commits $c_{1,k} \in [0, C_k]$ to deployment in period 1, holding the remainder for period 2. In Proposition 4, we show that if productivity is above a certain bust threshold, $\rho \geq \bar{\rho}$, deployment resumes, drawing on the retained budget $c_{2,k}(\rho) \leq C_k - c_{1,k}$ and on period-2 borrowing $B_{2,k}(\rho)$ against mark-to-market collateral, up to the optimum that rises with the realised draw. When $\rho < \bar{\rho}$, the remainder is held as cash and no borrowing takes place in period 2. Investment in each period is then

$$I_{1,k} = E_k + c_{1,k} + B_{1,k} \tag{4}$$

$$I_{2,k}(\rho) = c_{2,k}(\rho) + B_{2,k}(\rho) \tag{5}$$

Debt capacity in each period is bounded by pledgeable income, following Holmström and Tirole (1997): the lender requires the coalition to retain enough of its payoff to internalise effort costs. The period-1 bound is taken in expectation over the productivity draw and the period-2 bound is marked to market against the realised ρ,

$$B_{1,k} \leq \frac{\mu \mathbb{E}\big[\pi_k \Theta(\rho)/K\big]}{R_d}, \tag{6}$$

$$B_{2,k}(\rho) \leq \frac{\mu \pi_k \Theta(\rho)/K}{R_d}, \tag{7}$$

where $\mu \in (0,1)$ is the pledgeable fraction.[5] A moral-hazard microfoundation appears in Appendix B.3.

When the realised ρ falls below an endogenous threshold $\bar{\rho}$, period-2 deployment corners at zero and period-1 capital is subject to impairment. Of period-1 circular deployment $c_{1,k}$, only fraction $\alpha \in (0,1)$ survives as a productive investment claim; the complementary fraction $1-\alpha$ is destroyed through partnership unwind (a non-renewal of the compute-purchase commitment). Of period-1 debt-financed capital $B_{1,k}$, a fraction is lost to a fire-sale of specialised infrastructure. The fire-sale is an industry equilibrium: in a common bust all n coalitions liquidate at once into a market whose natural buyers, the other hyperscalers, are constrained at the same time, so the recovery falls with the quantity sold. A downward-sloping demand for the specialised capital then yields a loss rate that is linear in the sector's debt-financed stock,

$$\zeta_{\text{eff}} \;=\; \min\big\{\, \zeta + \delta\, B_1,\; \zeta_{\max} \,\big\}, \tag{8}$$

so the bust destroys $\zeta_{\text{eff}}\, B_{1,k}$ of coalition k's debt-financed capital. Here B_1 is the sector's (symmetric) debt-financed stock per coalition, ζ is the volume-independent specificity loss, $\delta = c\, n$ combines the price-impact slope c and the buyer concentration n, and $\zeta_{\max}$ caps the loss at the scrap floor. Each coalition takes the sector loss rate as given, so the equilibrium rate is the fixed point $\zeta_{\text{eff}} = \zeta + \delta B_1$ (Appendix B.4 derives (8) from a cash-in-the-market clearing condition). The circular-survival parameter α remains a reduced-form summary of

[5] Circular flows between hyperscalers and labs could generate book revenues mistaken for genuine income due to their opacity, inflating the collateral behind period-2 borrowing. The main analysis abstracts from this book-inflation channel, which we leave to future work.

within-coalition heterogeneity in distress outcomes.[6] [7]

2.3 Problems and equilibrium

The coalition's problem is to choose period-1 actions $(\gamma_k, c_{1,k}, B_{1,k})$, subject to the period-1 debt-capacity constraint (6), taking rivals' choices as given and anticipating that the period-2 actions $(c_{2,k}(\rho), B_{2,k}(\rho))$ will optimise against the realised ρ subject to the period-2 collateral constraint (7). The coalition's expected payoff is

$$V_k^{coal} \;=\; \mathbb{E}\left[\pi_k \,\frac{\Theta(\rho)}{K}\right] \;-\; r_f\, W_k \;-\; R_d\left(B_{1,k} + \mathbb{E}[B_{2,k}(\rho)]\right) \;-\; \tfrac{\kappa}{2}\,\gamma_k^2\, W_k \;-\; \tfrac{\phi}{2}\left(\tfrac{c_{1,k}}{C_k}\right)^2 C_k. \quad (9)$$

The first term is the expected contest revenue, the second the opportunity cost of cash, the third expected debt service (period-2 borrowing is positive only when $\rho \geq \bar{\rho}$ by the optimisation derived below), and the last two convex friction terms in the circular share and the deployment ratio. The convex-cost parameters $\kappa, \phi > 0$ capture, respectively, the rising governance cost of scaling circular financing and the adjustment cost of front-loading deployment. The coalition maximises (9) subject to the budget identities (4)–(5), the debt-capacity constraints (6)–(7), and the circular cap $0 \leq c_{1,k} \leq C_k$, $0 \leq c_{2,k}(\rho) \leq C_k - c_{1,k}$.

In period 2 the coalition observes the productivity draw ρ and decides how much further to build. When the news is good $(\rho \geq \bar{\rho})$, expansion pays: the coalition deploys additional capacity, drawing on its retained circular budget $c_{2,k}(\rho)$ and fresh borrowing $B_{2,k}(\rho)$, until either the marginal return is driven down to the cost of capital R_d or it exhausts the collateral limit (7), whichever binds first. When the news is bad $(\rho < \bar{\rho})$, expansion does not pay even at the first dollar, so the coalition holds back from deploying and borrowing. The threshold $\bar{\rho}$ separating the two regimes rises with the scale of the period-1 build, which we characterise in Proposition 4.

Definition 1 (Symmetric Nash equilibrium)**.** *A symmetric pure-strategy Nash equilibrium is a common action profile* $(\gamma^*, c_1^*, B_1^*, \{c_2^*(\rho), B_2^*(\rho)\}_{\rho\in[0,2]})$ *such that each coalition k, taking*

[6] α follows from an equity-side default boundary in the spirit of Leland (1994), where partnership continuation switches off once productivity falls below an endogenous threshold; the industry fire-sale loss rate ζ_{eff} follows Shleifer and Vishny (1992): specialised assets sold when their natural buyers are constrained recover below use-value. We distinguish this discount from secular GPU obsolescence, a separate, additive force. Note that part of the debt-financed stock (power and shell capital) recovers closer to par, so the loss base $\zeta_{\text{eff}} B_1$ is arguably conservative. Related collateral-cycle channels appear in Brunnermeier and Pedersen (2009) and Bianchi (2011).

[7] The co-impairment of E, c_1 and B at the coalition level does not violate seniority; it reflects aggregation across sub-units (lab partnerships, AI-investment SPVs), each with its own standard waterfall. Coalition-level α, ζ are weighted averages over sub-unit outcomes; $\zeta > 0$ because equity-poor sub-units default with debt impairment before parent-level equity is exhausted.

the other coalitions' strategies as given, maximises V_k^{coal} *in* (9) *subject to the budget identities* (4)–(5)*, the financing constraints* (6)–(7)*, and the deployment bounds* $0 \le c_{1,k} \le C_k$, $0 \le c_{2,k}(\rho) \le C_k - c_{1,k}$, $B_{1,k}, B_{2,k}(\rho) \ge 0$*, where, for each realised* ρ*, the period-2 actions* $(c_{2,k}(\rho), B_{2,k}(\rho))$ *maximise* $\pi_k\,\Theta(\rho)/K - R_d\,(c_{2,k}(\rho) + B_{2,k}(\rho))$ *given* $(\gamma_k, c_{1,k}, B_{1,k})$.

A symmetric equilibrium exists and is unique within the symmetric class.[8]

The planner differs from the coalitions in that it internalises the business-stealing contest. While each coalition races to capture sector revenue treating its rivals' actions as fixed, the planner recognises that the race is zero-sum, so the compute one coalition adds comes partly at the expense of the others. Reining it in lowers the over-investment and, with it, the fragility of the investment boom. Facing the same real frictions κ and ϕ and commanding the full instrument set $(\gamma, c_1/C, B_1)$, the planner's allocation is the benchmark against which we measure the equilibrium's distortions.

Definition 2 (Social planner)**.** *A benevolent social planner chooses symmetric coalition actions* $(\gamma, c_1/C, B_1)$ *to maximise the aggregate payoff*

$$V^{plan} = \mathbb{E}\big[\rho A\, J(\rho)^{\beta}\big] \;-\; n\Big[r_f W + R_d\big(B_1 + \mathbb{E}[B_2(\rho)]\big) + \tfrac{\kappa}{2}\gamma^2 W + \tfrac{\phi}{2}(c_1/C)^2 C\Big], \qquad (10)$$

treating $J(\rho) = nQ(\rho)$ *as endogenous, subject to the constraints of Definition 1.*

3 Analytical results

This section characterises the symmetric equilibrium in closed form and derives the main analytical results. We show how the race for AI dominance leads to an investment boom that carries the seeds of its own bust. The contest leads firms to build more compute than is collectively worthwhile (Proposition 1). The distortion also reaches the financing of the buildout, as firms lean excessively on circular financing (Proposition 2) and rush to commit early to seize the head start (Proposition 3). The larger the resulting boom, the more vulnerable the system is to a bust.

The excesses of the boom also shape the character of the bust. The contest wedge that inflates the boom determines the severity of the bust, through leverage and fire-sale

[8] The period-2 problem is a concave deployment choice with a single break-even threshold $\bar{\rho}$. Because the productivity density is continuous, the regime switch is continuous and the period-1 payoff V^{coal} is continuously differentiable in (γ, c_1, B_1). The convex frictions κ and ϕ dominate the revenue curvature at the calibration, so V^{coal} is concave and each best response is single-valued and continuous; a symmetric fixed point then exists by Brouwer's theorem. The equilibrium fire-sale loss rate solves $\bar{\zeta} = \zeta + \delta B_1(\bar{\zeta})$; the best-response B_1 is bounded and decreasing in the rate while the right side is increasing in B_1 and capped at $\zeta_{\max}$, so the fixed point exists and is unique. Uniqueness within the symmetric class follows from single-crossing of the first-order conditions in (γ, c_1, B_1). We do not rule out asymmetric equilibria.

amplification (Propositions 4 and 5). The circularity of the financing dictates where the bust concentrates (Proposition 6). The race to commit capital early, rather than the contest wedge, pins the frequency of the bust (Corollary 1). A boom that is "too much, too circular and too early" translates into a bust that is "too deep, too lop-sided and too frequent".

3.1 Over-investment in compute

The first result establishes the contest's level distortion in isolation, abstracting from financing frictions ($\gamma = 0$, no collateral constraint) and the head-start advantage that dictates investment timing. At the symmetric equilibrium, the marginal benefit of one unit of coalition compute is, for the coalition,

$$MR_{coal}(\rho) \;=\; \frac{\partial}{\partial Q_k}\left[\frac{Q_k}{J(\rho)}\,\Theta(\rho)\right]_{\text{sym}} \;=\; \frac{n-1+\beta}{n}\cdot\frac{\Theta(\rho)}{J(\rho)}, \tag{11}$$

and, for the planner,

$$MR_{plan}(\rho) \;=\; \frac{\partial\Theta(\rho)}{\partial Q_k}\bigg|_{\text{sym}} \;=\; \beta\frac{\Theta(\rho)}{J(\rho)}, \tag{12}$$

the social marginal return, which internalises that all n coalitions act symmetrically. Their ratio is the contest wedge,

$$\lambda \;\equiv\; \frac{MR_{coal}(\rho)}{MR_{plan}(\rho)} \;=\; \frac{n-1+\beta}{n\,\beta}, \tag{13}$$

strictly above unity for $n \geq 2$ and $\beta < 1$, and independent of ρ (the $\Theta(\rho)/J(\rho)$ factor cancels). In the unsaturated-demand limit $\beta \to 1$ the wedge collapses to unity: marginal compute is freely scalable, the share-stealing margin disappears, and the contest externality vanishes. The single parameter λ shapes every result that follows, amplifying the private benefit of compute, circular financing, and period-1 deployment relative to the social value while leaving their costs unchanged. Exponentiated through the period-1 first-order condition, it delivers a closed-form over-investment ratio.

Proposition 1 (Over-investment)**.** *At the symmetric Nash equilibrium with $\gamma = 0$ and unlimited debt capacity, and no head start ($\varepsilon = 0$), aggregate effective compute exceeds the planner's level by*

$$\frac{J^*}{J^{plan}} \;=\; \lambda^{\nu/(1-\nu\beta)}, \tag{14}$$

and aggregate dollar investment by

$$\frac{I^*}{I^{plan}} \;=\; \lambda^{1/(1-\nu\beta)} \;>\; \frac{J^*}{J^{plan}}. \tag{15}$$

The contest externality vanishes as $\beta \to 1$.

Proof. See Appendix A.1 $\square$

A positive head start widens the gap, by adding a premium to the coalition's committed investment and none to the planner's, so the ratios above serve as floors. Section 3.3 takes up the head start.

The distortion behind this over-investment is the foundation for the results that follow. The contest wedge not only drives the scale of investment, but also tilts firms into circular financing and early deployment, thus shaping the composition and timing of the buildout and thereby its fragility, as we show next.

3.2 Excessive circularity

The second result characterises the choice of circular share. Each coalition sets γ to balance the benefit of circular financing against the convex structural cost $\frac{\kappa}{2}\gamma^2 W$ of running a larger share. Circular financing earns the strategic benefit ψ in the boom and is impaired in the bust, and netting the two across states gives the social marginal value of circular financing,

$$g_\gamma \equiv \mathbb{E}\left[MR_{plan}(\rho)\frac{\partial Q_k}{\partial \gamma}\right],$$

where $MR_{plan}(\rho) \equiv \partial\Theta(\rho)/\partial Q_k$ is the marginal social product of compute. Note that g_γ is a function of the share γ itself. A planner values circular financing at exactly this social rate.

A coalition values the same circular share more than the planner does. In the contest, a larger circular share lifts not only its own output, the margin relevant to the planner, but also its share of the prize, capturing revenue that would otherwise have gone to its rivals. This is the business-stealing force behind the level distortion of Section 3.1, now operating on the circular share. A planner nets it to zero, since one coalition's gain in share is another's loss, while a coalition keeps it in full. Its marginal benefit therefore exceeds g_γ, implying a higher circular share than under the planner.

Proposition 2 (Excessive circular financing)**.** *The optimal circular shares satisfy*

$$\gamma^*_{plan} = \left[\frac{g_\gamma}{\kappa W}\right]_0^1, \qquad \gamma^*_{coal} = \left[\frac{\lambda g_\gamma}{\kappa W}\right]_0^1$$

where $[x]_0^1 \equiv \min\{1, \max\{0, x\}\}$ and g_γ is evaluated at each party's own optimum. The coalition holds weakly more circular financing than the planner,

$$\gamma^*_{coal} \geq \gamma^*_{plan}. \tag{16}$$

Proof. See Appendix A.2 □

The inequality is weak because both parties may deploy zero circular financing. The excess turns strict only when $(\lambda - 1)g_\gamma > 0$, that is, when $g_\gamma > 0$ and the coalition's inflated valuation of circular financing exceeds the planner's. The contest wedge, which scales the social value g_γ by λ with the same sign, is the single deciding factor. When circular financing is socially worthwhile, the coalition wants more of it than the planner does. When it is socially destructive at the margin, $g_\gamma < 0$, the wedge runs the other way and neither employs any circular financing.

3.3 Excessive front-loading

The third result concerns the timing of investment, the split between what a coalition commits in period 1 and what it defers to period 2. The two dates differ in their exposure to the bust. Period-2 capital is deployed only after the productivity draw is known, so it is never sunk into a downturn. Yet committing early wins contest share, through the head-start $I^{\varepsilon}_{1,k}$ in (3). A coalition and a planner both weigh this exposure against the genuine returns to building early; the coalition alone also chases the head-start, and over-commits as a result.

Proposition 3 (Excessive front-loading)**.** *Fix γ and c_1/C. Then*

$$I_1^* \;\geq\; I_1^{plan},$$

with equality when $\varepsilon \leq \varepsilon^*$ and strict inequality when $\varepsilon > \varepsilon^*$. The threshold ε^* is negative when the boom-state early-deploy return plus the bust salvage covers the gross cost of debt, so that $I_1^* > I_1^{plan}$ for every $\varepsilon \geq 0$. When that return falls short, $\varepsilon^* > 0$, and a strictly positive contest head-start is required for excess front-loading.

Proof. See Appendix A.3 □

Proposition 3 shows that coalitions commit too early, not just too much: their capex is front-loaded beyond what a planner would choose. Three forces shape this timing. The technological early-deploy premium arising from θ is common to coalitions and the planner, and encourages both to front-load. At the same time, committing early sinks capital before the productivity draw, exposing it to a potential bust. Aversion to bust provides a motive to wait and defer investment. This motive is stronger for the planner, because even in a downturn a coalition's surviving capacity still earns it a share of the contest. The last force is the zero-sum contest head-start from ε, which affects only the coalition, since being early lifts its standing in the race.

3.4 Boom-bust dynamics

This section turns to coalitions' investment and financing decisions once the productivity shock is realised. In good states deployment resumes, drawing on the retained budget and on borrowing against marked-to-market collateral. In bad states period-2 deployment shuts down and front-loaded circular and debt-financed capital is destroyed. The boundary between the two regimes is endogenous, and it rises with the size of the period-1 boom: a larger boom widens the range of shocks that tip the sector into the bust. Thus, the boom sows the seed for the bust. We first establish the bust and its threshold (Proposition 4), and then trace how the three excesses of the boom shape its character: the depth of the losses (Proposition 5), the frequency of the event (Corollary 1), and where the destruction concentrates (Proposition 6).

Proposition 4 (Boom-bust dynamics)**.** *Period-2 deployment breaks even at the threshold*

$$\bar{\rho} = \frac{R_d \, I_1^{1-\nu\beta} \, n^{1-\beta}}{\lambda\theta\nu\beta \, A \, (1+\psi\gamma)^{\beta}}, \tag{17}$$

which is increasing in period-1 investment I_1 since $\nu\beta < 1$. Below $\bar{\rho}$ the period-2 marginal return falls short of R_d, so every coalition corners at $c_2 = B_2 = 0$ and $I_2 = 0$ (the constraint (7) *is then slack). The bust destroys $(1-\alpha)c_1$ of circular and $\zeta_{\text{eff}} B_1$ of debt-financed capital at the industry fire-sale rate* (8)*, leaving the surviving capital stock*

$$s \, I_1 = E + \alpha c_1 + (1 - \zeta_{\text{eff}}) B_1, \tag{18}$$

so output collapses to $Q_k = (s \, I_1)^{\nu}$.

Proof. See Appendix A.4. □

The depth of the bust described in Proposition 4 is endogenous to the boom. The fire-sale ties the loss rate to sector leverage, so the severity of the bust is itself shaped by how the boom was financed, and each coalition prices this into its borrowing.

Proposition 5 (Fire-sale fragility)**.** *Under the industry fire-sale* (8)*, the equilibrium loss rate is the fixed point $\zeta_{\text{eff}} = \zeta + \delta B_1$, increasing in the period-1 leverage. The coalition internalises this rate and chooses less debt than under a constant recovery. Nonetheless, fixing γ and c_1/C as in Proposition 3, it over-invests relative to the planner, $I_1^* > I_1^{\text{plan}}$, whenever the contest wedge $\lambda > 1$ and committed debt is positive ($\varepsilon > \varepsilon^*$). A larger period-1 investment carries both more debt-financed capital and a higher loss rate.*

Proof. See Appendix A.5. □

The price-taking coalition internalises the going recovery rate but not the fact that raising own leverage worsens the fire-sale for the whole sector. A planner who internalised this pecuniary externality, in the spirit of Lorenzoni (2008) and Bianchi (2011), would choose still lower leverage, so the over-investment in Proposition 5 is conservative. The fire-sale externality thus compounds the contest externality, though we do not characterise the constrained-efficient allocation here.

We next turn to bust frequency. A bust occurs when the productivity draw falls below the threshold $\bar{\rho}$ in (17). This threshold has two determinants, namely the capital committed before the draw and the value of deploying after the draw. Proposition 4 isolates the first, showing that more committed capital pushes up $\bar{\rho}$. Most of the forces that enlarge the boom, however, also raise the value of deploying in period 2, pulling the threshold down. For the contest wedge, productivity and the coalition synergy, the two movements offset. The race to commit early is the exception, as it rewards capital committed before the draw and moves the threshold in one direction only. The next result makes the decomposition exact.

Corollary 1 (Bust frequency)**.** *At an interior symmetric equilibrium, the probability of a bust satisfies*[9]

$$\Pr\big(\rho < \bar{\rho}^*\big) = \underbrace{1-\theta}_{\substack{\textit{early-deploy}\\ \textit{discount}}} + \frac{\theta}{R_d}\Bigg[\underbrace{\frac{\varepsilon(n-1)}{n}\,\frac{\mathbb{E}\big[\pi_k\Theta(\rho)/K\big]}{I_1}}_{\substack{\textit{head-start premium}\\ \textit{(race to commit early)}}} + \underbrace{(1-\zeta_{\text{eff}})\,\lambda\,\nu\beta A\,n^{\beta-1}(sI_1)^{\nu\beta-1}\,\mathbb{E}\big[\rho\,\mathbf{1}\{\rho<\bar{\rho}^*\}\big]}_{\substack{\textit{salvage on capital}\\ \textit{surviving the bust}}}\Bigg]. \tag{19}$$

The contest wedge λ enters only through the salvage term, the value of capital that survives a bust. To first order, it does not enter through the scale of the boom it creates. The contest's effect on the bust frequency is therefore led by the race to commit early, ε.

Proof. See Appendix A.6. □

The corollary interprets the bust probability as the price that coalitions pay to deploy early and enjoy a head start. Each term reflects how the marginal committed dollar is valued. The first constant term is the break-even condition when early capital pays off only in booms: coalitions commit until the expected return equals its cost, $(1-\Pr(\rho<\bar{\rho}^*))\,R_d/\theta = R_d$, giving $1-\theta$. The second term contains ε because the head start pays committed capital a

[9] Exact under the convention of Appendix A.3 that each coalition takes the unwind threshold as market-given. When the coalition instead internalises its own threshold, the marginal commitment tips marginal states into the bust and the right-hand side gains a boundary discount, derived in Appendix A.6.

contest share on top of its output. The third term is the salvage value: a fraction of the committed dollar survives, and the coalition prices this at its private contest margin, thus the presence of λ. This salvage cushion is small under destructive busts, as it is the product of the surviving fraction, the impaired marginal product and the weight on bad draws.

The fragility of the boom is therefore linked to the front-loaded commitment. Remove the head start that rewards front-loading, and both the associated debt and the elevated bust odds unwind. The contest wedge strengthens both the boom and the resolve to deploy through weaker draws, leaving a relatively limited effect on the bust frequency. Instead, the contest wedge determines how much a bust destroys, through the fire-sale losses $\zeta_{\text{eff}} B_1$ of Propositions 4 and 5, and where the destruction concentrates, through the financing mix, to which we turn next.

The composition dimension is where the contest's imprint is clearest. How capital is committed is itself subject to distortions. The contest draws coalitions into circular stakes (Proposition 2), of which a fraction $1 - \alpha$ is lost in the unwind, and into front-loaded debt (Proposition 3), priced to the fire-sale risk. The next result holds the scale of the boom fixed and shows that the contest also determines the financing mix.

Proposition 6 (The contest creates a fragile financing mix)**.** *Hold the scale of the boom fixed at $I_1 = \bar{I} > W$, along with the deployment ratio c_1/C, and let the circular share γ^* be interior. The bust loss $L^{\text{bust}} = (1-\alpha)c_1 + \zeta_{\text{eff}} B_1$ then strictly increases with the contest wedge,*

$$\left.\frac{\partial L^{\text{bust}}}{\partial \lambda}\right|_{\bar{I}} > 0 :$$

the wedge raises the circular share, and a higher circular share converts surviving equity into destructible capital. An economy that reaches the same scale $\bar{I}$ without the contest wedge busts less severely.

Proof. See Appendix A.7. □

The margin is the circular share rather than the timing of deployment. Shifting a dollar between circular and debt at a fixed level moves the loss by only $[(1-\alpha) - \zeta_{\text{eff}}]\gamma W$, negligible at the calibration (Appendix A.7).

This endogenous fragility is not a collateral spiral. Even with the borrowing limits slack during the boom, the bust still operates through concavity (a higher build needs a stronger draw), composition (more circular and front-loaded capital being exposed) and the fire-sale depth (liquidation grows with debt). Amplification from a binding financing constraint requires heterogeneity and an idiosyncratic shock, which Section 5 develops. Keeping the symmetric model on the slack side of the constraint isolates the contest and the fire-sale cleanly.

3.5 How the distortions compound

The contest and circular channels raise the coalition's desired investment above the planner's. Combining the period-1 first-order condition of Proposition 1 with the circular boost, the over-investment in the desired level factors into

$$\frac{I^*}{I^{plan}} = \underbrace{\lambda^{1/(1-\nu\beta)}}_{\text{contest wedge}} \cdot \underbrace{(1+\psi\gamma)^{\beta/(1-\nu\beta)}}_{\text{circular boost}}, \tag{20}$$

with planner benchmark

$$I^{plan} = \left(\frac{\beta\nu A\, n^{\beta-1}}{R_d}\right)^{1/(1-\nu\beta)}. \tag{21}$$

Each factor exceeds unity when its channel is active, and they enter multiplicatively, so the channels reinforce one another. The exponents translate wedges at the margin into gaps in levels. The marginal value of investment declines with scale as $I^{\nu\beta-1}$, and investment expands until this marginal value meets the cost of capital R_d. Inverting that condition, any wedge in the margin enters the level raised to the power $1/(1-\nu\beta)$. With $\nu\beta < 1$, the levels respond more than proportionately to the wedges.

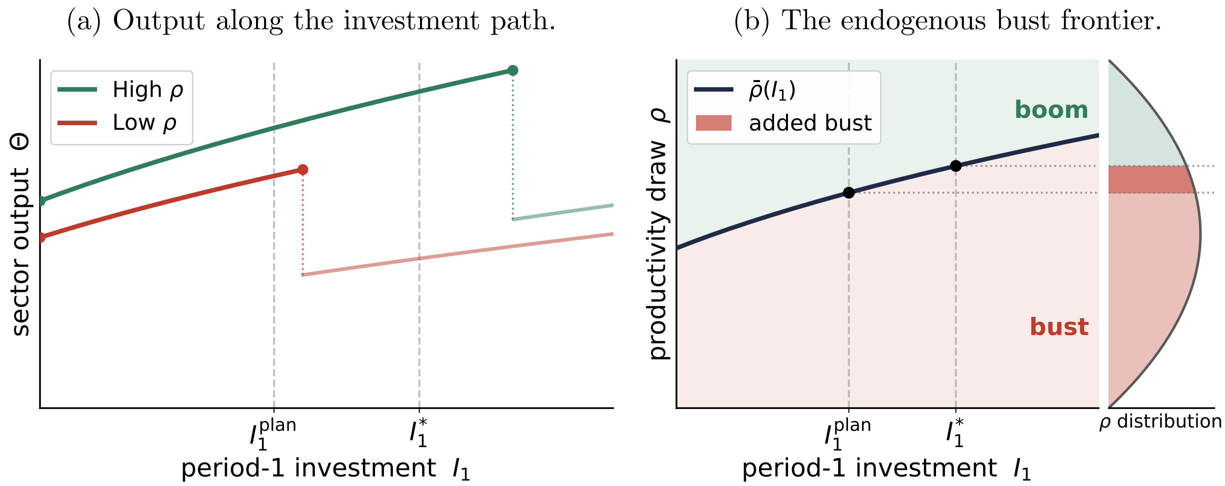
The two distortions compound. The contest sets the scale of the boom, and the head-start of Proposition 3 sets its timing, pulling the enlarged investment forward into committed period-1 debt that the planner does not take on. The coalition's boom is therefore both larger than the planner's and more heavily front-loaded into capital the bust can destroy.

The second period inherits the period-1 boom, but its course depends on the realised ρ,

$$I_2^*(\rho) = \begin{cases} 0, & \rho < \bar{\rho}, \\ \dfrac{1}{\theta}\left(\left[\dfrac{\lambda\theta\nu\beta\, A\, n^{\beta-1}(1+\psi\gamma)^{\beta}\,\rho}{R_d}\right]^{1/(1-\nu\beta)} - I_1\right), & \rho \geq \bar{\rho}, \end{cases} \tag{22}$$

which is continuous at the threshold, where the bracket equals I_1 and $I_2^* = 0$. Above the threshold the coalition draws on its retained budget and borrows against marked-to-market collateral, expanding period-2 capital until the marginal return falls to the cost of capital R_d; deployment rises with the realised ρ while the collateral constraint stays slack. Below the threshold the marginal return fails to cover R_d at any positive scale, deployment stops, and the period-1 stock is impaired as circular and debt-financed capital unwind. These two regimes shape the path of output, as shown in Panel (a) of Figure 2. An investment boom in period 1 exposes output to a large fall under a bad productivity draw in period 2, which would have been preventable in the absence of a large boom. Indeed, because the planner sets a lower period-1 investment $I_1^{\text{plan}} < I_1^*$, free from the contest and financing wedges, it averts the bust for a range of bad draws that would have afflicted the coalition.

Figure 2: Investment, the bust threshold, and output



Note: Panel (a) plots sector output Θ against period-1 investment I_1 for a favourable draw (High ρ, green) and an adverse draw (Low ρ, red). Each draw follows the boom branch until I_1 reaches the level at which its realisation falls short of the bust threshold, where the period-2 financing collapse forces output down to the lower branch. Panel (b) places the same mechanism in (I_1, ρ) space, where the bust frontier $\bar{\rho}(I_1)$ rises with I_1. A draw that survives at the planner level I_1^{plan} can fall below the frontier at the higher equilibrium level I_1^*. The shaded band in the ρ distribution is the probability mass converted from boom to bust by this over-investment, holding the deployment rule fixed at the coalition's frontier. Both panels share the horizontal axis and the markers I_1^{plan} and I_1^*.

The bust is avoidable because the threshold is not fixed. It rises with period-1 investment, with elasticity $1-\nu\beta$ (equation (17)), so a larger boom must be supported by a stronger draw. Panel (b) of Figure 2 shows this upward-sloping frontier $\bar{\rho}(I_1)$, the dividing line between the boom and bust regimes. A larger boom raises fragility along two dimensions. A higher threshold raises the probability of a bust, given by $\bar{\rho}^2(3-\bar{\rho})/4$ under the beta distribution $\rho \sim \text{Beta}(2,2)$ on $[0,2]$. The coalition's threshold $\bar{\rho}^*$ exceeds the planner's $\bar{\rho}^{\text{plan}}$, raising the bust probability by the dark area in Panel (b). This comparison holds the deployment rule fixed at the coalition's frontier. Under each agent's own rule the frequency effect nets out (Corollary 1), and only the second dimension operates. Furthermore, a larger period-1 stock leaves more capital vulnerable when the bust arrives, so the loss conditional on a bust is larger as well. The expected loss thus rises more than proportionately with the boom. A larger boom sows the seed for a more disruptive subsequent bust.

4 Quantitative analysis

This section calibrates the model and quantifies the symmetric equilibrium. It establishes the magnitude of the over-investment, the asymmetry of the bust, and the welfare cost of the contest.

4.1 Calibration

We calibrate the model to the deal data and the observed equilibrium. Two parameters come straight from the deal data and anchor the calibration: the equity-weighted circular share ($\gamma = 0.20$) and the period-1 deployment rate ($c_1/C = 0.71$, the share of committed compute that coalitions draw down within eighteen months of signing). Each coalition optimises its committed period-1 borrowing B_1, and two free parameters are pinned to features of the equilibrium: the productivity scale A to period-1 sector capex and the deployment friction ϕ so that $c_1/C = 0.71$ is optimal. The circularity-cost convexity κ is normalised so the constrained planner makes no use of circular financing ($\gamma_{\text{plan}} \to 0$).[10] At the optimised borrowing the marginal return on period-1 investment equals the gross cost of debt ($MR_{B_1} = R_d$), so the equilibrium leverage is privately optimal. Table 1 summarises all values and their rationales. We discuss the real-economy and financial parameters in turn.

The real-economy parameters govern the contest, the technology, and the bust. There are $n = 5$ frontier coalitions, corresponding to the number of major hyperscalers (Amazon, Google, Meta, Microsoft, Oracle) and labs (Anthropic, DeepMind, MetaAI, OpenAI, xAI). The production concavity is the cost share of compute equipment in data-centre construction; industry cost decompositions put that share between 0.6 and 0.8, and we take the central value of $\nu = 0.7$.[11] The early-deploy premium is anchored to frontier-accelerator depreciation: a one-year deploy lag costs roughly a sixth of the five-to-six-year useful lives hyperscalers report, putting late capacity at about 85% of an early unit, justifying $\theta = 0.85$.[12] The contest head-start $\varepsilon = 0.20$ is set in the low-middle of the range that market-share data bracket. Between 2023 and 2025 the enterprise segment flipped toward the smaller early committer while the consumer segment continued to reward incumbency heavily, so the implied head-start elasticity runs from near zero in enterprise to well above one in consumer markets (Appendix D). At $\varepsilon = 0.20$ the model delivers a period-1 debt share of one third,

[10] Circular financing has a social benefit as it resolves the bilateral hold-up between hyperscaler and lab. As a normalisation, we set κ so that the convex cost just offsets this benefit. Fixing the planner's benchmark at zero simply attributes the entire observed share to the contest, since excess circularity obtains regardless.

[11] Epoch AI's one-gigawatt study attributes 60 per cent of total cost to servers. TrendForce puts the server share of capital cost near 60 per cent in a reference build. IoT Analytics reports 78 per cent of 2024 data-centre spending on IT equipment against 12 per cent on facility infrastructure.

[12] Appendix C reports the headline magnitudes across $\theta \in [0.70, 0.90]$.

inside the range that issuance and lease disclosures imply, a moment the calibration does not target.[13] The strategic benefit $\psi = 0.30$ prices the hold-up value of circular financing, which Appendix B.1 suggests lies between 0.20 and 0.40 across the disclosed range of customisation costs.

The sector returns to scale $\beta = 0.80$ is a deliberately conservative upper bound. In the model, the revenue growth decomposes as $\Delta \ln \Theta = \Delta \ln(\rho A) + \beta \Delta \ln J$, namely demand shifts plus β times usage growth. Tokens processed grew between seven- and twenty-fold per year over 2024–26, while AI revenues grew three- to four-fold. Setting demand shifts, which have, if anything, increased over this period, to zero attributes all revenue growth to usage and returns the largest β consistent with the data, $\beta \leq \Delta \ln \Theta / \Delta \ln J$. Under this most generous assumption, combining the slowest disclosed usage growth with the fastest revenue growth still gives a bound for β of roughly 0.7 and central readings of around 0.4 (see Appendix D for details). The bound is biased upward further if raw token counts understate the growth of quality-adjusted usage. Prices corroborate a low β as inference prices at fixed capability fell about ten-fold per year while total spending rose only three- to four-fold, so revenue expanded far less than usage. We nonetheless set β above these bounds to be conservative about the quantitative strength of the paper's mechanism, as the contest wedge $\lambda = (n - 1 + \beta)/(n\beta)$ and over-investment magnitude rise as β falls. At $\beta = 0.80$ the model sits closer to the frictionless limit $\beta \to 1$ where the contest externality vanishes, so the headline distortions should be read as floors rather than point estimates.

In the bust, the two committed claims are impaired differently. A fraction $\alpha = 0.40$ of circular capital survives the partnership default, consistent with default-boundary recovery in the spirit of Leland (1994). Debt-financed capital is sold into an industry fire-sale, with the loss rate given by (8). The loss has two parts. Asset specificity alone costs $\zeta = 0.50$, whatever the volume sold, a conservative value relative to historical precedents (e.g. recoveries of specialised fiber assets during the 2001 dotcom and 2002 telecom episodes were less than 10 cents on the dollar). The loss then rises with the volume liquidated, at the intensity $\delta = 0.0015$ in (8), until it reaches the cap $\zeta_{\max} = 0.80$, a scrap recovery of one fifth. The intensity follows from the depth of the resale market, $\delta = (\zeta_{\max} - \zeta)\, n/\bar{Q}$. Here $\bar{Q} = \$1.0$ tn is the book value of distressed AI capital that the secondary market absorbs before prices hit the scrap floor. This depth is deliberately thin. The precedent is the crypto-mining busts of 2018 and 2022, when miners' fire-sold graphics cards cut secondhand prices by half or more. Appendix C carries $\bar{Q} \in [0.8, 2.5]$ tn.

[13] Big-5 capital expenditure was about \$450B in 2025 against \$121B of investment-grade bond issuance, with lease and SPV structures carrying a further large share of the financing, including \$662B of lease commitments not recorded on balance sheets according to Moody's. Bond issuance alone puts the debt share near 27%; lease and SPV structures increase it to the 30–40% range.

The remaining financial parameters fix the balance sheet and the cost of capital. We let each period correspond to a year, matching the frequency of capex guidance, bond issuance and accelerator generations.[14] Coalition equity $W = \$200$B is representative of annual AI outlay by a leading hyperscaler (Figure 1, panel a). The net risk-free rate $r_f = 0.05$ matches the recent US policy rate. The pledgeable fraction $\mu = 0.90$ follows the moral-hazard limit of Appendix B.3, under which lenders advance income net of the borrower's agency rent, here 10%, within the standard range (Tirole, 2006). Under this pledgeability, the boom-side borrowing limits (6)–(7) are slack: the highly profitable build-out can always pledge enough future revenue. This is by design. Holding the pledgeable-income limit slack for the representative, revenue-rich coalition isolates the contest and the fire-sale in the symmetric model. The same limit binds in Section 5, where an idiosyncratic loss impairs a backer's pledgeable income and forces it to retrench. The financing constraint is thus dormant in the aggregate boom and active in idiosyncratic distress. The friction that operates in the symmetric model is instead on the *bust* side: the industry fire-sale (8), whose loss rate the coalition internalises in its leverage choice. Because it takes the going recovery rate as given, the coalition internalises the fire-sale price but not its own contribution to it. Internalising it in full would lower the coalition's leverage by about a fifth while leaving the headline magnitudes essentially unchanged, because each coalition's own contribution to the sector loss rate is small (Appendix C).

4.2 Numerical results

At the calibrated baseline the contest wedge is $\lambda = 1.20$, and Proposition 1 delivers compute over-investment of 1.34 times and dollar over-investment of 1.51 times. These structural ratios describe the frictionless benchmark: they depend only on (ν, β, n), and are independent of the calibrated parameters (A, κ, ϕ). They are also conservative. Across the build-out cost-share range $\nu \in [0.60, 0.78]$ the dollar ratio runs from 1.42 to 1.62. It reaches 1.8 times at the $\beta = 0.7$ upper end that the revenue and usage evidence implies, and about 3 times at its central readings near 0.4 (Appendices C and D).

The full model delivers a similar distortion in levels. At the equilibrium each coalition commits \$280B of period-1 investment, implying sector period-1 capex of $nI_1 \approx \$1.4$tn. The planner in the same environment holds no circular stakes, takes on no committed debt, and commits its equity of \$200B, so the equilibrium over-invests by a factor of 1.40, in line with the structural benchmark. The coalitions front-load this capex, with the share of period-1 investment rising from 43% without the contest head-start to 60%, carried by \$92B of

[14] The lengths of the boom and the bust are unknowable in advance and are not calibration inputs. The headline magnitudes are fairly robust to this baseline (Appendix C).

Table 1: Model parameters.

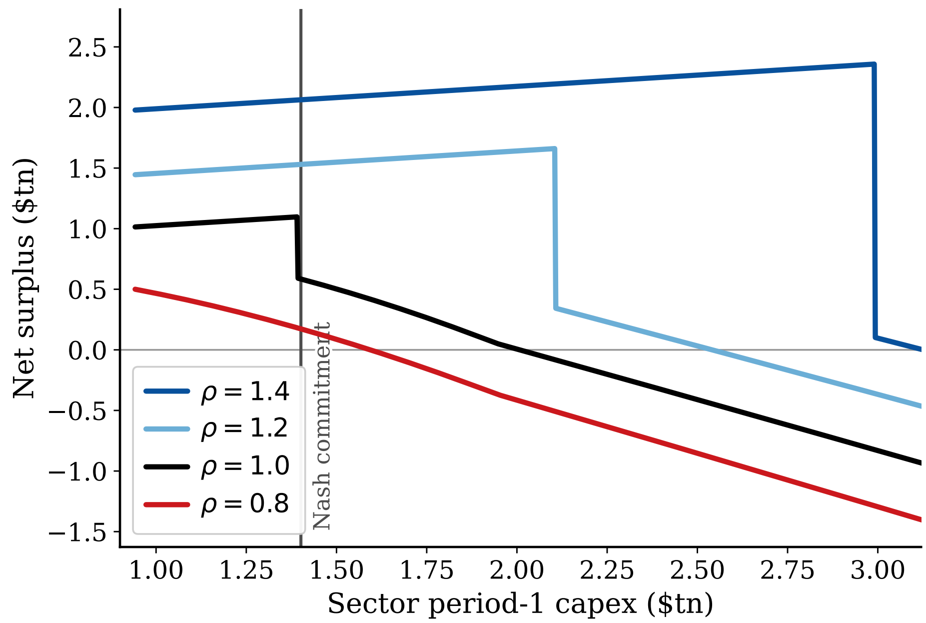
Symbol	Description	Value	Basis
Real economy			
n	AI coalitions	5	Major hyperscaler-lab partnerships
ν	Production concavity	0.70	Build-out cost share
θ	Early-deploy premium	0.85	Frontier-accelerator depreciation
ε	Contest head-start	0.20	Market share persistence (App. D)
ψ	Strategic benefit	0.30	Hold-up band midpoint (App. B.1)
A	Productivity scale	28.82	Calibrated to capex scale
β	Sector returns to scale	0.80	Upper edge of revenue/usage bound (App. D)
α	Circular survival	0.40	Default boundary, Leland (1994)
ζ	Fire-sale loss, specificity	0.50	Dotcom/telecom collapse
δ	Fire-sale intensity	0.0015	Market depth $\bar{Q} = \$1.0$tn (Shleifer–Vishny)
Financial structure			
W	Coalition equity	\$200B	Hyperscaler capex scale
r_f	Net risk-free rate	0.05	US benchmark
R_d	Gross cost of debt	1.05	$1 + r_f$
κ	Circularity-cost convexity	1.50	Normalisation ($\gamma_{\text{plan}} \to 0$)
ϕ	Deployment friction	1.03	Calibrated to $c_1/C = 0.71$
μ	Pledgeable fraction	0.90	Pledgeable-income form (App. B.3)

committed debt, about 33% of the period-1 investment. The capital it funds bears the brunt of the bust.

The endogenous bust threshold sits at an interior point $\bar{\rho} = 1.00$, just above the median productivity draw. Because the threshold rises with period-1 investment, it is the coalitions' front-loaded commitment, driven by the race for the head start, that pushes the sector closer to the bust (Corollary 1). The bigger the boom, the stronger the productivity draw needed to sustain it. Under the baseline calibration, the likelihood of a bust is close to even odds. And since the productivity shock is common, all coalitions cross the threshold together, producing a discrete, sector-wide fall in output rather than a gradual decline.

Corollary 1 decomposes these odds as a price. The floor $1 - \theta$ contributes 0.15 of the bust probability. The head-start premium adds about 0.31 and the salvage value 0.12, while the boundary discount subtracts 0.10, together reproducing the equilibrium probability of one half. Each coalition internalise its own unwind threshold, which activates the boundary discount. The race is the dominant factor: without the head start ($\varepsilon = 0$) the coalition takes on no debt at all and the bust probability falls to 0.38. The contest wedge, by contrast, hardly affects the bust frequency. A planner holding the same circular share and deployment rate busts with essentially the same odds, 0.51 against the equilibrium 0.50. Two channels net out: the planner disregards the head-start premium, which lowers the bust odds, but it also deploys less under weak draws than contestants who overvalue deployment, which raises the odds. What separates the two is the depth: conditional destruction of $378 billion against $85 billion. The contest mainly changes what the bust destroys rather than the bust probability.

Figure 3: Investment boom-bust



Notes: Each line traces ex-post net surplus, sector revenue less the cost of capital, against committed period-1 capex, for a fixed productivity draw ρ. A draw rises along the boom branch, then drops sharply once the commitment is large enough to push it into a bust. Higher draws sustain more commitment but suffer larger busts. The vertical line marks the Nash commitment. Both axes are in trillions of dollars.

The scale of period-1 capex and the productivity draw jointly determine whether the build-out pays off. Figure 3 plots the sector's net surplus, namely the total revenue minus the cost of deploying the capital ($\Theta - R_d I$), against how much it commits in the first period, one curve per productivity draw. A stronger draw allows larger capex commitments to translate into higher net surplus. But there are still limits: for each productivity draw, a sufficiently large capex in period 1 could tip the sector into a bust, causing the surplus to fall off a cliff. At the baseline Nash commitment of $1.4tn, the median draw ($\rho = 1$) already tips into the bust, resulting in a discrete drop in net surplus. Thus, higher capex has two consequences. First, it raises the bust probability, as the rising threshold overtakes more draws. Second, conditional on a bust, it lowers each draw's net surplus. As a result, higher capex ultimately lowers expected net surplus.

Figure 4 runs the comparative statics on the three forces behind the boom-bust: the cost of debt R_d, the contest head-start ε, and the productivity of late-deployed capacity θ. The left column plots net surplus against committed capex, marking the Nash commitment on each curve, and so shows how each force moves the equilibrium scale of the build-out and the payoff it earns. The right column plots output per coalition against the productivity draw, showing how each force moves the bust threshold and the level of output.

The cost of debt R_d provides a potential link to monetary policy. Higher rates (top row of Figure 4) curb borrowing, so the Nash commitment falls and the surplus curve shifts down with the higher cost of capital. In productivity space the threshold barely moves. Consistent with Corollary 1, the interest rate scales the boom and the bust threshold in offsetting proportion, so it shifts the bust probability only through the premia, by a couple of percentage points across $R_d \in [1.03, 1.10]$. Tighter monetary policy thus cools the size of the build-out, and with it the capital at stake in a bust, more than the likelihood of one.

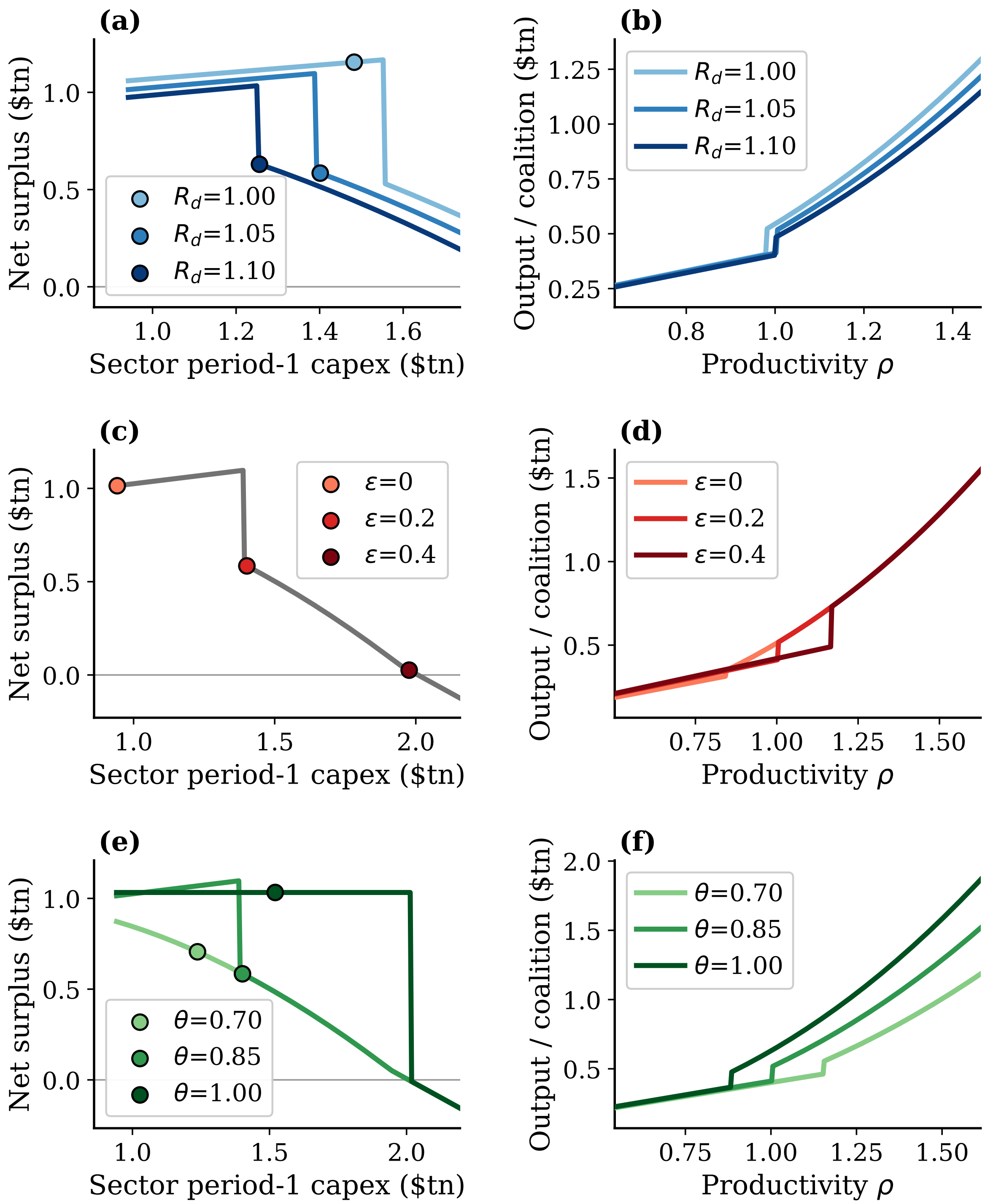
The contest head-start ε drives the over-commitment (middle row of Figure 4). It slides the equilibrium along the surplus curve to the right. The stronger the contest head-start, the further the sector commits and the more vulnerable it is to falling off the cliff. In productivity space the threshold climbs with the contest head-start, so the same force makes the sector build more and bust sooner. Without the head-start, the fragility weakens considerably.

The early-deploy premium determines the position of the cliff (bottom row of Figure 4). The more useful second-period capacity is (higher θ), the lower the productivity needed to keep deploying it, so the threshold falls, output rises, and the sector can build further before triggering a bust. As θ approaches one, late capacity is nearly as good as early, the threshold drops, making it easier to sustain a boom. The faster capacity becomes obsolete, the higher the bar and the more fragile the build-out.

As small differences in productivity realisations can have large impacts due to boom-bust cliffs, the same capex commitment can lead to two very different outcomes. Figure 5 plots the distribution of the sector's realised net surplus across productivity draws, distinguishing the boom outcomes (blue) from the bust outcomes (red), at four levels of committed capex. The distributions are all bimodal: a cluster of high-surplus booms on the right, a cluster of low-surplus busts on the left, and a gap in between that represents the cliff.

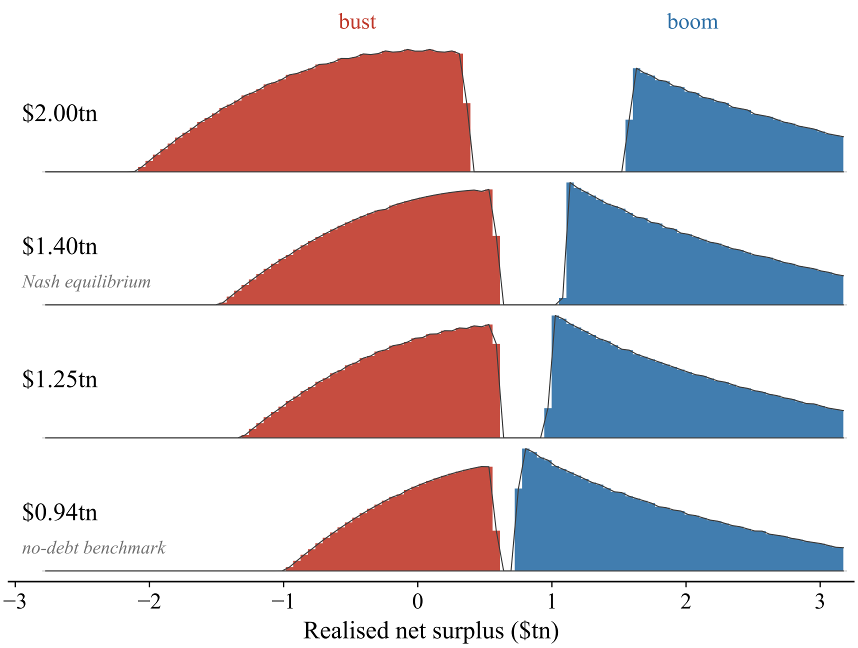
Committing more not only makes a bust more likely, it also worsens the bust. As investment increases from the no-debt benchmark of \$0.94tn through \$1.25tn to the Nash commitment of \$1.40tn and on to \$2.00tn, the bust-regime probability grows due to the rising productivity threshold, while the distribution slides left, as a heavier debt load eats further into each bust outcome. At the no-debt benchmark the busts are mild, clustered just above zero. At the Nash equilibrium, a bust is more likely to generate outright negative than positive net surplus. At \$2.00tn the lower tail deteriorates further still. Higher capex drives

Figure 4: Comparative statics



Notes: Each row varies one parameter, holding the rest at baseline: the cost of capital R_d (top), the strength of the early-mover advantage ε (middle), and the relative productivity of late capital θ (bottom). The left column plots ex-post net surplus against committed period-1 capex at the central draw $\rho = 1$, with the dots marking the equilibrium commitments. The right column plots output per coalition against productivity ρ at that equilibrium, with the kink marking the bust threshold.

Figure 5: Distribution of net surplus

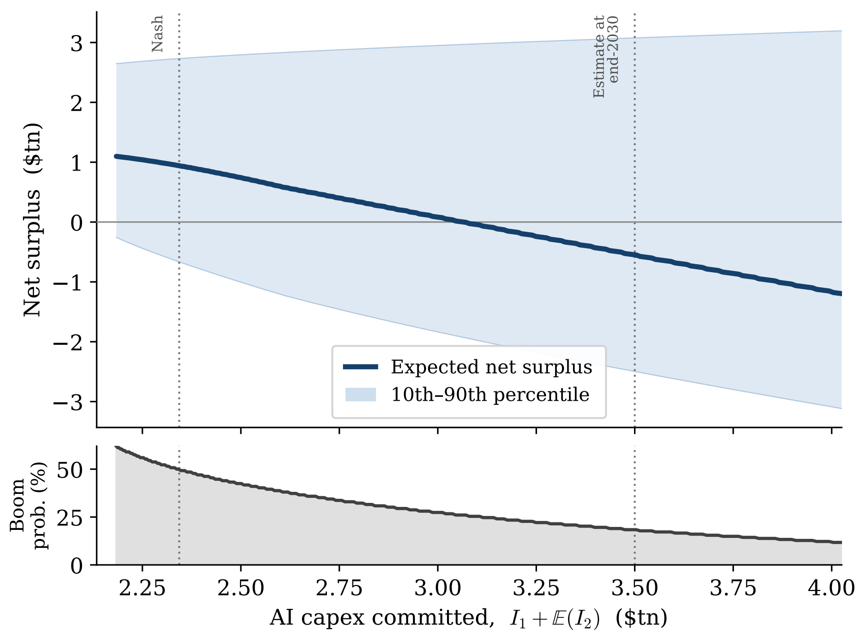


Notes: Each row is the distribution of realised net surplus over the Beta(2,2) productivity draw, for a given level of committed period-1 capex. The endogenous cliff splits every distribution into two sections, corresponding to boom (blue) and bust (red).

a wedge between boom and bust outcomes, widening the gulf between their densities and creating larger cliffs. The contest drives the sector to commit excessively, while the leverage that funds it deepens the fall when a boom ends in a bust.

Conditional on a bust, the immediate loss is concentrated in the financing structure. Each coalition writes off the fraction $1-\alpha$ of its circular commitments and ζ_{eff} of its debt-financed capital: at the calibration, \$85bn of collapsed circular positions and \$293bn of fire-sold debt-financed capital, about \$378bn across the sector. More than three-quarters of that is the debt fire-sale; the leverage that pulled the build-out forward is precisely what makes its unwinding costly. Weighted by the bust probability of about one-half, this committed-capital destruction runs to roughly \$190bn, on top of the broader collapse in revenue on impaired capital that the distribution in Figure 5 shows.

Figure 6: Average net surplus and total capex



Notes: The solid line is expected net surplus, $\Theta - R_d I$. The shaded band spans the tenth to ninetieth percentile of realised outcomes. The lower panel is the probability of avoiding a bust. The horizontal axis is total committed capex, $I_1 + \mathbb{E}(I_2)$: the period-1 commitment plus its expected period-2 deployment. Vertical lines mark the model's Nash commitment and the \$3.5tn midpoint of the \$3–4tn that industry guidance (Huang, 2025) puts on end-of-decade AI-infrastructure capex.

The boom-bust fragility means the average net economic benefit from the AI build-out declines with the scale of the build-out, as shown in Figure 6 together with an 80-percent probability band. Some recent industry estimates suggest that the total cumulative AI capex between 2025 and 2030 could reach \$3–\$4tn (Huang, 2025). Under the baseline calibration, our model suggests expected net surplus may turn negative before the total capex $I_1 + \mathbb{E}(I_2)$ reaches around \$3tn. There are of course sufficiently favourable productivity draws that can sustain a large capex expansion, as shown in the upper part of the fan. But that likelihood diminishes as the build-out grows, e.g. at \$4tn total capex, the probability of a boom falls to around 12% under the baseline calibration.

These numerical estimates are only suggestive and the model is not designed to pinpoint the sustainable AI capex scale. That would depend on the productivity scale A that AI can ultimately generate, which remains uncertain and is difficult to assess precisely. What the model illustrates is the potential extent of excesses and fragility of optimism, which is significant even if the technology brings about genuine productivity gains.

5 Network propagation

The model so far assumes symmetric coalitions as a simplification to focus on the time dimension of the AI investment cycle. This section turns to the cross-section. Relationships between AI firms are complex due to supply chains connections and financial linkages, with circular financing standing out in particular. When labs and their hyperscaler backers fund one another, a single failure can cascade across the network. We therefore generalise beyond common shocks and ask whether an idiosyncratic shock to one lab can propagate through balance sheet connections to affect others.

Under a bipartite network, the financing constraint that stays slack along the symmetric boom comes to the forefront. With heterogeneous backers and an idiosyncratic loss, the pledgeable-income limit (6) binds: an impaired backer is forced to retrench and withdraws funding from the other labs it supports, and the retrenchment propagates. The network model is effectively the coalition case unbundled. The backer's borrowing cap is the period-1 pledgeable-income constraint (6); its pledgeable income is the moral-hazard object of Appendix B.3 at the same $\mu = 0.90$; the stakes on its balance sheet are the circular positions of Section 2.2 marked to market; and the bust that strands them is the threshold of Proposition 4 at the Section 4 calibration. The topology of who backs whom and at what scale is calibrated, and the quantitative results are illustrative rather than predictive. What we study is the contagion that the binding constraint generates under heterogeneous ownership.

Propagation hinges on how a firm meets a loss it did not originate. A lab's bust strands the sunk capital committed by its backers. A backer can absorb that loss without retrenching from investment as long as its borrowing constraint stays slack. Whether this is the case depends on two primitives. The first is the backer's cushion of pledgeable income, which sets how far it can borrow before the limit binds. The second is the composition of the backer's liability: equity is marked down and absorbed, whereas debt forces the borrowing backer to retrench and withdraw funding from the other labs it supports, possibly triggering further failures there. A shared backer is thus the conduit through which one cluster's bust reaches another.

Because failures beget failures, the cascade is a fixed point. Two network quantities determine the system's fragility to busts: the *scale* of committed capital and the *concentration* of exposure across backers. The result is sharply nonlinear. Below a threshold an isolated bust is absorbed; above it the same bust spills over across the network. The threshold falls as financing concentrates on common counterparties. Fragility to busts is thus a property of the network's scale and shape, and a cascade can affect even solvent firms.

5.1 Network model

We relax the symmetric coalition assumption and consider a *bipartite* network consisting of AI labs $l \in \mathcal{L}$ and backers $h \in \mathcal{H}$ (e.g. hyperscalers, neoclouds and chipmakers).[15] Backer h provisions compute worth $e_{hl} \geq 0$ to lab l and holds an equity stake $\sigma_{hl} \geq 0$. We take e_{hl} and σ_{hl} as fixed numbers, empirically calibrated to data. Each active lab has available compute worth $\sum_h e_{hl} > 0$. h's collateral is the sum of its pledgeable income P_h (also calibrated) and the surviving stakes on labs

$$V_h(\Phi) = P_h + \sum_{l \notin \Phi} \sigma_{hl}, \qquad V_h^0 \equiv V_h(\varnothing), \tag{23}$$

where $\Phi \subseteq \mathcal{L}$ is the set of labs that have busted. Let backer h invest $b_h(s) \equiv s \sum_l e_{hl}$ in the build-out, with the multiplier $s \geq 1$ spanning counterfactual cases of larger build-out relative to the baseline $\sum_l e_{hl}$. The backer funds the build-out with debt $D_h(s)$ up to the pledgeable limit,

$$D_h(s) = \min\{\, b_h(s),\ \mu V_h^0 \,\}, \tag{24}$$

with the residual $b_h - D_h \geq 0$ raised as equity.

The shock is, as in Section 3.4, the period-2 productivity draw. Lab l is busted when $\rho_l < \bar{\rho}$, giving $\Phi_0 = \{l : \rho_l < \bar{\rho}\}$. Let $\mathcal{F} \subseteq \mathcal{H}$ and $\Phi \subseteq \mathcal{L}$ be the sets of retrenched backers and busted labs, respectively. Define the mapping $T = (T_{\mathcal{H}}, T_{\mathcal{L}})$ as:

$$h \in T_{\mathcal{H}}(\mathcal{F}, \Phi) \iff D_h(s) > \mu V_h(\Phi), \tag{retrench}$$

$$l \in T_{\mathcal{L}}(\mathcal{F}, \Phi) \iff \rho_l < \bar{\rho} \ \text{ or } \ \frac{\sum_{h \in \mathcal{F}} e_{hl}}{\sum_h e_{hl}} > \tau. \tag{bust}$$

[15] A bipartite network is a graph whose nodes can be divided into two distinct, non-overlapping sets, say $\mathcal{H}$ and $\mathcal{L}$, and every edge must connect a node in $\mathcal{H}$ to a node in $\mathcal{L}$. No edge can connect two nodes within the same set.

The first line defines $T_{\mathcal{H}}(\mathcal{F}, \Phi)$ as the backers whose debt breaches their limit once the labs in Φ have stranded their stakes (collateral $V_h(\Phi)$ in place of V_h^0). The second defines $T_{\mathcal{L}}(\mathcal{F}, \Phi)$ as the labs that fail either on their own productivity draw or because backers in $\mathcal{F}$ have pulled more than a share $\tau \in (0, 1)$ of their compute. Each map responds to the opposite side: $T_{\mathcal{H}}$ to which labs have busted, $T_{\mathcal{L}}$ to which backers have retrenched.

A backer retrenches when its debt overshoots its limit. By (23) this condition means the bust shrinks the borrowing limit by more than the unused borrowing capacity $z_h(s) \equiv \mu V_h^0 - D_h(s)$,

$$\mu \sum_{l \in \Phi} \sigma_{hl} > z_h(s) \geq 0. \tag{25}$$

Call h a *bridge from l to l'* if it holds an equity stake in l and supplies compute to l', $\sigma_{hl} > 0$ and $e_{hl'} > 0$. A bust at l affects the backer's balance sheet through h's stranded stake, potentially causing a retrenchment that then starves l' of compute. Call h *pivotal* for l' if it supplies more than a share τ of that lab's compute, $e_{hl'} / \sum_{h'} e_{h'l'} > \tau$. Let $(\mathcal{F}^*(s), \Phi^*(s))$ be the least fixed point of T above $(\varnothing, \Phi_0)$ and $N(s) = |\mathcal{F}^*(s)| + |\Phi^*(s)|$.

Proposition 7 (Network propagation)**.** *Fix a network with a pivotal bridge h^* from l_0 to l_1.*

(i) Fragility rises with the build-out s in two senses.

(a) (Cascade cliff.) *For the single trigger $\Phi_0 = \{l_0\}$, $N(s)$ is a non-decreasing step function with $\Phi^*(s) = \{l_0\}$ for $s \leq s^* \equiv \inf\{s : |\Phi^*(s)| \geq 2\} < \infty$, and a jump $\lim_{s \downarrow s^*} N(s) \geq N(s^*) + 2$.*

(b) (Systemic risk.) *If $\mathcal{L} = \{l_0, l_1\}$ and (ρ_{l_0}, ρ_{l_1}) places positive probability on l_0 busting while l_1 survives, the co-bust probability $F_{\text{sys}}(s) = \Pr(l_0$ and l_1 both bust$)$ is non-decreasing in s and, for $s > s^*$, strictly exceeds the s-independent comovement floor $\Pr(\rho_{l_0} < \bar{\rho},\ \rho_{l_1} < \bar{\rho})$.*

(ii) (Concentration.) *Keep $\mathcal{L} = \{l_0, l_1\}$ and let h^* be the only bridge from l_0 to l_1. Holding pledgeable incomes P_h, stakes σ_{hl}, and each lab's total funding $\sum_h e_{hl}$ fixed, concentrating funding onto h^* (raising e_{h^*,l_0} and e_{h^*,l_1} at its co-funders' expense) weakly lowers s^*, bringing both the cliff and the rise in systemic risk forward. It is least when h^* funds both labs in full.*

Proof. See Appendix A.8. □

The proposition highlights the nonlinear destabilising effects of the investment scale arising from network connections. A backer borrows against fixed pledgeable income and equity stakes in labs, and must increase debt to finance a larger build-out. A larger build-out raises this debt toward the collateral ceiling. The cascade starts when an adverse productivity draw induces a lab's bust, hitting the equity stake of the stretched backer and pushing its collateral below its outstanding debt and forcing it to pull funding. That withdrawal could trigger another lab's failure, which then devalues the stakes of other backers, eroding their collateral and tipping the next backer over its own limit. When strong enough, this cascading process could induce a jump at the threshold scale s^* where multiple firms go down at once. The financing connections could make two labs that never transact fail in a single episode. The cross-holding that financed the boom is also the same one that transmits the bust. The degree of fragility then depends not only on the scale of investment, but also how much funding passes through critical nodes of the network.

The retrenchment condition (25) is the composition margin of Proposition 6 in network form. What shrinks a backer's borrowing limit in the bust is precisely the circular stake that secured demand in the boom, so the excess circularity of Proposition 2 tightens the cascade condition one for one. The financing mix the contest creates is not only the costliest to destroy. It is also the channel that transmits the destruction.

5.2 Network simulation

We illustrate the theoretical mechanism in a stylised quantitative exercise, in which a hypothetical single failure of a lab propagates through the network and cascades to other firms. The pseudo-realistic network consists of two major labs, ℓ_1 and ℓ_2, and six backers, h_1 to h_6 (four hyperscalers, a chipmaker and a neocloud).[16] The labs are financed by an overlapping group of backers, and the same firm can fund one lab while holding a stake exposed to another. To quantitatively discipline the simulation, these financial contracts are calibrated to empirical data whenever available. But given the partial data coverage and the stylised nature of the exercise, the simulation is meant to be illustrative rather than predictive. To retain focus on the mechanism, we anonymise all backers and labs.

The model is calibrated using the public deal record and firms' financial statements, current as of June 2026. Each backer-lab tie comprises two links. The first is an equity stake that the backer has invested in the lab. This is anchored to the backer's disclosed ownership share at the lab's most recent valuation.[17] The second is a compute commitment, the dedicated capacity the backer finances and builds for the lab, in exchange for a commitment by the lab to buy. We pin this to the compute-purchase contracts the two sides have announced.

[16]We abstract from the vertically integrated labs that fund their compute in-house, as these are not *directly* connected to the main network. The viability of these integrated labs and their mother companies of course may pose systemic risks in practice through mechanisms beyond the scope of this paper, including through financial institutions.

[17]Stakes are the backer's disclosed economic share of the lab's latest post-money round.

Each backer can borrow up to a fraction $\mu = 0.9$ of its pledgeable income, which is its own operating cash flow together with the marked-to-market value of its lab stakes.[18] The scale s measures the build-out against the contracts already signed. At $s = 1$ each backer builds what it has announced and borrows what its cash flow cannot cover. A higher s pushes the build past those contracts and raises the debt each backer carries against an unchanged income base. The few quantities undisclosed by the deals are drawn from uniform distributions over assumed plausible ranges: these include how ℓ_2's compute is split between its two suppliers, the marked-to-market value of h_3's stake in ℓ_2, and the share of compute a lab can lose before it fails (bust tolerance).[19]

We trigger the cascade with a hypothetical failure of ℓ_1, and let all other firms adjust endogenously. This choice is expositional, aimed to illustrate the transmission mechanism when the sector comes under stress, under a stylised yet data-disciplined setup. We interact the hypothetical failure with the investment scale s, following Proposition 7 which shows that the sector's fragility grows with the build-out scale.

Figure 7 shows the simulation results. Panel (a) shows the cascading between key nodes in a stylised network. At the baseline scale $s = 1$, ℓ_1's failure would already trigger busts in its two thinnest suppliers, h_2 and h_6 ("fragile at shock"). As the scale s expands, the strain additionally reaches two major backers, h_1 and h_3, whose debt exceeds their borrowing capacity after their ℓ_1 stakes are written down ("cascade from scale"). Finally, in scenarios where h_3 supplies a large share of ℓ_2's compute and/or if the latter has a lower bust tolerance ("cascade from adversity"), the distress would cross the bridge into ℓ_2 as well.

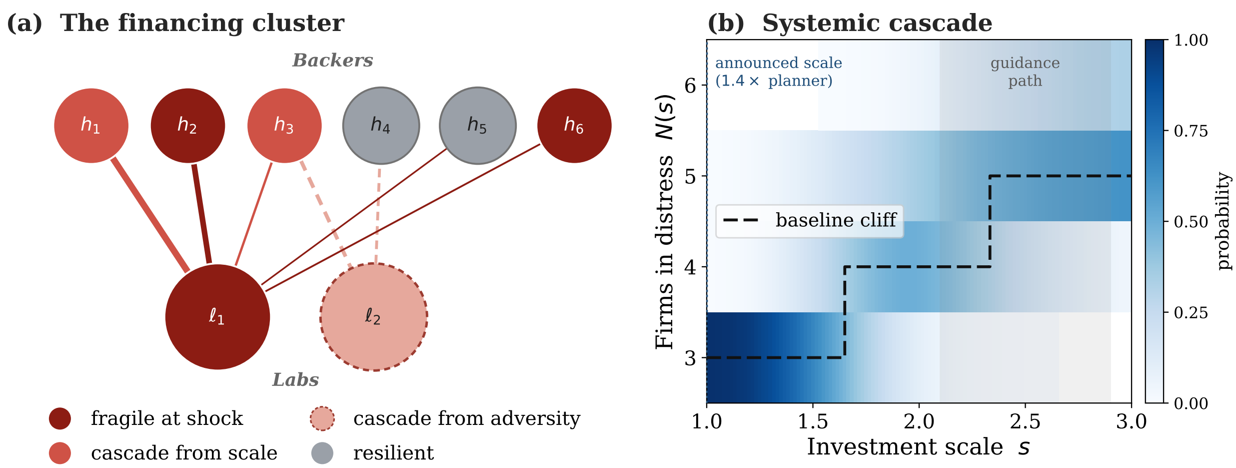
The complementary panel (b) of Figure 7 shows the number of firms in distress, $N(s)$, as a function of the scale s, accounting for uncertainty about variables without tight empirical anchors: the split of ℓ_2's compute between its two suppliers, the marked-to-market value of h_3's stake in ℓ_2, and bust tolerance. The dashed line traces the central scenario, where the assumed random variables take their mean values. This step function is increasing with s, as shown in Proposition 7. The general case where all the variables are stochastic also exhibits this monotonicity. The distribution function for $N(s)$, the shaded areas, shows that the probability of a larger cascade rises with the investment scale s.

The shock scale s places the network results against the benchmarks of the core model.

[18] Pledgeable income is each backer's trailing-twelve-month operating cash flow, discounted one period at R_d, from its 2026 10-Q and 10-K filings, with periods ending 31 March to 31 May 2026. The internal funds it draws on first are its free cash flow over the same period.

[19] h_4's compute commitment to ℓ_2 is drawn from a range derived from the disclosed bounds of its infrastructure build-out and the estimated capital expenditure per unit of capacity. Similarly, the marked value of h_3's stake in ℓ_2 is bounded between its previously reported historical position and its implied value during the company's most recent funding round. The unobservable loss tolerance parameter and the log standard deviation for lognormal debt dispersion are set to assumed baseline values.

Figure 7: Network cascade



Notes: The figure shows the endogenous cascading effects of an exogenous failure of lab ℓ_1, computed by iterating the retrenchment-and-bust map to its fixed point, given a build-out scale s. Panel (a) shows the cascade between key nodes (ℓ_1, ℓ_2 the two labs, h_1 to h_6 their backers) with edges representing financing relationships. "Fragile at shock" are immediately impaired by the shock. "Cascade from scale" fail only if the scale s is sufficiently large. "Cascade from adversity" requires large s as well as adverse scenarios to materialise. Panel (b) plots the cascade size $N(s)$ (the number of firms in distress), as a function of scale s. Shades show the probability $\Pr(N(s) = k)$ over 8,000 draws of stochastic variables (the split of ℓ_2's compute between its two suppliers, the marked value of h_3's stake in ℓ_2, and the bust tolerance). The dashed line is the deterministic baseline with random variables fixed at their central values. The dotted vertical line marks the announced scale $s = 1$, which the core model places at 1.4 times the planner-equivalent commitments; the shaded band is the \$3–4 tn end-of-decade guidance path under a proportional mapping. The remaining equity stakes and compute commitments, together with pledgeable incomes and cash flows, are calibrated from the deal data, with $\mu = 0.9$.

The announced contracts at $s = 1$ correspond to the equilibrium boom, and the over-investment of 1.4 times in Section 4.2 puts the planner-equivalent commitment at $s \approx 0.7$. Scaling back to the planner level would not spare the thinnest suppliers, whose fragility is a matter of composition rather than scale. But it would lessen the risk of contagion. In the central calibration the cascade reaches the first major backer at $s \approx 1.6$ and the second at $s \approx 2.3$, both well above the planner scale. A proportional mapping of the \$3–4 tn end-of-decade guidance onto the roughly \$1.4 tn announced places the continuation of the race at $s \approx 2.1$–2.9, past the first threshold and straddling the second, the shaded region in Panel (b). The contest's level distortion does not push the network into the contagion region by itself. The continuation of the race along the guidance path does, while the financing mix the contest creates decides who is fragile at any scale.

This simulation is illustrative and subject to several caveats. Financial deals are privately negotiated, disclosed in pieces through company announcements, regulatory filings and press reports, and frequently revised. While we have assembled deal data from public sources on a best-effort basis, they may be incomplete or superseded by later announcements. The quantitative results should thus be taken as illustrative rather than definitive. The exercise is designed to demonstrate how the financing network shapes systemic fragility, but is not appropriate for gauging individual firms' vulnerability. Importantly, the analysis also abstracts from the financing of AI investment activities by banks or nonbank entities such as private credit and special purpose vehicles, which have played an important role in a number of cases. Just as a common exposure to the same AI firms can propagate shocks across the network, sharing exposures to financial institutions could also produce vulnerabilities. We leave this important channel to future research.

6 Conclusion

The current AI boom shares many of the key features that characterised investment boom-bust cycles accompanying genuine technological breakthroughs in the past. Exuberance about the technology and potential rewards provide strong incentives for entrepreneurs to participate in the race and invest heavily in order to win it. To facilitate the expansion, firms also seek leverage and adopt more innovative financing models. The specialised capital that powers the boom is also what a downturn must liquidate, all at once and into a thin market, so the losses deepen with the sector's own leverage. We show that, in aggregate, these could lead to excessive investment, which makes the system endogenously more vulnerable to a disruptive retrenchment. Of course, there remains much uncertainty about the eventual productivity payoffs from AI, and it may well be that these would be more than sufficient to justify large-scale deployment being planned today. This time could well be different. That said, given the significant amount of resources at stake and the copious lessons from history, a more balanced perspective may also be of value. Our paper aims to articulate one argument as a contribution to this important debate.

The paper's second contribution is to couple two forces usually studied apart, competition and financial fragility. On the one hand, the theory of contests typically treats rivalry as an allocative matter. On the other, macro-finance treats fragility as the amplification of shocks that originate outside the financial system. One lesson of the paper is that, in a financed investment race, they become part of the same phenomenon. The race to invest early increases the chance of the boom turning into bust, while the contest externalities set how painful the bust would be. The competitive drive that over-builds the boom is the same force that shapes its fragility. Contest is not merely wasteful in the static sense, as in the classic business-stealing externality. It is a source of endogenous financial fragility with dynamic consequences. The wedge between private and social returns that leads firms to over-invest is the wedge that leaves the system exposed. The AI boom is a particularly relevant case study, but the same mechanism may also be relevant in a broader macro-finance context.

A Proofs

A.1 Proof of Proposition 1

Proof. At the symmetric Nash equilibrium with $\gamma = 0$ and unlimited debt, the coalition's first-order condition in I is $MR^s_{coal} \cdot \nu I^{\nu-1} = R_d$, while the planner's is $MR^s_{plan} \cdot \nu I^{\nu-1} = R_d$. Taking ratios at the symmetric equilibrium uses $MR_{coal}/MR_{plan} = \lambda$ from (13) and the equilibrium-evaluated identity $\Theta/J = AJ^{\beta-1}$. Solving for the ratio of equilibrium investments yields $I^*/I^{plan} = \lambda^{1/(1-\nu\beta)}$ and correspondingly $J^*/J^{plan} = (I^*/I^{plan})^\nu = \lambda^{\nu/(1-\nu\beta)}$. As $\beta \to 1$, $\lambda \to 1$ and both ratios converge to unity. Strict concavity $\nu < 1$ gives $\lambda^{1/(1-\nu\beta)} > \lambda^{\nu/(1-\nu\beta)}$. □

A.2 Proof of Proposition 2

Proof. Each party's first-order condition equates the marginal benefit of the circular share to the marginal cost $\kappa\gamma W$, with corner projections onto $[0, 1]$: $g_\gamma = \kappa\gamma W$ for the planner and $\lambda g_\gamma = \kappa\gamma W$ for the coalition. Since g_γ is evaluated at the chosen allocation, which moves with γ, these conditions characterise the optima as fixed points, the expressions in the proposition. The marginal cost rises linearly from zero, the marginal benefit is continuous in γ, and the convex cost dominates the curvature of the benefit, by an order of magnitude at the calibration, so each payoff is single-peaked in γ and the first-order condition picks out the optimum. The ordering follows from evaluating the coalition's marginal payoff at the planner's optimum. At an interior γ^*_{plan} the planner's condition gives $g_\gamma = \kappa\gamma^*_{plan}W > 0$ there, so the coalition's marginal payoff at the same point is $(\lambda - 1)\, g_\gamma \geq 0$: the coalition's payoff is still rising at the planner's optimum, and its peak lies weakly to the right. If the planner corners at zero, $g_\gamma \leq 0$ at zero, the coalition's marginal payoff λg_γ is also non-positive, and both shares are zero. If the planner corners at one, $g_\gamma \geq \kappa W > 0$ at one, the coalition's marginal payoff there is also positive, and both shares are one. In every case $\gamma^*_{coal} \geq \gamma^*_{plan}$. □

A.3 Proof of Proposition 3

Proof. Fix γ and c_1/C, so $I_1 = (1-\gamma)W + c_1 + B_1$ with the first two terms common to both problems; the only free instrument is early debt $B_1 \geq 0$, and since I_1 rises with B_1 it suffices to compare B^*_1 and B^{plan}_1. The unwind threshold $\bar{\rho}$ is set in the market and taken as given by the individual borrower, so each value is differentiable in B_1. Under the price-taking industry fire-sale of Proposition 5, the coalition treats the bust recovery rate as the constant equilibrium value $\bar{\zeta} = \zeta_{\text{eff}}$; we write ζ for $\bar{\zeta}$ throughout this proof, so the marginal bust cost of debt is the constant ζ, consistent with (8).

A unit of B_1 raises I_1 by one and is repaid at R_d in every state. In a boom output enters only through $S = I_1 + \theta I_2$, so $\partial Q/\partial I_1 = \theta^{-1}\partial Q/\partial I_2$; each party deploys period-2 capacity to its own marginal return R_d, so the boom value of the unit is R_d/θ for both, the induced change in period-2 borrowing being netted by the envelope. Differentiating the weight I_1^{ε} at the symmetric profile adds, to the coalition alone, $H(\varepsilon) = \varepsilon(n-1)\Pi/(nI_1) \geq 0$, increasing in ε with $H(0) = 0$, where $\Pi = \mathbb{E}[\pi_k\Theta(\rho)/K]$ is the expected contest payoff over both branches, since the salience weight earns share in bust states as well. In a bust the unit leaves $(1-\zeta)$ of surviving capital, with social marginal revenue product $M = A\beta\nu\, n^{\beta-1}X^{\nu\beta-1}$, $X = (1-\gamma)W + \alpha c_1 + (1-\zeta)B_1$ the surviving stock, which at the $B_1 = 0$ evaluation used for the interiority threshold below reduces to $(1-\gamma)W + \alpha c_1$; the coalition values its output at the contest margin λM and the planner at M, with $\lambda = (n-1+\beta)/(n\beta) \geq 1$. With F the bust probability, p the productivity density, and $\Psi = \int_0^{\bar{\rho}} \rho\, p\, d\rho > 0$,

$$\frac{\partial V^{coal}}{\partial B_1} = (1-F)\frac{R_d}{\theta} + H(\varepsilon) + (1-\zeta)\lambda M\Psi - R_d, \qquad \frac{\partial V^{plan}}{\partial B_1} = (1-F)\frac{R_d}{\theta} + (1-\zeta)M\Psi - R_d.$$

Subtracting, $\partial V^{coal}/\partial B_1 - \partial V^{plan}/\partial B_1 = H(\varepsilon) + (1-\zeta)(\lambda-1)M\Psi \geq 0$, so the coalition's marginal value of B_1 is everywhere at least the planner's. Each value is single-peaked in B_1, so $B_1^* \geq B_1^{plan}$ and $I_1^* \geq I_1^{plan}$.

Since $\bar{\rho} > 0$ gives $F > 0$ and $\lambda > 1$ ($n \geq 2$, $\beta < 1$), that difference is strictly positive, so the coalition's optimum strictly exceeds the planner's whenever it is interior, i.e. whenever its marginal value at $B_1 = 0$ is positive, $(1-F)R_d/\theta + H(\varepsilon) + (1-\zeta)\lambda M\Psi > R_d$. As H rises from $H(0) = 0$ this holds iff $\varepsilon > \varepsilon^*$, with

$$\varepsilon^* = \frac{nI_1}{(n-1)\Pi}\Big[R_d - (1-F)\frac{R_d}{\theta} - (1-\zeta)\lambda M\Psi\Big].$$

For $\varepsilon \leq \varepsilon^*$ both sit at $B_1 = 0$ and $I_1^* = I_1^{plan}$; for $\varepsilon > \varepsilon^*$, $I_1^* > I_1^{plan}$. The bracket is the shortfall of the early-debt return below its cost, so $\varepsilon^* \leq 0$ iff $(1-F)R_d/\theta + (1-\zeta)\lambda M\Psi \geq R_d$ and $\varepsilon^* > 0$ otherwise. □

A.4 Proof of Proposition 4

Proof. A coalition deploys period-2 capital as long as its private marginal return covers the gross cost R_d. With the contest wedge λ on the compute gradient and symmetric revenue

$\pi_k\Theta/K = \rho A n^{\beta-1} Q_k^{\beta}$, $Q_k = (I_1 + \theta I_2)^{\nu}(1+\psi\gamma)$, that return is

$$\lambda \frac{\partial(\pi_k \Theta/K)}{\partial I_2} = \lambda\theta\nu\beta\, \rho A n^{\beta-1}(I_1+\theta I_2)^{\nu\beta-1}(1+\psi\gamma)^{\beta}, \tag{26}$$

decreasing in I_2 since $\nu\beta < 1$. The corner $I_2 = 0$ is optimal exactly when the first unit fails to cover its cost. Setting (26) at $I_2 = 0$ equal to R_d and solving for ρ gives (17), whose exponent $1 - \nu\beta > 0$ makes $\bar{\rho}$ increasing in I_1.

For $\rho < \bar{\rho}$ the return lies below R_d at every $I_2 \geq 0$, so $c_2 = B_2 = 0$ is the unconstrained optimum. Since the surviving stock still earns $\pi_k\Theta > 0$, the right-hand side of (7) stays strictly positive, so the constraint slacks. The shock is common and the coalitions symmetric, so all cross $\bar{\rho}$ at once.

The unwind destroys $(1-\alpha)c_1$ of circular and the fire-sale $\zeta_{\text{eff}} B_1$ of debt-financed capital, while equity-financed capital E is unencumbered, leaving (18). With $I_2 = 0$ and the partnership boost gone, output is $Q_k = (s\, I_1)^{\nu}$. □

A.5 Proof of Proposition 5

Proof. A price-taking coalition treats the loss rate as a constant $\bar{\zeta}$ and bears a bust loss $\bar{\zeta} B_1$. In a symmetric equilibrium the sector liquidates nB_1, which by (8) (with $\delta = cn$) clears at the rate $\zeta + \delta B_1$; consistency requires $\bar{\zeta} = \zeta + \delta B_1$. The optimal B_1 falls as the rate $\bar{\zeta}$ rises, so the map from $\bar{\zeta}$ to $\zeta + \delta B_1(\bar{\zeta})$ is decreasing and bounded above by $\zeta_{\max}$, and the fixed point exists and is unique.

Relative to a flat recovery ζ, the coalition faces a higher marginal bust cost on debt, $\partial(\bar{\zeta} B_1)/\partial B_1 = \bar{\zeta} > \zeta$, so its optimal B_1 is lower. The fire-sale is, however, a cost borne *symmetrically*: the coalition and the planner face the same equilibrium recovery rate $\bar{\zeta}$, so it enters their debt first-order conditions identically and lowers both their borrowing. It therefore cannot close the gap between them. That gap is set by the contest wedge: from the proof of Proposition 3 the marginal-value difference is $H(\varepsilon) + (1-\bar{\zeta})(\lambda - 1)M\Psi \geq 0$, which is non-negative for every recovery rate $\bar{\zeta} < 1$ and strictly positive when $\lambda > 1$. Hence the fire-sale attenuates the level of the boom but not the over-investment, and $I_1^* > I_1^{\text{plan}}$ whenever $\lambda > 1$ and the equilibrium carries positive committed debt ($\varepsilon > \varepsilon^*$), fixing γ and c_1/C as in Proposition 3.

Finally, the bust loss $\zeta_{\text{eff}} B_1 = (\zeta + \delta B_1)B_1$ is increasing and convex in B_1, and the equilibrium B_1 rises with the boom, so a larger build carries a deeper bust. □

A.6 Proof of Corollary 1

Proof. The decomposition. The interior period-1 first-order condition of Appendix A.3 is $(1-F)R_d/\theta + H(\varepsilon) + (1-\zeta_{\text{eff}})\lambda M\Psi = R_d$, where F is the bust probability, $H(\varepsilon) = \varepsilon(n-1)\,\mathbb{E}[\pi_k\Theta(\rho)/K]/(nI_1)$ the head-start term, $M = \nu\beta A\, n^{\beta-1}(sI_1)^{\nu\beta-1}$ the bust marginal revenue product of the surviving stock, and $\Psi = \mathbb{E}[\rho\,\mathbf{1}\{\rho < \bar{\rho}^*\}]$. Rearranging for F gives (19). The condition, and hence (19), is exact under the stated convention that the coalition takes the unwind threshold as market-given. When the coalition instead internalises its own threshold, a marginal committed dollar also raises the threshold by $\partial\bar{\rho}/\partial B_1 = (1-\nu\beta)\bar{\rho}/I_1 > 0$ and tips the states just above $\bar{\rho}$ from boom to bust, so the bracket in (19) gains the boundary discount $-\Lambda$, where

$$\Lambda \;=\; p(\bar{\rho})\,\frac{\partial\bar{\rho}}{\partial B_1}\left[\pi_k\Theta(\bar{\rho})/K\Big|_{\text{boom}} - \pi_k\Theta(\bar{\rho})/K\Big|_{\text{bust}}\right] \;\geq\; 0,$$

with p the productivity density, positive because the bust impairs committed capital and suspends the partnership synergy. Setting the bracket to zero anchors the bust probability at $1-\theta$.

The wedge enters only through the salvage term. The wedge appears in (19) explicitly only in the bust-state term. Its remaining influence is indirect, through the equilibrium objects I_1^* and $\bar{\rho}^*$, and this influence nets out. By (17) the coalition's threshold carries λ in its denominator, while the planner deploys to the social margin, so at any common level $\bar{\rho}^{\text{plan}}(I_1) = \lambda\,\bar{\rho}(I_1)$. In the frictionless benchmark of Proposition 1, $I_1^*/I_1^{\text{plan}} = \lambda^{1/(1-\nu\beta)}$ by (15), so

$$\frac{\bar{\rho}^*(I_1^*)}{\bar{\rho}^{\text{plan}}(I_1^{\text{plan}})} = \frac{1}{\lambda}\Big(\frac{I_1^*}{I_1^{\text{plan}}}\Big)^{1-\nu\beta} = 1:$$

the wedge enlarges the boom and deepens period-2 deployment in the same proportion, leaving the range of draws that end in a bust unchanged, while the capital destroyed in each bust is larger by the full over-investment factor. The contest's effect on the bust frequency is therefore carried by the head start $H(\varepsilon)$; the wedge's direct appearance in the bust-state term is second-order at the calibration of Section 4. □

A.7 Proof of Proposition 6

Proof. Write $E = (1-\gamma)W$, $c_1 = (c_1/C)\,\gamma W$, and, from $I_1 = (1-\gamma)W + c_1 + B_1 = \bar{I}$, $B_1 = \bar{I} - (1-\gamma)W - (c_1/C)\gamma W$. Holding $\bar{I}$ and the deployment ratio c_1/C fixed,

$$\frac{dE}{d\gamma} = -W, \qquad \frac{dc_1}{d\gamma} = \tfrac{c_1}{C}\,W, \qquad \frac{dB_1}{d\gamma} = \big(1-\tfrac{c_1}{C}\big)W,$$

which sum to zero, so $\bar{I}$ is genuinely held fixed, and $B_1 > 0$ throughout because $\bar{I} > W$. Equity survives in full while c_1 loses $1-\alpha$ and B_1 loses the going rate ζ_{eff}, which the coalition takes as given, so

$$\left.\frac{\partial L^{\text{bust}}}{\partial \gamma}\right|_{\bar{I}} = \Big[(1-\alpha)\tfrac{c_1}{C} + \zeta_{\text{eff}}\big(1-\tfrac{c_1}{C}\big)\Big]W > 0,$$

a convex combination of $1-\alpha > 0$ and $\zeta_{\text{eff}} > 0$: raising the circular share at a fixed scale converts surviving equity into destructible capital.

Next, at a fixed scale the circular share solves the first-order condition $\lambda\, n_\gamma(\gamma) = \kappa\gamma W$, where n_γ is the revenue-based net marginal value of γ, which the contest scales by λ while the governance friction $\frac{\kappa}{2}\gamma^2 W$ carries no λ. Interiority gives $n_\gamma = \kappa\gamma^* W/\lambda > 0$, and the implicit function theorem yields

$$\left.\frac{\partial \gamma^*}{\partial \lambda}\right|_{\bar{I}} = \frac{n_\gamma}{\kappa W - \lambda\, n'_\gamma} > 0,$$

with the denominator positive by the second-order condition. Combining the two derivatives, $\partial L^{\text{bust}}/\partial\lambda\big|_{\bar{I}} = (\partial L^{\text{bust}}/\partial\gamma)(\partial\gamma^*/\partial\lambda) > 0$, the claim. Setting $\lambda = 1$ recovers the no-contest circular share at the same scale, which is strictly smaller, so an economy reaching $\bar{I}$ without the contest wedge busts less deeply.

Finally, holding $\bar{I}$ and γ fixed, $dc_1/d(c_1/C) = \gamma W$ and $dB_1/d(c_1/C) = -\gamma W$, so $\partial L^{\text{bust}}/\partial(c_1/C) = [(1-\alpha) - \zeta_{\text{eff}}]\gamma W$, which vanishes as $\zeta_{\text{eff}} \to 1-\alpha$ and is negligible at the calibration: the composition margin is the circular share, not the timing of deployment. □

A.8 Proof of Proposition 7

Proof. The map T is monotone. As more labs bust, $V_h(\Phi)$ falls (while D_h is unaffected), so more backers breach their limit; as more backers retrench, the failed funding share $\sum_{h\in\mathcal{F}} e_{hl}/\sum_h e_{hl}$ rises, so more labs bust. Hence T is isotone in the state $(\mathcal{F}, \Phi)$. Raising s raises $D_h(s)$ but not the failed share, so T, and hence $N(s)$, are non-decreasing in s. Being integer-valued, $N(s)$ is a step function, and since a backer exactly at its limit is solvent (retrenchment is strict), its jumps fall just past each threshold, so it is left-continuous.

(i.a) Cliff. For $s < s^*$ at most one lab has busted, and it is the trigger, so $\Phi^* = \{l_0\}$. Strictness leaves no backer strictly over its limit even at the boundary s^* (there $D_h \le \mu V_h(\{l_0\})$), so $\Phi^*(s^*) = \{l_0\}$ as well, and the cascade stays contained for all $s \le s^*$. The bridge h^* has $b_{h^*}(s) \to \infty$, so its unused capacity $z_{h^*}(s)$ falls to zero; by (25) it retrenches once $z_{h^*}(s) < \mu\sigma_{h^*,l_0}$, at a finite scale, and being pivotal for l_1 its withdrawal alone carries l_1 past τ. So $s^* < \infty$. Just past s^* a second lab enters Φ^*, and since the failed share moves only when $\mathcal{F}^*$ does, that crossing forces a new backer to retrench at the same scale; the cascade

therefore jumps by at least two, $\lim_{s\downarrow s^*} N \geq N(s^*) + 2$.

(i.b) Systemic risk. For each realisation the cascade is non-decreasing in s, so F_{sys} is too. With $\mathcal{L} = \{l_0, l_1\}$ the co-bust event splits as $A \dot{\cup} B$: the s-independent floor $A = \{\rho_{l_0} < \bar{\rho},\ \rho_{l_1} < \bar{\rho}\}$, and the event B in which exactly one lab is productivity-busted and the other is reached by propagation. This split is exhaustive. There is no third trigger, and a bust cannot start without a productivity one. With none, the cap keeps every $D_h(s) \leq \mu V_h(\varnothing)$, so no backer retrenches and the cascade stays empty. Without a shared backer B is empty and $F_{\text{sys}} = \Pr(A)$. With the bridge, part (i.a) carries the l_0 trigger to l_1 for every $s > s^*$. Because the bridge runs one way, a stake in the trigger l_0 and funding to the victim l_1, the single event $\{\rho_{l_0} < \bar{\rho},\ \rho_{l_1} \geq \bar{\rho}\}$ lies in B and has positive probability by non-degeneracy. Hence $F_{\text{sys}} = \Pr(A) + \Pr(B) > \Pr(A)$ for $s > s^*$.

(ii) Concentration. Because h^* is the only bridge, the trigger reaches l_1 only through it: every other funder of l_1 holds no stake in l_0, so l_0's bust never makes it retrench, and every other holder of l_0's stake does not fund l_1. So l_1 busts exactly when h^* retrenches, and, the cascade being monotone in s, s^* is the scale at which it does. With only l_0 busted, h^* retrenches once its debt exceeds its reduced limit, $D_{h^*}(s) > \mu V_{h^*}(\{l_0\})$; below the cap $D_{h^*}(s) = s(e_{h^*,l_0} + e_{h^*,l_1})$, so

$$s^* = s^{\text{ret}}_{h^*} \equiv \frac{\mu V_{h^*}(\{l_0\})}{e_{h^*,l_0} + e_{h^*,l_1}}.$$

Concentrating funding onto h^*, with stakes and pledgeable incomes fixed, leaves $V_{h^*}(\{l_0\})$ unchanged and raises the denominator, so $s^{\text{ret}}_{h^*}$, and with it s^*, falls weakly; it is least when h^* funds both labs in full, $e_{h^*,l_j} = \sum_h e_{h,l_j}$. $\square$

B Microfoundations

B.1 Hold-up microfoundation of ψ

This appendix derives the strategic benefit ψ from a hold-up problem inside the partnership. The hyperscaler chooses relationship-specific customisation $x \in [0, \bar{x}]$ (custom training-stack engineering, GPU-cluster reservations, joint research staff) at convex cost $\xi x^2/2$. Customisation raises the partnership's surplus multiplicatively, $S(x) = (1 + x)\, V_{\text{base}}$, where V_{base} is the annual bilateral flow at stake in the relationship. We take x as non-contractible (Grossman and Hart, 1986): the parties cannot write ex-ante contracts on the customisation level, and at renegotiation they split $S(x)$ by symmetric Nash bargaining. An equity stake σ in the partner pays a contractual dividend $\sigma S(x)$ ahead of the renegotiated pot, so the hyperscaler's total share is $s^H(\sigma) = \sigma + (1 - \sigma)/2 = (1 + \sigma)/2$.

Maximising $s^H(\sigma)\, S(x) - \xi x^2/2$ gives $x^*(\sigma) = (1 + \sigma)\, V_{\text{base}}/(2\xi)$. At $\sigma = 0$ this is $x_{\text{HU}} = V_{\text{base}}/(2\xi) = x_{\text{FB}}/2$, the canonical Grossman and Hart (1986) underinvestment: the hyperscaler bears the full cost of customisation but captures only half of its return. Evaluating total welfare $W(x) = (1 + x)V_{\text{base}} - \xi x^2/2$ at the two investment levels gives the gain from the stake,

$$\Delta W(\sigma) \;\equiv\; W\big(x^*(\sigma)\big) - W(x_{\text{HU}}) \;=\; \frac{V_{\text{base}}^2}{8\xi}\big[(1+\sigma)(3-\sigma) - 3\big] \;=\; \frac{V_{\text{base}}^2\,\sigma(2-\sigma)}{8\xi}, \quad (27)$$

rising from $\Delta W(0) = 0$ to $\Delta W(1) = V_{\text{base}}^2/(8\xi)$ at full equity, where $x^*(1) = x_{\text{FB}}$.

The reduced form in the body summarises this gain. At the symmetric equilibrium the boost multiplies each coalition's revenue by $(1 + \psi\gamma)^\beta$, so the pair's annual gain from its circular position is, to first order, $\beta\psi\gamma\, V_{\text{base}}$. The microfoundation prices the same position as the stake $\bar{\sigma}$ it has accumulated, with annual gain $\Delta W(\bar{\sigma})$ in (27). Equating the two identifies ψ:

$$\beta\psi\gamma\, V_{\text{base}} \;=\; \frac{V_{\text{base}}^2\,\bar{\sigma}(2-\bar{\sigma})}{8\xi} \qquad \Longrightarrow \qquad \psi \;=\; \frac{\bar{\sigma}(2-\bar{\sigma})\, V_{\text{base}}}{8\beta\gamma\,\xi}. \quad (28)$$

The inputs are pinned by the body and the deal record. The sector returns to scale and the circular share take their calibrated values, $\beta = 0.8$ and $\gamma = 0.20$. The stake is a representative disclosed position, $\bar{\sigma} = 0.3$. The flow at stake is anchored to the largest disclosed compute agreement, about \$50B a year. The cost parameter ξ prices a full customisation upgrade, since raising the relationship's flow by one hundred per cent costs $\xi/2$. We carry $\xi \in [\$50, \$100]$B, a band consistent with the dedicated builds the major pairings have disclosed, which range from a \$22 billion single-supplier contract to a \$100 billion decade-long programme (Section 5). Substituting into (28) gives $\psi = 0.40$ at $\xi = \$50$B and $\psi = 0.20$ at $\xi = \$100$B. We adopt the midpoint of this window, $\psi = 0.30$, as the working baseline.

B.2 Microfoundation of the early-mover weight

We microfound the contest's allocation of fixed sector revenue. The share in (3) is the logit choice probability of a buyer who values coalition k at its capacity plus an early-mover salience term,

$$u_k = \ln Q_k + \varepsilon \ln I_{1,k} + \xi_k, \tag{29}$$

where $\ln Q_k$ values capacity, $\varepsilon \ln I_{1,k}$ is the salience of an early and established provider through default choice, switching costs, and the data lock-in early usage builds, and ξ_k is an i.i.d. extreme-value taste shock. The buyer chooses coalition k with the logit probability

$$\frac{\exp(\ln Q_k + \varepsilon \ln I_{1,k})}{\sum_j \exp(\ln Q_j + \varepsilon \ln I_{1,j})} = \frac{I_{1,k}^{\varepsilon} Q_k}{\sum_j I_{1,j}^{\varepsilon} Q_j}, \tag{30}$$

the per-lottery win probability in (3), with expected slots π_k equal to K times it. Coalition k's expected revenue is this probability times $\Theta(\rho)$.

Three properties follow. First, ε is revenue-neutral at the sector level: the probabilities sum to one, so it reallocates Θ toward early-committing coalitions without changing it. Second, the salience loads period-1 capital alone, so a buyer's choice depends on $I_{2,k}$ only through Q_k, and late capacity cannot buy back early share. Third, $\varepsilon = 0$ collapses the probability to Q_k/J, the plain Tullock contest.

B.3 Moral-hazard microfoundation of B

The Period-1 debt capacity (6) arises from a standard pledgeable-income limit (Holmström and Tirole, 1997; Tirole, 2006). The lab can divert resources at private benefit b per dollar invested. Let R^{lab} and R^{lender} be the lab's and lender's expected gross payoffs, with revenue identity $R^{\text{lab}} + R^{\text{lender}} = \tilde{Y}$ where $\tilde{Y} \equiv \mathbb{E}[Q]\, A\, J^{\beta-1}|_{\text{eqm}}$. The lab's incentive compatibility is $R^{\text{lab}} \geq bI$, and the lender's participation is $R^{\text{lender}} \geq R_d\, B$. At the binding IC, $R^{\text{lab}} = bI$ implies $R^{\text{lender}} = \tilde{Y} - bI$, and the PC yields $B \leq (\tilde{Y} - bI)/R_d$. Local-linearising $\tilde{Y} \approx \tilde{Y}'(I^*)\, I$ around the calibrated I^* (necessary because $Q = I^{\nu}$ with $\nu < 1$ makes $\tilde{Y}$ globally concave) gives the form (6) with pledgeable fraction $\mu \equiv 1 - b/R_d$. The calibrated $\mu = 0.90$ corresponds to $b/R_d = 0.10$, in the standard range for opaque intermediation (Tirole, 2006).

In Period 2, the lender re-collateralises against the realised draw, $Q(\rho)\, A\, J^{\beta-1}|_{\text{eqm}}$; the same constraints at the realised ρ yield (7). We hold μ fixed across realisations.

B.4 Industry fire-sale microfoundation of ζ_{eff}

We microfound the loss rate (8) as a cash-in-the-market industry fire-sale in the spirit of Shleifer and Vishny (1992). In a common bust the n coalitions liquidate their debt-financed capital simultaneously into a market for specialised assets whose natural buyers—the industry peers—are constrained at the same time. The distressed capital sells along a downward-sloping inverse demand, with recovery per unit

$$P(Q) = \rho_H - c\,Q \qquad \text{(floored at a scrap value } \rho_f\text{)}, \tag{31}$$

where ρ_H is the recovery on the first unit sold, set by the highest-value remaining buyer and $c > 0$ is the price impact, steeper the thinner and more constrained the buyer pool. Sector supply is $Q = nB_1$ (each coalition liquidates B_1), so the uniform clearing price is $P = \rho_H - c\,nB_1$ and the loss rate is

$$\zeta_{\text{eff}} = 1 - P = (1 - \rho_H) + c\,n\,B_1 = \zeta + \delta\,B_1, \tag{32}$$

with the volume-independent specificity loss $\zeta = 1 - \rho_H$ and intensity $\delta = c\,n$ collecting the price-impact slope and the buyer concentration. The price reaches the scrap floor when the sector liquidates $\bar{Q}$, where $\rho_H - c\bar{Q} = \rho_f$, so $\delta = (1 - \zeta - \rho_f)\,n/\bar{Q}$; the market depth $\bar{Q}$ is the calibration lever.

Each coalition takes the sector recovery rate as a given price, so the equilibrium rate is the fixed point $\zeta_{\text{eff}} = \zeta + \delta B_1$ evaluated at the symmetric B_1. Price-taking is the standard treatment of a common-pool recovery price (Shleifer and Vishny, 1992; Lorenzoni, 2008; Bianchi, 2011); the marginal effect of a coalition's own leverage on that price, which it ignores, is the fire-sale externality, left uncharacterised here.

The application to AI is sharp rather than loose. The natural buyers are a concentrated set of hyperscalers, correlated in a common bust, so the two Shleifer–Vishny conditions, asset specificity and simultaneously-constrained buyers, hold with force; the crypto-mining busts of 2018 and 2022 are a precedent in this very asset class. Two caveats discipline the loss base. Secular GPU obsolescence is a distinct, additive force and is not part of ζ_{eff}. And part of the debt-financed stock, power interconnects and datacentre shells, recovers nearer par, so applying ζ_{eff} to the full B_1 is conservative-to-generous.

C Sensitivity

The quantitative results rest on three numbers: the over-investment ratio, the bust probability, and the expected destruction of committed capital. This appendix asks how far each one moves as the externally set parameters vary, and how much it owes to the two anchors taken from the deal record (the circular share and the deployment rate). We vary one parameter at a time over a defensible range and hold the calibrated parameters at their Table 1 values. Table 2 collects the results.

Table 2: Sensitivity to parameters.

Parameter	Description	Range	OI nominal	$F(\bar{\rho})$	$E[L^{\text{bust}}]$ (\$B/yr)
Baseline			1.51	0.50	189
ν	Production concavity	$[0.60, 0.78]$	[1.42, 1.62]	[0.55, 0.48]	[96, 377]
β	Sector returns to scale	$[0.40, 0.90]$	[2.99, 1.26]	[0.50, 0.51]	[75, 294]
θ	Early-deploy premium	$[0.70, 0.90]$	[1.51, 1.51]	[0.61, 0.46]	[159, 183]
ε	Contest head-start	$[0.10, 0.30]$	[1.51, 1.51]	[0.42, 0.58]	[72, 384]
ψ	Strategic benefit	$[0.20, 0.40]$	[1.51, 1.51]	[0.49, 0.51]	[158, 215]
α	Circular survival	$[0.30, 0.50]$	[1.51, 1.51]	[0.48, 0.50]	[168, 182]
ζ	Fire-sale loss, specificity	$[0.40, 0.60]$	[1.51, 1.51]	[0.52, 0.47]	[193, 151]
$\bar{Q}$	Market depth (tn)	$[0.8, 2.5]$	[1.51, 1.51]	[0.50, 0.52]	[191, 194]

Note: Sensitivity of the headline outcomes to parameters, each varied over a plausible range with other calibrated parameters held at their Table 1 values. OI is the nominal over-investment ratio, $F(\bar{\rho})$ the bust probability, and $E[L^{\text{bust}}]$ the expected destruction of committed capital across the sector. Each cell reports the outcome at the low and high ends of the range. For the ν and β rows, the productivity scale A, which is calibrated jointly with these two parameters, is adjusted at each value to hold the market size at baseline quantities fixed, so the rows vary the curvature rather than the size of the market.

Over-investment depends only on the production concavity, the sector returns to scale, and the number of coalitions. Across the build-out cost share it runs from 1.42 to 1.62. It climbs higher as the sector returns fall toward the central readings of the revenue and usage evidence in Appendix D, reaching about 3× at $\beta = 0.4$, and higher still as the field of coalitions widens (Figure 8, panel a).

The fire-sale enters through the market depth $\bar{Q}$. A thinner market deepens the bust but, because the coalition is a price-taker, barely changes its leverage, so the over-investment is invariant across $\bar{Q} \in [0.8, 2.5]$tn while expected destruction moves only between \$191B and

\$194B (the $\bar{Q}$ row). The headline is also robust to the internalisation convention. Price-taking is the standard treatment of a common-pool recovery price (Shleifer and Vishny, 1992; Lorenzoni, 2008; Bianchi, 2011). Were each coalition instead to internalise its full price impact, its realised over-investment would fall only from 1.40× to 1.31×.

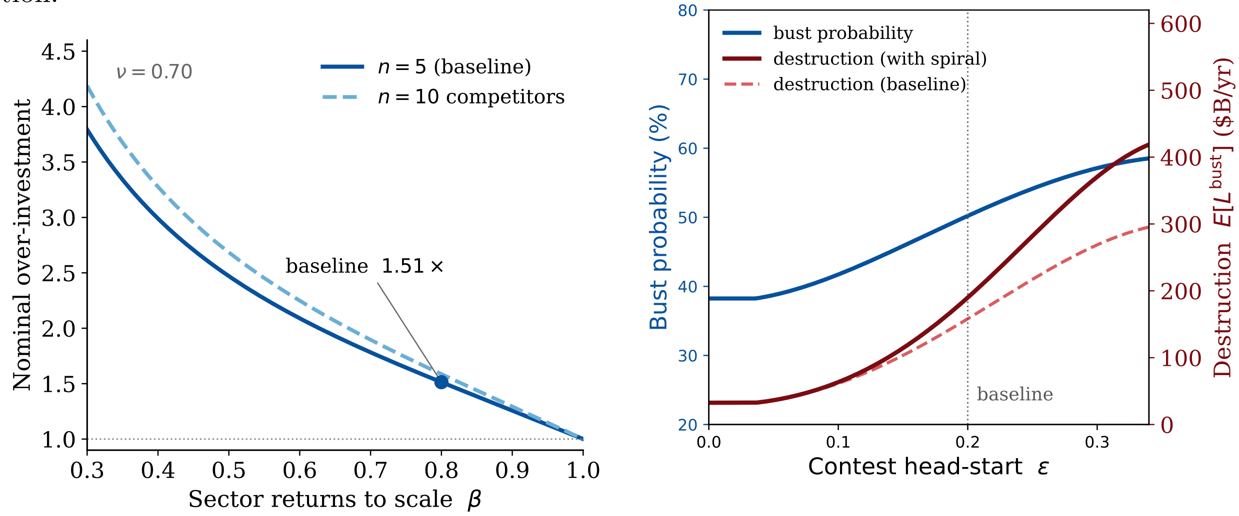
The contest head-start is the key determinant of fragility. It sets how far the sector front-loads, raising the period-1 share from about one half to three quarters. That committed capital is what the bust destroys, so the same lever drives the bust probability, from 42% to 58%, and expected destruction, from \$72bn to \$384bn (Figure 8, panel b). Remove the head-start and the boom-bust all but disappears. Strengthen it and the destruction more than quintuples. The fragility is a product of the race to commit early, not of the technology or the financing on their own.

The other parameters move the outcomes within a narrower band. A higher early-deploy premium makes late capacity more useful, so the sector sustains a boom at a lower productivity draw and the bust probability falls. The strategic benefit barely registers. The two bust parameters, the survival rate of circular capital and the fire-sale loss on debt, scale the destruction only modestly, because firms that face a harsher bust carry less committed debt into it.

The pattern across the table follows Corollary 1. The bust probability is insulated from the level forces: it stays within about two percentage points of one half across the debt-cost range $R_d \in [1.03, 1.10]$ and the synergy range $\psi \in [0, 0.6]$, within five across a field of $n \in \{3, 10\}$ coalitions, and a twenty-five per cent rise in the productivity scale A moves it by under seven. The sharpest case is the β row: the contest wedge swings from 2.20 to 1.09 across its range, and the bust probability stays within about two percentage points of one half. The timing forces are what move it, the head start from 0.42 to 0.58 and the early-deploy premium by fifteen points across their ranges. Even the number of coalitions works through the race rather than the wedge: a wider field raises the bust probability because the head-start premium scales with $(n-1)/n$, while the smaller wedge λ is frequency-neutral. The one asymmetry sits on the downside of fundamentals. With period-1 equity already committed, a productivity scale a quarter below the calibration raises the bust probability to 0.59, the equity-floor effect that makes commitment in excess of fundamentals, rather than scale itself, the fragility object.

Figure 8: Sensitivity of the headline outcomes

(a) Over-investment across returns and competition. (b) Fragility and the contest head-start.



Note: Panel (a) plots the nominal over-investment ratio against the sector returns to scale β, for the baseline field of five coalitions and for ten, holding the build-out cost share at $\nu = 0.70$. The dot marks the baseline calibration. The ratio depends only on (ν, β, n). It rises as returns fall and as the field widens, reaching three to four at the low-elasticity floor. Panel (b) plots the bust probability (solid, left axis) and expected destruction $E[L^{\text{bust}}]$ (dashed, right axis) against the contest head-start ε. Both climb steeply with the head-start, which sets how far the sector front-loads into committed, bust-exposed debt. The no-head-start benchmark and the baseline are marked.

Table 3: Robustness to circular share γ and the deployment rate c_1/C.

Scenario	OI	$F(\bar{\rho})$	$E[L^{\text{bust}}]$ ($B/yr)
Baseline	1.51	0.50	189
$\gamma = 0.18$ (-10%)	1.51	0.49	162
$\gamma = 0.22$ $(+10\%)$	1.51	0.50	194
$c_1/C = 0.64$ (-10%)	1.51	0.50	183
$c_1/C = 0.78$ $(+10\%)$	1.51	0.49	174

Note: Robustness of the headline outcomes to $\pm10\%$ perturbations of the two data anchors, the circular share γ and the deployment rate c_1/C, with the calibrated productivity scale held fixed. Over-investment is structurally invariant by Proposition 1. The bust probability moves by less than one percentage point, and expected destruction by under 15%.

The model charges one gross rate on all debt, although period-1 debt is outstanding longer. Repricing period-1 debt at the two-period compounded rate leaves the over-investment ratio of Proposition 1 intact at 1.51 and moves the bust probability only slightly from 0.50 to 0.48. The adjustments fall largely on levels: committed debt drops from \$92B to \$72B, expected destruction from \$189B to \$144B, and period-1 investment from \$280B to \$260B, with the planner benchmark unchanged as it commits no early debt. Beyond this asymmetry, the annual period is a normalisation. The anchors are per-period objects, so a longer period scales revenue flows and compounded rates together and leaves the investment margins unchanged. The exception is the early-deploy discount, a ratio tied to the lag length rather than to units. It deepens with longer periods, and its swept range covers readings of up to two years.

The two anchors are not knife-edge. A ten per cent change in the circular share or the deployment rate leaves over-investment untouched, moves the bust probability by less than one percentage point, and moves expected destruction by under 15% (Table 3). The headline magnitudes are properties of the mechanism, not of the two figures read off the deal record.

D Evidence on the demand curvature and the head start

This appendix records the public evidence behind the calibration of the two contest-side parameters, β and ε. It is deliberately minimal: every number is a company disclosure or a published industry series, combined only through the model's own algebra.[20]

The demand curvature β is bounded from above. Sector revenue is $\Theta = \rho A J^{\beta}$, so along the observed path $\Delta \ln \Theta = \Delta \ln(\rho A) + \beta\, \Delta \ln J$. Demand shifted upward over 2024–26, so $\Delta \ln(\rho A) \geq 0$ and

$$\beta \;\leq\; \frac{\Delta \ln \Theta}{\Delta \ln J}.$$

We proxy J with tokens processed and Θ with AI revenue. Google disclosed 480 tn tokens per month in May 2025, more than 1.3 quadrillion in October 2025 (a stated twenty-fold rise over a year) and 3.2 quadrillion in May 2026 (a seven-fold rise), while enterprise generative-AI spending tripled in 2025, from \$11.5 billion to \$37 billion, and the leading labs' run-rate revenues grew three-to-four-fold. Taking the fastest defensible revenue growth against the slowest defensible token growth gives $\beta \leq \ln 4/\ln 6.7 \approx 0.73$. Central readings give $\beta \leq \ln 3.2/\ln 20 \approx 0.39$. Both choices are biased against the conclusion, and raw tokens understate the growth of quality-adjusted compute services, which biases the bound up again. Prices corroborate the pattern: inference prices at fixed capability fell about ten-fold per year while total spending rose, the signature of demand expanding along a concave revenue schedule. The baseline $\beta = 0.80$ therefore sits above every reading of the bound, and since λ and over-investment rise as β falls, the headline magnitudes are lower bounds:

β	0.80 (baseline)	0.73	0.60	0.39
λ	1.20	1.30	1.53	≈ 2.3
OI (dollar)	1.51×	1.70×	2.09×	≈ 3.1×

The head start ε is bracketed by share persistence. The head start is the elasticity of contest share to committed period-1 capacity (Appendix B.2). Between 2023 and 2025 the enterprise API market flipped toward the smaller early committer: the leading lab's share fell from 50 to 27 per cent while its challenger's rose from 12 to 40 per cent, although the

[20] Token figures: Google I/O and Alphabet earnings disclosures (May 2025, October 2025, May 2026). Enterprise spending and shares: Menlo Ventures, *2025: The State of Generative AI in the Enterprise*, December 2025. Lab run rates: press disclosures compiled in Epoch AI's public revenue dataset. Fixed-capability prices: Appenzeller, *LLMflation*, a16z, November 2024. Consumer shares: Similarweb monthly trackers, May 2026. Full URLs and as-of dates are catalogued in the replication package.

leader's coalition held roughly two and a half to three times the committed compute. Read through the logit of Appendix B.2, the enterprise-segment ε is near zero. The consumer segment, by contrast, still rewarded incumbency with a share multiple of about six on the same commitment ratio, an implied elasticity well above one, though it eroded quickly over the year. The compute contest of the model is weighted toward the enterprise side, so the baseline $\varepsilon = 0.20$ sits in the low-middle of this bracket, and the sensitivity range $[0.1, 0.3]$ of Appendix C covers the calibration's neighbourhood within it. One moment the calibration does not target then falls into place: at $\varepsilon = 0.20$ the model's period-1 debt share is one third, inside the 30–40 per cent that bond issuance and lease disclosures imply for the observed build-out.

References

Acemoglu, D. (2024). The simple macroeconomics of AI. *NBER Working Paper* 32487.

Acemoglu, D., A. Ozdaglar, and A. Tahbaz-Salehi (2015). Systemic risk and stability in financial networks. *American Economic Review*, 105(2):564–608.

Adrian, T. and N. Boyarchenko (2015). Intermediary leverage cycles and financial stability. Federal Reserve Bank of New York Staff Reports 567, August 2012, revised February 2015.

Aghion, P. and P. Howitt (1992). A model of growth through creative destruction. *Econometrica*, 60(2):323–351.

Akcigit, U. and W. R. Kerr (2018). Growth through heterogeneous innovations. *Journal of Political Economy*, 126(4):1374–1443.

Aldasoro, I., S. Doerr, and D. Rees (2026). Financing the AI boom: from cash flows to debt. *BIS Bulletin* 120, January.

Allen, F. and D. Gale (2000). Financial contagion. *Journal of Political Economy*, 108(1):1–33.

Allen, F., A. Babus, and E. Carletti (2012). Asset commonality, debt maturity and systemic risk. *Journal of Financial Economics*, 104(3):519–534.

Bank for International Settlements (2026). Progress and peril. *Annual Economic Report 2026*, Chapter I, June.

Bernanke, B. and M. Gertler (1989). Agency costs, net worth, and business fluctuations.*American Economic Review*, 79(1):14–31.

Bianchi, J. (2011). Overborrowing and systemic externalities in the business cycle. *American Economic Review*, 101(7):3400–3426.

Bolton, P. and D. Scharfstein (1990). A theory of predation based on agency problems. *American Economic Review*, 80(1):93–106.

Brander, J. A. and T. R. Lewis (1986). Oligopoly and financial structure. *American Economic Review*, 76(5):956–970.

Brunnermeier, M. and L. Pedersen (2009). Market liquidity and funding liquidity. *Review of Financial Studies*, 22(6):2201–2238.

Caballero, R. J. (2026). Speculative growth and the AI "bubble". *NBER Working Paper* 34722.

Caballero, R. J. and E. Farhi (2018). The safety trap. *Review of Economic Studies*, 85(1):223–274.

Caballero, R. J. and A. Simsek (2013). Fire sales in a model of complexity. *Journal of Finance*, 68(6):2549–2587.

Cabrales, A., P. Gottardi, and F. Vega-Redondo (2017). Risk-sharing and contagion in networks. *Review of Financial Studies*, 30(9):3086–3127.

Clark, D. J. and C. Riis (1996). A multi-winner nested rent-seeking contest. *Public Choice*, 87(1–2):177–184.

Crouzet, N. and J. C. Eberly (2021). Intangibles, markups, and the measurement of productivity growth. *Journal of Monetary Economics*, 124:S92–S109.

Eisfeldt, A. L., G. Schubert, and M. B. Zhang (2023). Generative AI and firm values. *NBER Working Paper* 31222.

Elliott, M., B. Golub, and M. Jackson (2014). Financial networks and contagion. *American Economic Review*, 104(10):3115–3153.

Eren, E., I. Krohn, and K. Todorov (2026). Financing the AI infrastructure boom: on- and off-balance sheet borrowing. *BIS Quarterly Review*, March.

Geanakoplos, J. (2010). The leverage cycle. *NBER Macroeconomics Annual*, 24:1–65.

Gertler, M. and N. Kiyotaki (2010). Financial intermediation and credit policy in business cycle analysis. *Handbook of Monetary Economics*, 3:547–599.

Gilbert, R. J. and D. M. G. Newbery (1982). Preemptive patenting and the persistence of monopoly. *American Economic Review*, 72(3):514–526.

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

Holmström, B. and J. Tirole (1997). Financial intermediation, loanable funds, and the real sector. *Quarterly Journal of Economics*, 112(3):663–691.

Huang, J. (2025). Remarks on end-of-decade AI infrastructure investment. NVIDIA FY2026 second-quarter earnings call, August 2025.

Kiyotaki, N. and J. Moore (1997). Credit cycles. *Journal of Political Economy*, 105(2):211–248.

Konrad, K. A. (2009). *Strategy and Dynamics in Contests*. Oxford University Press.

Korinek, A. and J. Vipra (2025). Concentrating intelligence: Scaling and market structure in artificial intelligence. *Economic Policy*, 40(121):225–256.

Leland, H. E. (1994). Corporate debt value, bond covenants, and optimal capital structure. *Journal of Finance*, 49(4):1213–1252.

Lorenzoni, G. (2008). Inefficient credit booms. *Review of Economic Studies*, 75(3):809–833.

Loury, G. (1979). Market structure and innovation. *Quarterly Journal of Economics*, 93(3):395–410.

Mankiw, N. G. and M. D. Whinston (1986). Free entry and social inefficiency. *RAND Journal of Economics*, 17(1):48–58.

Reinganum, J. F. (1989). The timing of innovation: Research, development, and diffusion. In R. Schmalensee and R. D. Willig (Eds.), *Handbook of Industrial Organization*, vol. 1, pp. 849–908. North-Holland.

Shleifer, A. and R. W. Vishny (1992). Liquidation values and debt capacity: A market equilibrium approach. *Journal of Finance*, 47(4):1343–1366.

Tirole, J. (2006). *The Theory of Corporate Finance*. Princeton University Press.

Tullock, G. (1980). Efficient rent seeking. In J. M. Buchanan, R. D. Tollison, and G. Tullock (Eds.), *Toward a Theory of the Rent-Seeking Society*, pp. 97–112. Texas A&M University Press.

Van Nieuwerburgh, S. (2026). Financing the AI buildout. *Journal of Economic Perspectives*, forthcoming.

Wachter, J. and J. Wachter (2026). What investment data implies about the AI transition. *NBER Working Paper* 35290.