# The Macroeconomics of AI Capacity: Insights from a Two-Asset Growth Model

Jonathan Rice[1]

February 1, 2026

**Abstract**

Specialised AI hardware depreciates far faster than conventional capital. This paper embeds that distinction in a growth model with two capital stocks: conventional capital and AI-related compute capacity. Because replacement investment in short-lived hardware is large, even small stock misalignments can be corrected quickly by adjusting gross investment. Calibrated to advanced economies in the mid-2020s, the model helps rationalise boom-bust patterns in data-centre spending. Macro outcomes are sensitive to the rate at which hardware loses value: a two-percentage-point reduction in quarterly depreciation raises welfare by 0.36% in consumption-equivalent terms, while an equal-sized compute tax shrinks the AI stock by around one-fifth.

**JEL codes:** E22, E25, O33, O41

[1]European Systemic Risk Board (ESRB). The views expressed are the author's and do not necessarily reflect those of the European Systemic Risk Board.

# 1 Introduction

The marginal cost of querying a trained large-language model has fallen sharply since the release of ChatGPT in late 2022. At the same time, rapid product cycles in specialised AI hardware (such as GPUs) render installed capacity economically obsolete much faster than conventional capital. This combination of high fixed costs, low marginal usage costs and accelerated obsolescence places AI-related assets in a distinct category for macroeconomics. Standard one-capital frameworks struggle to capture the associated investment dynamics and distributional effects: they cannot explain, for example, why surges in specialised-hardware spending unwind over a relatively short span of time, or how an influx of low-cost digital services can materially shift consumption patterns and welfare while leaving the aggregate labour share largely unchanged. In the baseline Cobb-Douglas specification, the aggregate labour share is fixed by the chosen exponents rather than being an outcome of AI adoption (holding technology shares fixed).

The AI Index (2025) reports that the price of querying a model achieving GPT-3.5-level performance fell from about \$20 to \$0.07 per million tokens between November 2022 and October 2024.[2] At a usage price of \$0.07 per million tokens, a 700-token exchange costs roughly 0.005 cents. These usage costs are of the same order as widely cited estimates of a web search's electricity use (around 0.3 Wh) and recent measurements of median energy use for a Gemini text prompt (0.24 Wh), although such figures vary by workload and methodology.[3] API prices differ widely by model and date: low-cost tiers such as Gemini Flash are now only a small multiple of benchmark resource costs, whereas frontier multimodal models remain substantially more expensive.[4] On the hardware side, capital outlays on accelerators and data-centre infrastructure have followed boom-bust patterns tied to semiconductor fabrication cycles, and accounting lives at the main hyperscalers have been revised repeatedly in response to these developments.

In this paper, we construct a neoclassical growth model in which a stock of AI capacity joins traditional capital as a second reproducible asset. "AI capacity" is interpreted as the effective installed compute needed to deploy existing models: a bundle of specialised hardware and tightly coupled software whose depreciation is dominated by hardware obsolescence and intensive utilisation. The model weights themselves are non-rival, but installed capacity is rival and congestible. Once installed, AI capacity delivers digital services at low (but

[2] Stanford Institute for Human-Centered Artificial Intelligence (2025), Section 5.1, Figure 5.3.
[3] Hölzle (2009); Elsworth et al. (2025); Google Cloud (2025).
[4] Google AI (2025); OpenAI (2025).

non-zero) marginal resource cost and, separately, raises productivity in the tangible-goods sector. On the demand side, the consumption bundle is a constant-elasticity (CES) aggregate of tangible goods and digital services. On the supply side, production is Cobb-Douglas with an additional AI factor that enters with its own output elasticity; the baseline does not assume direct labour-augmentation.

The allocation can be described by a social planner who internalises the capacity constraint and the resource cost of inference. Two user-cost conditions characterise the steady state analytically. The condition for conventional capital pins down its user cost and yields a closed-form expression for the steady-state stock of K as a power function of the AI stock. Substituting this expression into the Euler equation for AI capacity reduces the problem to a single marginal-benefit map in one variable. Appendix A shows that this map is strictly decreasing under mild assumptions on congestion and depreciation, which guarantees a unique interior solution for the steady-state AI stock.

To study dynamics, we log-linearise the Euler block together with the capital-accumulation equations around the steady state, keeping investment endogenous in the feasibility constraint. Because the co-state variable depends on both current and next-period capital stocks, the linearised Euler system is second order. Forming a companion system and selecting the stable eigenspace yields a reduced two-dimensional law of motion for the log deviations of $K$ and $M$. As a transparent comparison, we construct partial equilibrium (PE) benchmarks that hold investment fixed at steady-state replacement levels, so that stock deviations decay passively at the survival rates $(1 - \delta)$ and $(1 - \delta_M)$. These benchmarks isolate the role of hardware durability: in the baseline calibration, passive decay alone implies a half-life of roughly seven quarters for AI capacity versus about eleven years for conventional capital, a six-to-one gap driven entirely by the depreciation differential. The general equilibrium (GE) responses diverge from these benchmarks because the planner re-optimises investment. For short-lived AI hardware, replacement flows are large relative to small stock deviations, so the planner can close gaps much faster than passive decay by modestly adjusting gross investment. For diffusion from a low initial stock, the planner smooths the build-up over a longer horizon than the PE benchmark would suggest. Figures 1–3 display both paths, making visible the distinct roles of passive depreciation and active investment adjustment.

The analytical characterisation highlights which parameters matter most. The effective depreciation rate of AI hardware governs the size of the optimal AI stock and the half-life of an investment surge; the elasticity of substitution between digital and tangible uses controls the share of expenditure captured by low-cost digital services; and the output elasticity of AI

capacity determines how strongly those services feed back into aggregate output and incomes. Because the steady state and the GE dynamics can be written directly in terms of these primitives, a wide range of policy interventions can be evaluated using steady-state elasticities and linearised impulse responses rather than full numerical solution methods.

Two quantitative findings stand out. First, extending hardware life produces a disproportionately large increase in the steady-state AI stock and in the digital share of consumption, yet, unless AI is allowed to substitute directly for labour in production, the aggregate labour share remains close to the Cobb-Douglas benchmark. In the baseline calibration, raising the annual economic life of AI hardware from about three to four years increases the optimal AI-capacity stock by several tens of per cent and raises the digital expenditure share materially, while the labour share moves little. Second, the high depreciation rate typical of accelerators creates a sharp asymmetry between the GE and PE adjustment speeds. When AI hardware is hit by a positive stock shock, the planner can eliminate most of the excess within a few quarters by cutting gross investment modestly below replacement, because steady-state replacement flows ($\delta_M M^* \approx 9.5\%$ of the stock per quarter) are large relative to small deviations. The PE passive-decay benchmark, which abstracts from this behavioural response, implies a half-life of roughly seven quarters and provides an intuitive yardstick for the speed at which hardware-driven booms unwind once investment returns to replacement. These distinct timescales align with the volatility observed in data-centre expansions and suggest tight windows for macroprudential intervention.

The policy experiments exploit the mapping from hardware costs into the primitives of the model. A hardware tax or a regulatory surcharge on compute can be represented as an increase in the effective depreciation rate of AI capacity, while accelerated-depreciation allowances, repair mandates or durability-enhancing subsidies operate in the opposite direction. In the baseline calibration, a two-percentage-point surtax on the quarterly economic depreciation rate of AI capacity reduces the optimal AI stock by around one fifth, whereas a symmetric improvement in hardware longevity raises lifetime welfare by around 0.36 per cent in consumption-equivalent units. Because replacement flows for AI hardware are large, the transition to a new steady state is rapid: most of the welfare gain from a longevity improvement accrues within a few years.

Finally, the framework accommodates several extensions at low analytical cost. A simple nested labour aggregator maps growth in AI capacity into wage premia between routine and non-routine workers. Embedding the consumption aggregator in a Calvo pricing environment yields a New Keynesian Phillips curve in which hardware-cost shocks act as cost-push

disturbances whose persistence is tied to the durability of AI capacity. Introducing an additional state variable capturing model size separates the training and inference margins without disturbing the two-asset core. In each case, the key advantage of the present setup is that the capital block remains two-dimensional and analytically characterised, so that the extra margins can be layered on without losing tractability.

The remainder of the paper is organised as follows. Section 2 reviews related work. Section 3 presents the environment and the planner's allocation problem. Section 4 characterises the steady state and its comparative statics. Section 5 describes the calibration. Section 6 analyses the GE adjustment dynamics and compares them with PE benchmarks. Section 7 studies policy and measurement. Section 8 discusses extensions and robustness. Section 9 concludes.

## 2 Literature Review

Macroeconomic analysis of artificial intelligence draws on three interconnected strands of research: (i) task-based automation and its labour-market consequences, (ii) the accumulation and measurement of intangible capital, and (iii) the empirical diffusion of robotics and foundation models. This section situates the present framework within those literatures and clarifies its distinctive contribution.

The potential for capital to substitute for labour in specific tasks, thereby influencing aggregate factor shares, has long been recognised. **?** showed how sequential task automation can compress the wage share unless the economy continually creates new labour-complementary activities. Building on that insight, **?** develop a tractable production function that combines displacement and reinstatement of labour, while subsequent studies explore general equilibrium consequences of software-driven automation. Empirical work corroborates these mechanisms: exposure to industrial robots lowers employment and wages in US commuting zones (**?**), with analogous displacement effects documented across European regions (**?**). A common premise in much of this research, however, is that automated services are rivalrous, so their marginal production cost is comparable to the labour they replace. This paper instead models AI capacity as a high-depreciation capital asset: the underlying model weights are non-rival, but the installed hardware and associated infrastructure required to deploy them are rival and congestible, delivering very low marginal-cost services once in place.

That assumption links AI to the economic properties of intangible assets (software, designs, data) that can be replicated without diminishing their availability. Counting such spending

as investment rather than intermediate expense substantially alters measurements of capital deepening and productivity growth (**?**). Later work by **?** emphasises the scalability and non-rivalry of intangibles, helping to explain both the declining labour share (**?**) and the rise of "superstar" firms. Yet while a subset of two-capital DSGE and growth models introduces separate tangible and intangible (or knowledge) stocks, they rarely admit analytically transparent dynamics and instead rely on balanced-growth characterisations or numerical solution methods (e.g. **????**). Furthermore, these models typically characterise "knowledge capital" as depreciating slowly, which contrasts sharply with the rapid obsolescence of AI hardware. By contrast, the present two-capital framework keeps the steady state analytically characterised even when the two assets depreciate at very different rates: the user-cost conditions reduce the problem to a single monotone marginal-benefit map for AI capacity, and conventional capital is an explicit power function of the AI stock. Log-linearising the Euler equations and the accumulation rules around this steady state yields a two-dimensional system with interpretable eigenvalues, so that adjustment speeds and policy semi-elasticities can be expressed directly in terms of a small number of technology and preference parameters.

The empirical motivation is clear. Global industrial-robot density more than doubled, from 69 to 151 robots per 10,000 manufacturing workers, between 2015 and 2022 (**?**). On the AI side, the **?** (**?**) AI Index reports that the cost of querying a model delivering GPT-3.5-level performance fell from roughly $20 to $0.07 per million tokens within roughly two years (Nov 2022-Oct 2024). Capital outlays on specialised hardware such as GPUs follow similar boom-bust patterns tied to semiconductor fabrication cycles. The model's short-run investment dynamics provide a theoretical counterpart to these rapid swings, linking their half-life to hardware obsolescence.

On the policy front, accelerated-depreciation allowances can tilt incentives toward automation (**?**), and intangible booms may require counter-cyclical buffers to address financial-stability risks in AI-hardware lending. Hence, bridging intangible capital with high depreciation is of direct policy relevance.

In sum, while prior studies highlight rich micro evidence on automation and intangible capital, few embed AI's low marginal cost and rapid hardware decay in a transparent general equilibrium model. By doing so in an analytically tractable two-asset framework (with a steady state characterised by user-cost conditions and linear GE dynamics that admit clear eigenvalue interpretations) this paper provides simple mappings from technology primitives to macroeconomic quantities and policy-relevant welfare measures.

# 3 The Model

The economy lives in discrete time $t = 0, 1, 2, \ldots$ and is populated by a representative household and perfectly competitive firms. It features two reproducible assets: conventional physical capital $K_t$ and a stock of AI capacity $M_t$. The former depreciates slowly, the latter much faster. We interpret $M_t$ as effective installed AI capacity (specialised hardware and closely coupled infrastructure required to deploy existing models) whose economic depreciation is dominated by hardware obsolescence and intensive utilisation. Once installed, this capacity delivers digital services at low (but non-zero) marginal resource cost and, separately, raises the productivity of the tangible-goods sector.

**Preferences and labour** Household welfare is

$$\mathcal{U}_0 = \sum_{t=0}^{\infty} \beta^t \ln C_t, \qquad 0 < \beta < 1, \tag{1}$$

where $C_t$ is a composite consumption index. To preserve the analytical focus on the capital block, we assume inelastic labour supply and normalise the endowment:

$$L_t \equiv 1 \quad \text{for all } t.$$

(Section 8 discusses an extension with endogenous labour.) Composite consumption aggregates tangible goods $C_t^{\mathrm{T}}$ and digital services $D_t$ via

$$C_t = \left[\theta \left(C_t^{\mathrm{T}}\right)^{\frac{\sigma-1}{\sigma}} + (1-\theta)\, D_t^{\frac{\sigma-1}{\sigma}}\right]^{\frac{\sigma}{\sigma-1}}, \qquad 0 < \theta < 1,\ \sigma > 0, \tag{2}$$

with $\sigma$ the elasticity of substitution between tangible and digital streams.

**Technology** Output of the tangible-goods sector is

$$Y_t = A\, K_t^{\alpha}\, M_t^{\zeta}\, L_t^{1-\alpha-\zeta}, \qquad 0 < \alpha + \zeta < 1, \tag{3}$$

so the reproducible AI stock enters as a factor of production with elasticity $\zeta$. Because the Cobb-Douglas exponents are primitive, the wage share equals $1 - \alpha - \zeta$ and is invariant to movements in hardware costs or preferences.

AI capacity also enables a flow of digital services subject to capacity:

$$0 \le D_t \le \psi\, M_t^{\phi}, \qquad 0 < \phi \le 1,\ \psi > 0, \tag{4}$$

while each unit of $D_t$ requires tangible resources (electricity, cooling, supervision) at marginal cost $\chi > 0$.

**Resource constraint and capital accumulation** Let $I_t^K$ and $X_t$ denote gross investment in $K_t$ and $M_t$, respectively. Goods feasibility is

$$C_t^{\mathrm{T}} + I_t^K + X_t + \chi\, D_t = Y_t, \tag{5}$$

and the two assets evolve according to

$$K_{t+1} = (1-\delta)\, K_t + I_t^K, \qquad 0 < \delta < 1, \tag{6}$$

$$M_{t+1} = (1-\delta_M)\, M_t + X_t, \qquad 0 < \delta_M < 1. \tag{7}$$

**Planner's problem, multipliers and optimality conditions** Because markets are competitive and there are no externalities, the decentralised allocation coincides with the solution to a social planner who chooses $\{C_t^{\mathrm{T}}, D_t, I_t^K, X_t\}_{t\geq 0}$ to maximise (1) subject to (2)–(7) and given $(K_0, M_0)$. Let $\lambda_t$ be the current-value multiplier on (5) (goods units) and $\mu_t \geq 0$ the multiplier on the capacity constraint in (4). The FOC with respect to tangible consumption implies

$$\lambda_t \;=\; \frac{\partial \mathcal{U}_t}{\partial C_t}\,\frac{\partial C_t}{\partial C_t^{\mathrm{T}}} \;=\; \frac{1}{C_t}\,\frac{\partial C_t}{\partial C_t^{\mathrm{T}}}.$$

Define the intratemporal marginal rate of substitution

$$\mathrm{MRS}_t \equiv \frac{\partial C_t/\partial D_t}{\partial C_t/\partial C_t^{\mathrm{T}}} = \frac{1-\theta}{\theta}\left(\frac{C_t^{\mathrm{T}}}{D_t}\right)^{1/\sigma}.$$

The Kuhn-Tucker conditions yield:

**(i) Consumption-services mix with capacity**

$$\mathrm{MRS}_t \begin{cases} = \;\chi, & \text{if capacity is slack } \;(D_t < \psi M_t^{\phi},\; \mu_t = 0), \\ \geq\; \chi, & \text{if capacity binds } \;(D_t = \psi M_t^{\phi},\; \mu_t \geq 0). \end{cases} \tag{8}$$

**(ii) Euler equation for tangible capital**

$$\lambda_t = \beta\, \lambda_{t+1}\Big[(1-\delta) + \alpha\,\frac{Y_{t+1}}{K_{t+1}}\Big]. \tag{9}$$

**(iii) Euler equation for AI capacity** A unit of $M_{t+1}$ survives with probability $(1-\delta_M)$ and yields two dividends at $t+1$: a production dividend inside (3) and a services dividend equal to the net value of the induced extra flow of $D_{t+1}$ beyond its resource cost $\chi$. The envelope condition is

$$\lambda_t = \beta\,\lambda_{t+1}\Big[(1-\delta_M) + \zeta\,\frac{Y_{t+1}}{M_{t+1}} + \max\{0, \mathrm{MRS}_{t+1} - \chi\}\;\phi\,\psi\,M_{t+1}^{\phi-1}\Big]. \tag{10}$$

The services term vanishes under slack capacity ($\mathrm{MRS}_t = \chi$) and is positive when capacity binds ($\mathrm{MRS}_t > \chi$).

**Focus of the analysis (regime)** Given the high productivity of AI services and the specialised hardware required, we focus on the empirically relevant binding-capacity regime:

$$D_t = \psi M_t^{\phi} \quad \text{and} \quad \mathrm{MRS}_t > \chi$$

for the baseline results. Slack-capacity formulas are collected in the appendix, and the two regimes yield piecewise-defined local dynamics when the Euler system is linearised.

## Steady state

Let $K_{t+1} = K_t \equiv K^*$, $M_{t+1} = M_t \equiv M^*$ and $D^* = \psi(M^*)^{\phi}$ (binding). The Euler conditions become

$$1 = \beta\Big[(1-\delta) + \alpha\,\frac{Y^*}{K^*}\Big], \tag{11}$$

$$1 = \beta\Big[(1-\delta_M) + \zeta\,\frac{Y^*}{M^*} + \big(\mathrm{MRS}^* - \chi\big)\,\phi\,\psi\,(M^*)^{\phi-1}\Big], \tag{12}$$

with $Y^* = A\,(K^*)^{\alpha}\,(M^*)^{\zeta}$ (using $L_t = 1$) and

$$\mathrm{MRS}^* = \frac{1-\theta}{\theta}\left(\frac{C^{\mathrm{T}*}}{D^*}\right)^{1/\sigma}, \qquad C^{\mathrm{T}*} = Y^* - \delta K^* - \delta_M M^* - \chi D^*.$$

In the binding-capacity regime, the steady state is the solution to the simultaneous system composed of (11)–(12), the feasibility condition above, the production function (3) and the definition of $\mathrm{MRS}^*$. The first equality pins down the user cost of tangible capital; the second equates the price of one unit of AI capacity to the present value of its resale plus dividends. The key property for uniqueness is that the right-hand side of (12), the discounted marginal benefit of capacity, is strictly decreasing in $M^*$. As shown formally in Appendix A, this

monotonicity is preserved once the general-equilibrium adjustment of $\text{MRS}^*$ is accounted for, guaranteeing a unique interior solution.[5]

**Comparative statics (qualitative)** Lower $\chi$ weakly increases the likelihood that capacity binds and raises the services dividend in (12); a lower $\delta_M$ or higher $\zeta$ increases the production dividend. In all three cases the optimal $M^*$ rises. (Closed-form elasticities reported later are computed under binding capacity and rely on the appendix monotonicity result.)

## Local dynamics

Let $k_t \equiv \ln K_t - \ln K^*$ and $m_t \equiv \ln M_t - \ln M^*$. Log-linearising the Euler block together with (6)–(7) around the steady state yields a two-dimensional linear system

$$\begin{bmatrix} k_{t+1} \\ m_{t+1} \end{bmatrix} = J \begin{bmatrix} k_t \\ m_t \end{bmatrix}, \tag{13}$$

where $J$ is the Jacobian of the law of motion for $(K_t, M_t)$ evaluated at $(K^*, M^*)$. Each entry of $J$ is a unit-free elasticity built from $(\alpha, \zeta, \sigma, \theta, \delta, \delta_M, \phi, \psi, \chi)$ and steady-state ratios. The coefficients are regime-dependent: (i) if capacity is slack ($\text{MRS}_t = \chi$), the services term envelopes out and the off-diagonal elements reflect only the production channel (proportional to $\alpha\zeta$); the system becomes block-diagonal only when, in addition, $\zeta = 0$. (ii) Under binding capacity ($\text{MRS}_t > \chi$), an extra term proportional to $(\text{MRS}^* - \chi)\,\phi$ enters the $M$ equation. Appendix B describes the construction and sign structure of $J$ in more detail. Section 6 uses the eigenvalues of $J$ to characterise general-equilibrium adjustment speeds and to compare them with the simpler partial-equilibrium geometric benchmarks driven directly by the depreciation rates $(\delta, \delta_M)$.

**Remark (rivalry and interpretation)** AI capacity $M_t$ represents specialised hardware and associated infrastructure, which are rival inputs. The model abstracts from the micro-allocation of that compute between two uses, raising productivity in tangible production (Eq. (3)) and serving digital demand (Eq. (4)), by allowing the same stock to play both roles at the macro level. This abstraction preserves the tractability of the two-asset core while capturing the dual economic role of AI capital. A variant that explicitly allocates hardware across uses, $M_t = M_t^Y + M_t^D$ with $D_t \leq \psi(M_t^D)^\phi$, equalises marginal values across uses within the period and leaves the state dimension unchanged; its algebraic details are available on

[5] Appendix A derives conditions (satisfied by the baseline calibration) under which $M^* \mapsto (1-\delta_M) + \zeta Y^*/M^* + (\text{MRS}^* - \chi)\phi\psi(M^*)^{\phi-1}$ is strictly decreasing, even though $\text{MRS}^*$ is endogenous.

request.

# 4 Steady-state allocation and comparative statics

In the binding-capacity regime ($D_t = \psi M_t^{\phi}$), the steady state is characterised by the two user-cost equalities

$$1 = \beta\Big[(1-\delta) + \alpha\,\frac{Y^*}{K^*}\Big], \tag{14}$$

$$1 = \beta\Big[(1-\delta_M) + \zeta\,\frac{Y^*}{M^*} + \big(\mathrm{MRS}^* - \chi\big)\,\phi\psi\,(M^*)^{\phi-1}\Big], \tag{15}$$

where $D^* = \psi\,(M^*)^{\phi}$ and

$$\mathrm{MRS}^* = \frac{1-\theta}{\theta}\left(\frac{C^{\mathrm{T}*}}{D^*}\right)^{1/\sigma}, \qquad C^{\mathrm{T}*} = Y^* - \delta K^* - \delta_M M^* - \chi D^*.$$

Output is $Y^* = A\,(K^*)^{\alpha}\,(M^*)^{\zeta}$ (using $L = 1$).

**Analytical mapping for $K^*$** Equation (14) pins down the user cost of tangible capital and implies a simple power-law relationship between $K^*$ and $M^*$:

$$K^* = \Big[\Xi\,(M^*)^{\zeta}\Big]^{1/(1-\alpha)}, \qquad \Xi = \frac{\alpha\,A}{\beta^{-1} - (1-\delta)}. \tag{16}$$

Hence conventional capital is an explicit function of the AI stock: once $M^*$ is known, both $K^*$ and $Y^*$ follow immediately from (16) and the production function. Substituting (16) into (15) collapses the steady-state problem to a single scalar equation in $M^*$, where the right-hand side (the discounted marginal benefit of capacity, comprising resale, production, and net services) is strictly decreasing in $M^*$. As shown in Appendix A, this monotonicity is preserved even when the endogenous term $\mathrm{MRS}^*$ adjusts with $M^*$, guaranteeing a unique interior solution for AI capacity. In this sense the steady state is analytically characterised: one user-cost condition delivers the explicit mapping $K^*(M^*)$, and the other determines $M^*$ as the unique root of a monotone marginal-benefit map.

**Factor-income shares** Under perfect competition the real wage equals the marginal product of labour, $w^* = (1-\alpha-\zeta)\,Y^*$. Measured against value added $Y^*$ the labour share is therefore

$$s_L^* = 1 - \alpha - \zeta. \tag{17}$$

Because the Cobb-Douglas exponents $(\alpha, \zeta)$ are treated as primitive technology parameters, $s_L^*$ is invariant to movements in $\delta_M$, $\chi$ or $\sigma$ in the baseline model. Put differently, the constancy of the aggregate labour share is an assumption built into the production structure: shifts in hardware longevity, inference costs or substitution elasticities affect the scale of AI capital and the composition of expenditure, but do not alter the wage share unless $\zeta$ itself is allowed to vary. Extensions in which the effective output elasticity of AI capacity evolves with $M^*$ (for example via organisational capital or learning-by-doing) would naturally generate time-varying labour shares, but those mechanisms are kept outside the baseline for tractability.

**Comparative statics (signs)** Let $\mathcal{S} \equiv \{K^*, M^*, Y^*, D^*\}$. Differentiating the steady-state system implicitly and using the monotonicity result in Appendix A yields unambiguous signs for the key technology primitives:

$$\frac{\partial M^*}{\partial \delta_M} < 0, \qquad \frac{\partial M^*}{\partial \chi} < 0, \qquad \frac{\partial M^*}{\partial \zeta} > 0.$$

A lower economic depreciation rate for AI hardware $(\delta_M)$ or a higher output elasticity of AI capacity $(\zeta)$ both increase the marginal benefit of holding $M$, raising the steady-state AI stock. A higher marginal resource cost of digital services $(\chi)$ reduces the net services dividend in (15) and therefore lowers $M^*$.

Changes in the elasticity of substitution $\sigma$ operate only through $\text{MRS}^*$ and thus have an a priori ambiguous effect on $M^*$; the sign depends on the steady-state ratio $C^{\text{T}*}/D^*$ and the calibration of $(\theta, \sigma)$. Once $M^*$ is determined, $K^*$ follows from (16) and $Y^*$ from the production function, while $D^*$ is pinned down by capacity, $D^* = \psi(M^*)^\phi$.

**Interpretation** Hardware longevity $(\delta_M)$ and the marginal resource cost of inference $(\chi)$ are technological primitives; together with the production elasticity $\zeta$ they scale the optimal stock of AI capacity and, via the CES aggregator, the steady-state share of digital services in expenditure. Increases in longevity or productivity push up $M^*$, $K^*$ and $Y^*$, and raise the digital share, whereas higher marginal resource costs pull in the opposite direction. None of these movements, however, affects the labour share $1 - \alpha - \zeta$ in (17) as long as the Cobb-Douglas exponents are fixed. The next section calibrates the parameters to recent evidence and reports the implied steady-state levels, digital expenditure shares and the ratios that enter the dynamics and policy analysis.

# 5 Calibration

We calibrate the model to advanced economies circa mid-2025, with particular attention to the AI-capacity block. The model is solved at a quarterly frequency, with period parameters obtained from annual targets.[6] Throughout, we adopt the empirically relevant binding-capacity regime ($D_t = \psi M_t^{\phi}$ with $\text{MRS}_t > \chi$), so the shadow price of digital services $p_D \equiv \text{MRS}_t$ enters the Euler condition for $M_t$ and the steady state. As in Section 3, the labour endowment and technology level are normalised to $L_t = 1$ and $A = 1$. For the baseline parameter vector we verify ex post that the steady state indeed lies in the binding regime (Appendix A).

## 5.1 Foundational macro parameters

The annual discount factor $\beta_a$ targets a long-run real interest rate of 1% per year, $\beta_a = 1/1.01 \approx 0.99$, consistent with official US long-run assessments in the low single digits.[7] This implies a quarterly discount factor

$$\beta = \beta_a^{1/4} \approx 0.9975.$$

The tangible capital share is fixed at $\alpha = 0.30$, a conventional value in growth/RBC calibrations and consistent with PWT-based labour-share evidence for advanced economies.[8] For the depreciation rate of tangible capital we set $\delta_a = 0.06$ annually, in line with implied geometric rates across private fixed assets using BEA fixed-asset accounts and related methodology notes.[9] This implies a quarterly rate

$$\delta = 1 - (1 - \delta_a)^{1/4} \approx 0.0154.$$

## 5.2 AI-capacity block

**Economic depreciation of AI capacity** We choose an annual rate $\delta_{M,a} = 0.33$ (an economic service life of roughly three years) to capture the rapid obsolescence and heavy

[6] The conversion from annual rates (subscript $a$) to quarterly rates follows standard conventions: $\beta = \beta_a^{1/4}$ for the discount factor, and $\delta_i = 1 - (1 - \delta_{i,a})^{1/4}$ for depreciation rates $i \in \{K, M\}$.

[7] See, for example, Congressional Research Service (2025) on long-run real-rate assumptions in policy models.

[8] See Feenstra, Inklaar and Timmer (2015) and recent reviews of labour shares in advanced economies.

[9] See BEA Fixed Assets Accounts and BLS/OECD discussions of geometric depreciation; the implied aggregate rates across structures and equipment typically lie in the 5-7% range.

utilisation of specialised accelerators. This is an economic rate, distinct from accounting lives, which the major hyperscalers have extended to six years for servers (and networking) in 2023-2024.[10]

Market data strongly supports this rapid decay. Historical resale pricing for the Nvidia V100 (2017–2020 cycle) indicates a value loss of approximately 45% within two years of release. More recently, spot market rental rates for H100 GPUs, a direct proxy for the asset's economic yield, declined by roughly 50% in the first eight months of 2024 alone (**?**). This accelerated depreciation is structurally driven by the "Red Queen" effect of innovation: as GPU price-performance doubles approximately every 2.5 years, older vintages become energy-insolvent for frontier tasks, forcing a rapid collapse in their economic value (**?**).

The quarterly depreciation rate used in the model is

$$\delta_M = 1 - (1 - \delta_{M,a})^{1/4} \approx 0.0953,$$

so the quarterly survival rate of installed AI capacity is about $1 - \delta_M \approx 0.905$. This survival rate defines the PE passive-decay benchmark against which the GE dynamics are compared in Section 6. It also governs the steady-state replacement flow $\delta_M M^*$, which is large enough that the planner can close small stock gaps much faster than passive decay alone, as the impulse responses in Figures 1–3 illustrate.

**AI in production** The elasticity of AI capacity in the goods technology is set to $\zeta = 0.06$ as a transparent scenario capturing a modest but non-trivial contribution of $M_t$ to value added. Because macro evidence is nascent, we treat $\zeta$ as a disciplined scenario parameter rather than a tight estimate and report sensitivity in robustness checks.

**Digital services technology** We set $\phi = 0.7$ to capture aggregate congestion and bottlenecks (energy, cooling, bandwidth, engineering) that temper the micro-level scalability of inference. This curvature helps ensure a well-defined steady state under binding capacity and matters for the strength of the services dividend in the AI Euler equation, but does not affect the qualitative conclusions about relative adjustment speeds.

**Marginal resource cost of services** We calibrate a small positive $\chi$ consistent with recent evidence that the resource cost of LLM inference is extremely low. The AI Index

[10] Alphabet 2023 Form 10-K; Microsoft 2023 Form 10-K. Notably, Amazon's 2023 Form 10-K explicitly acknowledges that the "increased pace of technology development" in AI may require shorter useful lives for specialized hardware, validating this distinction.

(2025) documents a drop in GPT-3.5-level inference prices from about \$20 to \$0.07 per million tokens during Nov 2022–Oct 2024, and Google reports a median inference energy of roughly 0.24 Wh for a text prompt.[11] At retail electricity prices around \$0.10–0.15 per kWh, this corresponds to energy costs of the order of $10^{-5}$ dollars per prompt, highlighting just how small the physical resource cost is.

Because output in the model is measured in abstract goods units, there is no unique mapping from these dollar figures to $\chi$. We therefore set $\chi = 0.006$ as a conservative scenario value: this places the marginal resource cost of digital services well below unity in goods units, ensures that the steady state lies in the binding-capacity regime ($p_D^* > \chi$), and implies resource use for digital services that is easily reconciled with the inference-cost numbers above under any plausible normalisation of $Y^*$. The main results are robust to small perturbations around this value.

## 5.3 Consumption aggregator and digital-share target

We set the elasticity of substitution between tangible goods and digital services to $\sigma = 1.5$. This implies that the two streams are gross substitutes ($\sigma > 1$), consistent with estimates for related categories such as home-production substitution (often cited around 1.5–1.8).

We target a baseline digital expenditure share of $s_D = 0.10$. The BEA Digital Economy Satellite Account estimates that the digital economy accounted for about 10% of US GDP in 2022 (the latest benchmark year). In the model, $s_D$ is the expenditure share of digital services in the composite consumption aggregator rather than the value-added share of a "digital sector". We use the BEA figure as an order-of-magnitude target for this consumption share, recognising that the two concepts are not identical.

Given the exogenous parameters (including $\chi$), we jointly calibrate the CES weight $\theta$ and the service-efficiency parameter $\psi$ to match the target $s_D$ in the steady state. Because capacity binds, the shadow price $p_D^* = \mathrm{MRS}^*$ is endogenous. For given $(\theta, \psi)$ we solve for the steady state, obtain $p_D^*$, and then update $\theta$ so that the implied expenditure share satisfies

$$\frac{s_D}{1 - s_D} \;=\; \left(\frac{1-\theta}{\theta}\right)^{\sigma} \left(p_D^*\right)^{1-\sigma} \quad \Longrightarrow \quad \theta \;=\; \frac{1}{1 + \left[\left(s_D/(1-s_D)\right)\left(p_D^*\right)^{\sigma-1}\right]^{1/\sigma}}.$$

Simultaneously, we adjust $\psi$ so that $D^* = \psi\,(M^*)^{\phi}$ matches the level of digital expenditure

[11] Stanford HAI, AI Index 2025; Google Cloud (2025) on inference energy; Elsworth et al. (2025). See the Introduction for further discussion.

consistent with $s_D$ and the resource constraint. In practice this two-dimensional fixed-point problem converges quickly: starting from a reasonable guess for $(\theta, \psi)$, a few iterations are enough to hit the expenditure-share target to numerical tolerance. This procedure ensures that the calibration is fully consistent with the binding-capacity Euler condition and with the targeted digital share.

**External price checks (not targets)** Published API prices vary by model and date: low-cost tiers such as Gemini Flash are now only a small multiple of benchmark resource costs, while frontier multimodal models remain substantially more expensive per token. We do not attempt to match any specific price series; instead, we use these figures as external reasonableness checks that the chosen combination of $\chi$, $p_D^*$ and the digital share $s_D$ lies in a plausible range given observed API pricing and energy-use statistics.

**Table 1:** Baseline calibration (mid-2025)

| Parameter | Description | Value | Evidence / target |
|---|---|---|---|
| *A. Preferences and aggregation* | | | |
| $\beta_a$ | Discount factor (annual) | 0.99 | Targets $r \simeq 1\%$ real; consistent with policy-model long-run assessments. |
| $\sigma$ | Substitution elasticity ($C^{\mathrm{T}}$ vs. $D$) | 1.5 | Plausibly $> 1$; analogy to home-production substitution (≈1.5–1.8). |
| *B. Technology (tangible goods)* | | | |
| $\alpha$ | Tangible capital share | 0.30 | Standard growth calibration; PWT labour-share evidence. |
| $\delta_a$ | Tangible depreciation (annual) | 0.06 | Consistent with BEA fixed-asset implied geometric rates (5–7%). |
| $\zeta$ | AI elasticity in production | 0.06 | Scenario parameter. |
| *C. Technology (AI capacity and digital services)* | | | |
| $\delta_{M,a}$ | AI economic depreciation (annual) | 0.33 | Economic life ≈3 years; faster than 6-year accounting lives at hyperscalers. |
| $\phi$ | Congestion in $D$ production | 0.7 | Captures aggregate bottlenecks (energy, cooling, bandwidth, engineering). |
| $\chi$ | Marginal resource cost of $D$ | 0.006 | Scenario value; consistent with very low measured inference energy/cost. |
| *D. Targets and calibrated parameters* | | | |
| $s_D$ | Digital expenditure share (target) | 0.10 | BEA Digital Economy Satellite Account (≈10% of GDP, 2022). |
| $\theta$ | CES weight on $C^{\mathrm{T}}$ | Calibrated | Jointly with $\psi$ to hit $s_D$ target under binding capacity. |
| $\psi$ | Service efficiency | Calibrated | Jointly with $\theta$ to hit $s_D$ target under binding capacity. |

*Notes:* Quarterly parameters used in the dynamics and policy analysis are $\beta = \beta_a^{1/4} \approx 0.9975$, $\delta = 1 - (1 - \delta_a)^{1/4} \approx 0.0154$ and $\delta_M = 1 - (1 - \delta_{M,a})^{1/4} \approx 0.0953$. All results in Sections 6 and 7 are computed in the binding-capacity steady state associated with this parameter vector.

# 6 Adjustment dynamics

This section characterises the transition dynamics of the two-asset economy. The distinct depreciation rates of conventional capital and AI capacity imply widely separated adjustment speeds. We start from the linearised general equilibrium (GE) system derived in Section 3 and Appendix B, show how its eigenvalues relate to the underlying depreciation rates, and then compare the resulting impulse responses with simple partial equilibrium (PE) geometric benchmarks. Throughout, time is measured in quarters and all parameters are the quarterly values implied by the calibration in Section 5.

## 6.1 Linear general equilibrium dynamics

Let $k_t \equiv \ln K_t - \ln K^*$ and $m_t \equiv \ln M_t - \ln M^*$ denote log deviations of the two capital stocks around the steady state $(K^*, M^*)$ described in Section 4. Log-linearising the tangible-capital Euler equation (9), the AI-capacity Euler equation (10) and the accumulation equations (6)-(7) around $(K^*, M^*)$ yields the two-dimensional linear system

$$\begin{bmatrix} k_{t+1} \\ m_{t+1} \end{bmatrix} = J \begin{bmatrix} k_t \\ m_t \end{bmatrix}, \qquad J \equiv \begin{bmatrix} J_{KK} & J_{KM} \\ J_{MK} & J_{MM} \end{bmatrix}, \tag{18}$$

where $J$ is the Jacobian of the law of motion for $(K_t, M_t)$ evaluated at the steady state. Each entry of $J$ is a unit-free elasticity built from the primitive parameters $(\alpha, \zeta, \sigma, \theta, \delta, \delta_M, \phi, \psi, \chi)$ and steady-state ratios such as $Y^*/K^*$ and $Y^*/M^*$. Because investment is endogenous in the feasibility constraint, the Euler block is second order in $(K_t, M_t)$; the reduced two-state law of motion $x_{t+1} = Jx_t$ is obtained by forming a companion system and selecting the stable eigenspace (Appendix B). Here we focus on the economic content of the resulting GE dynamics and compare them with PE benchmarks that hold investment at replacement levels.

Intuitively, a positive deviation in $K_t$ lowers the marginal product of capital and hence the gross return (user cost) on tangible capital, which reduces the incentive to invest and induces a gradual return of $K_{t+1}$ towards $K^*$; this effect, together with depreciation, determines $J_{KK}$. The term $J_{MM}$ is governed by the high depreciation rate of AI capacity, the diminishing marginal product of $M$ and the curvature of the congestion term in the services dividend. The off-diagonal terms reflect general equilibrium linkages. Through the Cobb-Douglas technology, higher AI capacity raises the marginal product of tangible capital and vice versa, generating terms proportional to $\alpha\zeta$. When capacity binds, the services dividend contributes an additional cross-effect proportional to $(\text{MRS}^* - \chi)\phi$; when capacity is slack that component

is absent.

Let $\lambda_1$ and $\lambda_2$ denote the eigenvalues of $J$, ordered so that $0 < \lambda_1 \leq \lambda_2 < 1$. Because the model is saddle-path stable (Appendix B), both eigenvalues lie strictly inside the unit circle. The associated half-lives are

$$t_{1/2}(\lambda_i) = \frac{\ln 0.5}{\ln \lambda_i}, \qquad i = 1, 2.$$

In the baseline calibration of Section 5, the fast eigenvalue $\lambda_1$ is near zero ($\approx 7.7 \times 10^{-6}$), implying an extremely rapid adjustment mode through which the planner eliminates most stock deviations within one or two quarters by re-optimising gross investment. The slow eigenvalue $\lambda_2 \approx 0.980$ governs the gradual convergence of tangible capital. Appendix B reports the full numerical Jacobian. For comparison, the PE survival rates are $1-\delta_M \approx 0.905$ and $1-\delta \approx 0.985$; the GE eigenvalues diverge from these benchmarks because the corrected linearisation keeps investment endogenous in the feasibility constraint. Economically, the system decomposes into a fast mode through which the planner exploits the large replacement flows of short-lived AI hardware to close stock gaps quickly, and a slow mode associated with the sluggish adjustment of tangible capital.

Given an initial deviation $x_0 \equiv (k_0, m_0)'$, the GE solution is

$$x_t = J^t x_0 = \omega_1 \lambda_1^t v_1 + \omega_2 \lambda_2^t v_2, \tag{19}$$

where $v_i$ is the eigenvector associated with $\lambda_i$ and the weights $\omega_i$ are determined by the initial condition. The fast mode dominates the short-run response of AI capacity and the digital block; the slow mode governs the protracted adjustment of tangible capital and output.

## 6.2 Partial equilibrium geometric benchmarks

While (18) describes the exact linearised GE dynamics, it is useful to have a simpler benchmark that makes the role of depreciation rates completely transparent. Following the heuristic in the original version of the paper, consider a PE experiment in which investment in each asset is fixed at its steady-state replacement level. The capital-stock equations then become

$$K_{t+1} = (1-\delta)\, K_t + \delta K^*, \tag{1}$$

$$M_{t+1} = (1-\delta_M)\, M_t + \delta_M M^*. \tag{2}$$

Expressed in log deviations, a first-order approximation delivers

$$k_{t+1} \approx (1-\delta)\, k_t, \qquad m_{t+1} \approx (1-\delta_M)\, m_t. \tag{20}$$

The PE dynamics thus consist of two independent geometric decay processes, with roots equal to the survival rates of the respective capital stocks.

Under the baseline quarterly calibration $(\delta, \delta_M) = (0.0154, 0.0953)$, these roots are

$$\text{Fast root (AI capacity)}: \quad 1-\delta_M \approx 0.905, \qquad \text{Slow root (tangible capital)}: \quad 1-\delta \approx 0.985.$$

The implied half-lives are

$$t^M_{1/2} = \frac{\ln 0.5}{\ln(1-\delta_M)} \approx 6.9 \text{ quarters } (\approx 1.7 \text{ years}), \qquad t^K_{1/2} = \frac{\ln 0.5}{\ln(1-\delta)} \approx 44.8 \text{ quarters } (\approx 11 \text{ years}) \tag{21}$$

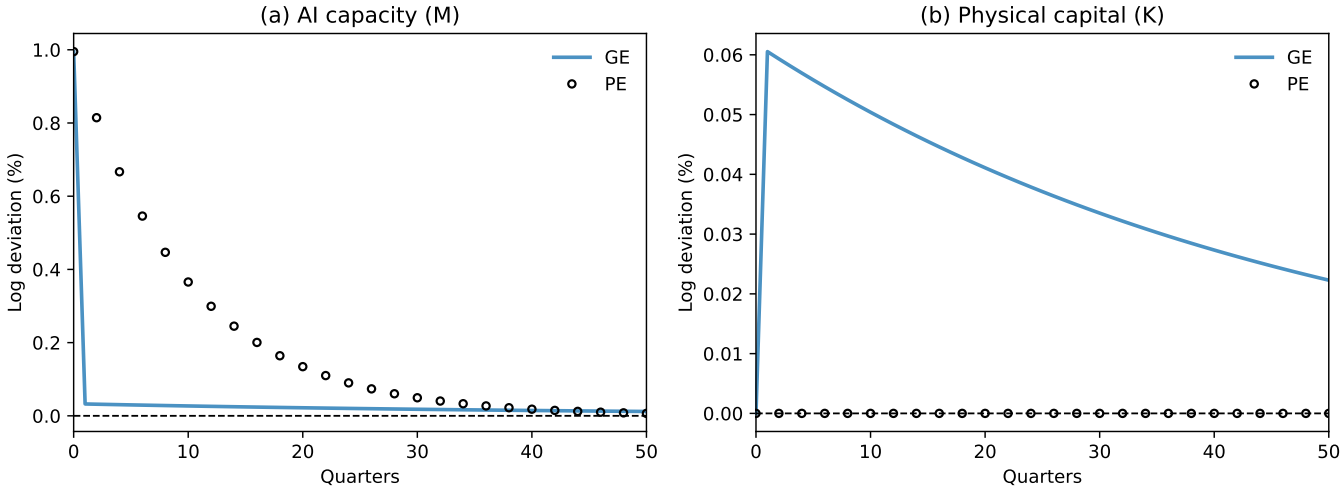
These PE half-lives provide a transparent yardstick for passive depreciation-driven decay when investment is held at its steady-state replacement level. In the corrected GE linearisation, investment re-optimisation introduces a much faster adjustment mode for AI capacity, so equilibrium convergence can be substantially quicker than passive decay, especially for high-depreciation hardware. Appendix B reports the Jacobian and eigenvalues.

### 6.3 Impulse responses: GE versus PE

To illustrate the link between the GE solution (18)–(19) and the PE benchmark (20), Figure 1 plots the response of AI capacity and tangible capital to a one-off 1% increase in $M_t$ at date 0. The PE benchmark holds investment fixed at steady-state replacement levels and therefore abstracts from endogenous movements in consumption and marginal utility; it captures only passive depreciation-driven decay. The comparison therefore measures how large the GE investment-adjustment feedbacks are in the baseline calibration. The solid lines show the GE impulse responses from the stable-manifold solution (Appendix B); the circle markers show the PE paths $m_t = (1-\delta_M)^t m_0$ and $k_t = 0$ (since only $M_0$ is shocked and PE investment in $K$ remains at replacement).

For AI capacity, the GE path adjusts markedly faster than the PE benchmark. The mechanism is straightforward: with $\delta_M \approx 0.095$ per quarter, steady-state replacement investment is about 9.5% of the stock. A 1% excess stock is therefore small relative to replacement flows, so the planner can eliminate most of the deviation within a few quarters simply by cutting gross investment modestly below replacement while keeping $X_t \geq 0$. The PE benchmark,

**Figure 1:** Adjustment after a 1% AI-capacity shock: general equilibrium vs. PE benchmark




Notes: The figure plots log deviations of $M_t$ (left panel) and $K_t$ (right panel) from the steady state following a 1% increase in $M_0$. Solid lines: GE dynamics from the stable-manifold solution (Appendix B). Circle markers: PE benchmark, which holds investment at steady-state replacement levels so that deviations decay passively at the survival rate. The PE benchmark isolates depreciation-driven decay; the GE response allows investment to adjust, which closes stock gaps much faster for short-lived AI hardware because replacement flows are large relative to small deviations. For $K$, the PE benchmark is zero by construction (only $M_0$ is shocked), whereas in GE the cross-effect through Cobb–Douglas production raises $K$ temporarily.

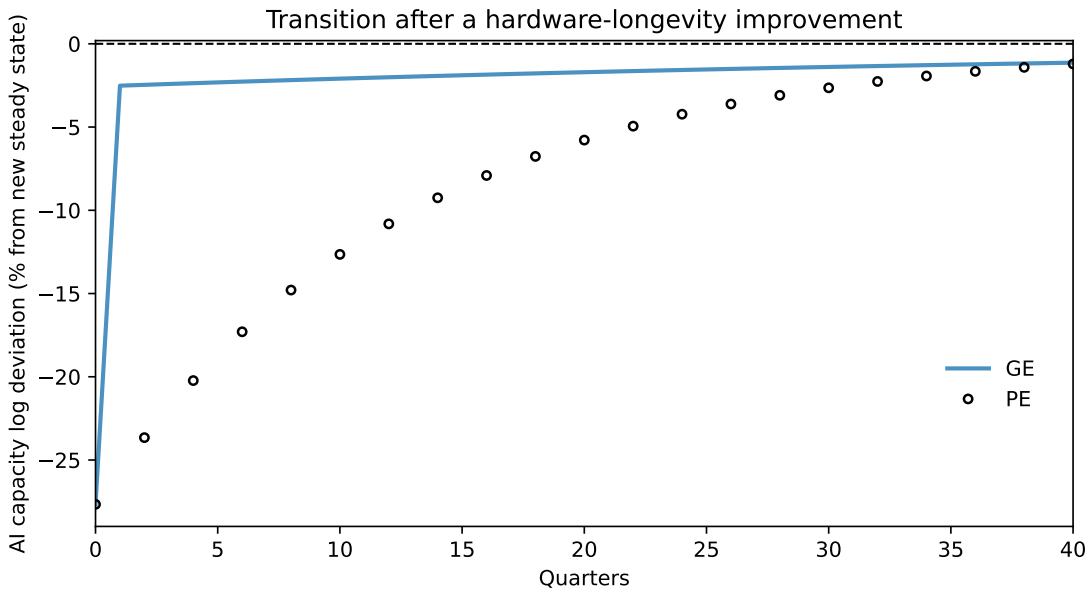
which holds $X_t = \delta_M M^*$ by construction, captures only the passive component of this unwind and therefore decays more slowly, with a half-life of $t_{1/2}^M \approx 6.9$ quarters. For tangible capital, the PE benchmark is zero because only $M_0$ is shocked, but the GE path shows a temporary increase in $K$ driven by the cross-effect of higher AI capacity on the marginal product of tangible capital, followed by a slow return governed by the second eigenvalue.

## 6.4 Hardware-longevity shocks

Consider next a permanent reduction in the quarterly depreciation rate of AI capacity, $\delta_M \mapsto \delta_M - \Delta$ with $\Delta > 0$, implemented at date 0. This shock shifts the steady state from $(K^*, M^*)$ to a new pair $(K^{**}, M^{**})$ with higher AI capacity and higher tangible capital. In the GE system, the transition dynamics for the log deviations from the new steady state, $\tilde{k}_t \equiv \ln K_t - \ln K^{**}$ and $\tilde{m}_t \equiv \ln M_t - \ln M^{**}$, again follow (18) with the Jacobian re-evaluated at $(K^{**}, M^{**})$. The GE dynamics at the new steady state reflect the changed depreciation rate, with the Jacobian re-evaluated at $(K^{**}, M^{**})$.

As a transparent comparison, the PE benchmark assumes that the gap to the new AI steady

**Figure 2:** Transition after a hardware-longevity improvement: GE vs. PE benchmark




Notes: The figure plots log deviations of AI capacity from the new steady state after a permanent two-percentage-point fall in $\delta_M$. Solid line: GE dynamics re-evaluated at $(K^{**}, M^{**})$. Circle markers: PE benchmark (22), which holds investment at replacement levels so that the gap closes passively at the new survival rate $1 - \delta_M + \Delta$. The GE path adjusts faster because the planner front-loads AI investment in response to the permanent longevity improvement, exploiting the large replacement flows to close the gap to the new steady state more quickly.

state closes at the new survival rate:

$$M_t - M^{**} \approx (1 - \delta_M + \Delta)^t \left(M_0 - M^{**}\right). \tag{22}$$

Figure 2 illustrates this transition for a two-percentage-point fall in the quarterly depreciation rate (from $\delta_M \approx 0.0953$ to $\delta_M - \Delta \approx 0.0753$). As in Figure 1, the GE path adjusts faster than the PE benchmark because the planner re-optimises investment rather than holding it at replacement. The PE half-life of the gap to the new steady state is approximately

$$t_{1/2}^{M,\text{new}} = \frac{\ln 0.5}{\ln(1 - \delta_M + \Delta)} \approx 8.9 \text{ quarters},$$

longer than in the baseline because the improvement in longevity slows adjustment.

## 6.5 Diffusion from a negligible initial stock

Finally, consider the adoption path of AI capacity when the economy starts from a negligible stock $M_0 \ll M^*$. In the PE benchmark, applying the recursion with $M_0 \approx 0$ and constant

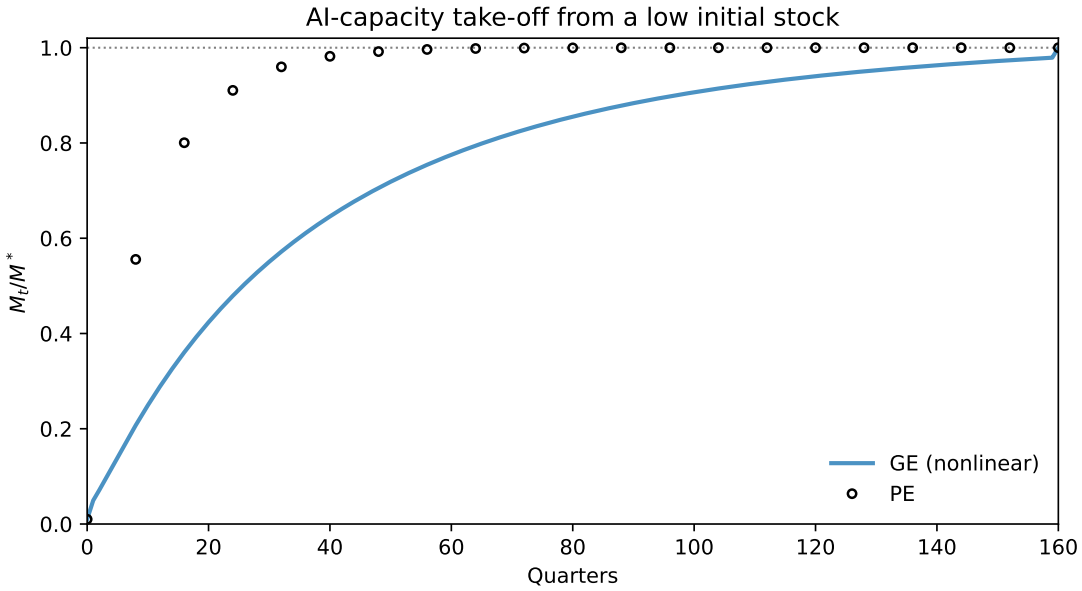
investment at its steady-state value implies

$$M_t \approx M^* \Big( 1 - (1 - \delta_M)^t \Big), \tag{23}$$

so that the gap to the steady state closes at the survival rate $1 - \delta_M$. With the baseline calibration, half of $M^*$ is reached after about $t^M_{1/2} \approx 6.9$ quarters and around 90% after roughly 23 quarters (just under six years).

In the GE system, the qualitative pattern is similar but the speed of diffusion differs. Starting from a low $M_0$ and a correspondingly low $K_0$, the planner must build up both capital stocks simultaneously while keeping consumption non-negative, which requires smoothing investment over a longer horizon than the PE benchmark assumes. As a result, the nonlinear perfect-foresight GE transition converges more slowly than (23): the PE benchmark implicitly front-loads investment at the steady-state replacement rate from date zero, whereas the GE planner optimally spreads the build-up across many periods. Figure 3 compares the two paths.

**Figure 3:** AI-capacity take-off from a low initial stock: GE vs. PE benchmark




Notes: The figure plots $M_t/M^*$ under a nonlinear perfect-foresight GE transition (solid line) and the PE diffusion benchmark (23) (circle markers) for an economy starting from $M_0 \approx 0$. The PE benchmark holds investment fixed at steady-state replacement levels, implicitly front-loading capacity accumulation; the GE planner optimally smooths the build-up of both capital stocks, resulting in a more gradual diffusion path.

### 6.6 Interpretation

Across these experiments, the comparison between GE and PE paths clarifies the distinct roles of passive depreciation and active investment adjustment. The PE benchmark holds investment at steady-state replacement levels and therefore captures only depreciation-driven decay; it provides a deliberately mechanical "no behavioural response" yardstick whose half-life is tied directly to the survival rate of each asset. The GE response allows investment to adjust endogenously, which can close stock gaps much faster for short-lived AI hardware (because replacement flows are large relative to small deviations) and can smooth accumulation over a longer horizon when starting far from steady state (because the planner internalises the resource cost of rapid build-up). The PE benchmarks remain useful precisely because of this contrast: they isolate the hardware-durability channel and provide an intuitive lower bound on the speed of passive unwind, against which the additional bite of GE investment re-optimisation can be measured.

## 7 Policy experiments and measurement issues

The tractability of the framework means that many policy levers map into a small set of technology parameters, chiefly the effective depreciation rate of AI hardware and the marginal resource cost of inference, $\chi$. Because the steady state is analytically characterised by two user–cost conditions (Section 4) and the linearised dynamics are governed by just two eigenvalues of the Jacobian $J$ (Section 6), a wide class of interventions can be evaluated with back-of-the-envelope expressions. For small shocks, the closed-form steady-state comparative statics in Appendix A and the geometric benchmarks from Section 6 provide very accurate approximations to the full general-equilibrium (GE) adjustment paths. Throughout this section we work with quarterly parameters. In particular, $\delta_M$ denotes the quarterly depreciation rate, and tax or longevity shocks are expressed in percentage points on this quarterly rate.

### Compute taxation and accelerated depreciation

A tax on specialised hardware raises the user cost of AI capacity in exactly the same way as a higher depreciation rate. Let $\tau$ denote an ad-valorem tax on the purchase price of GPUs (or an equivalent surcharge on investment in $M_t$), measured in percentage points on the quarterly rate. If firms write hardware off at the baseline $\delta_M$, the effective depreciation rate becomes

$$\tilde{\delta}_M = \delta_M + \tau. \tag{24}$$

Policies that shorten the accounting life of hardware, add regulatory charges to data-centre assets, or impose a levy on compute all operate through this effective rate.

Let $M^*(\delta_M)$ denote the steady-state AI-capacity stock in the binding-capacity regime. Appendix A shows that $M^*$ is uniquely defined and strictly decreasing in $\delta_M$. The semi-elasticity of steady-state capacity with respect to the hardware tax,

$$\varepsilon_M^{\tau} \equiv \frac{\partial \ln M^*}{\partial \tau} = \frac{\partial \ln M^*}{\partial \delta_M},$$

is therefore negative and can be computed directly from the implicit-function expression for $M^*$ in Appendix A. In the baseline quarterly calibration (Table 1), this semi-elasticity is sizeable: a one-percentage-point surtax on hardware ($\tau = 0.01$) lowers the optimal AI-capacity stock by just over ten per cent, and a two-percentage-point surtax cuts it by about twenty per cent (see panel (a) of Figure 4). Because the Cobb–Douglas exponents are fixed, the aggregate labour share $s_L^* = 1 - \alpha - \zeta$ remains unchanged; the tax acts purely on the scale of AI capital and on the distribution of capital income.

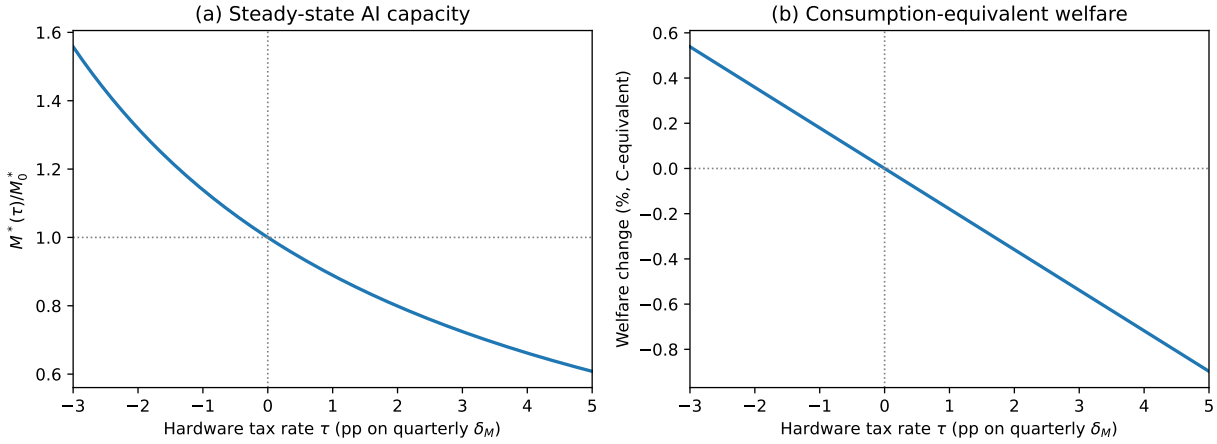
The same logic applies in reverse to accelerated-depreciation allowances. Expensing an additional $\Delta$ of hardware, or introducing a targeted investment credit for AI-related equipment, is equivalent (in present-value terms) to $\tau = -\Delta$ in (24) and therefore raises $M^*$ with the same semi-elasticity. In this sense, standard hardware taxes and accelerated-depreciation policies can be viewed as symmetric instruments that move the effective depreciation rate in opposite directions. The transition towards the new steady state is dominated by the fast GE eigenvalue $\lambda_1$ (Section 6), so the resulting adjustment in $M_t$ unfolds over a horizon of only a few years.

## Permanent hardware-longevity shock

A once-and-for-all reduction in the economic depreciation rate of AI hardware, $\delta_M \mapsto \delta_M - \Delta$ with $\Delta > 0$, captures improvements in hardware design, repairability, or cooling that extend the useful life of GPUs and related infrastructure. Linearising the steady-state condition for $M^*$ with respect to $\delta_M$ and aggregating the consumption stream along the transition path yields a closed-form welfare gain, expressed in consumption-equivalent units:

$$\Omega(\Delta) = \frac{\beta(1-\beta)\,\Delta}{\big(1 - \beta + \beta\,\delta_M\big)(1-\phi)}\,\frac{M^*}{C^*}, \tag{25}$$

**Figure 4: Steady-state effects of a hardware tax (or subsidy)**




*Notes:* Panel (a) plots the ratio $M^*(\tau)/M_0^*$ as the effective hardware tax $\tau$ (in percentage points on the quarterly $\delta_M$) varies. A two-percentage-point tax ($\tau = 0.02$) reduces the AI-capacity stock by approximately twenty per cent. Panel (b) converts the utility change into consumption-equivalent units using (25). For moderate tax rates, welfare responds approximately linearly, with losses of around 0.18 percentage points per percentage-point of tax.

where $C^*$ is steady-state consumption and $(\beta, \delta_M, \phi)$ are the quarterly parameters (with $\beta = \beta_a^{1/4}$).[12]

For the baseline calibration (implying quarterly $\beta \approx 0.9975$ and $\delta_M \approx 0.0953$, with $\phi = 0.7$), the coefficient in front of $M^*/C^*$ in (25) is approximately 0.086. Given the baseline steady-state ratio $M^*/C^* \approx 2.1$, this implies

$$\Omega(\Delta) \;\approx\; 0.18\,\Delta.$$

A two-percentage-point improvement in hardware longevity on the quarterly rate ($\Delta = 0.02$) therefore raises lifetime utility by roughly 0.36 per cent in consumption-equivalent terms (see panel (b) of Figure 4). Because the transition in $M_t$ is governed by the fast adjustment mode, a substantial fraction of this gain accrues in the first few years: the bulk of the adjustment in AI capacity is complete within roughly two half-lives (about three to four years), after which the economy is close to its new steady state.

[12]The formula uses log utility and the binding-capacity regime. The derivation approximates the transition of $M_t$ by the PE passive-decay benchmark with survival rate $1 - \delta_M \approx 0.905$, which provides a conservative (slower-than-GE) approximation to the transition speed. The numerator captures the discounted effect of a permanent change in $\delta_M$ on the user cost of AI capacity; the denominator $(1 - \phi)$ reflects the curvature of the capacity constraint $D_t = \psi M_t^{\phi}$.

## Competition policy

Dominant cloud providers often enjoy contractual and scale advantages that lower their private effective depreciation rate: bespoke maintenance contracts, favourable power-purchase agreements, or proprietary cooling solutions all allow hardware to remain in service longer than for a marginal entrant. Reversing part of that asymmetry, through data-portability mandates, interoperability standards, or limits on exclusive service contracts, can be viewed as a policy that raises the incumbent's effective $\delta_M$ towards the economy-wide baseline.

Within the present framework, any such intervention operates through the same $M^*(\delta_M)$ schedule as a hardware tax. Using the semi-elasticity discussed above, a five-percentage-point increase in an incumbent's effective quarterly depreciation rate would, in the baseline calibration, reduce its optimal capacity stock by on the order of 40 per cent. Because digital services are produced at very low marginal resource cost once capacity is installed, the welfare cost of a moderate reduction in scale is second order, whereas the potential pro-competitive gains lie largely outside the representative-agent model. The main message here is that the scale and sign of the mechanical capacity effect can be quantified in closed form once $\delta_M$ is pinned down.

## Macroprudential timing

The high depreciation rate of AI hardware also has implications for financial-stability tools. Suppose hardware investment is partly debt-financed and that policymakers wish to lean against AI-driven credit booms by introducing a counter-cyclical capital buffer. The PE passive-decay benchmark provides a conservative guide to the speed at which hardware-driven booms unwind once investment returns to replacement:

$$t_{1/2}^{\mathrm{PE}} = \frac{\ln 0.5}{\ln(1-\delta_M)} \approx 6.9 \text{ quarters},$$

or roughly one and three-quarter years in the baseline calibration. As Figures 1–2 illustrate, the GE adjustment can be substantially faster because the planner (or market participants) can cut gross investment below replacement, exploiting the large steady-state replacement flows that characterise short-lived assets. A simple prudential rule that accumulates capital buffers during hardware-investment booms and releases them on a schedule tied to the PE half-life would therefore provide a transparent and conservative timing benchmark. Unlike traditional Basel III buffers, which often release with multi-year lags, such a rule can be calibrated ex ante from observable hardware depreciation rates and replacement cycles,

making it both transparent and time-consistent.

### Summary

Across taxation, longevity shocks and prudential tools, two primitives dominate outcomes: the effective depreciation rate of AI hardware and the elasticity with which digital and tangible uses substitute in consumption. Policy measures that shift $\delta_M$ translate in a mechanically quantifiable way into changes in AI capacity, prices and lifetime welfare via the steady-state elasticities in Appendix A and the fast GE adjustment mode in Section 6. Measures that alter the substitution elasticity $\sigma$ primarily affect how much of the resulting surplus is recorded as digital rather than tangible expenditure. Additional layers, such as skill heterogeneity or nominal rigidities, can enrich distributional or Phillips-curve analysis without disturbing the core elasticities already derived in the two-asset block. In all cases, the key measurement tasks for policy are to pin down effective GPU depreciation, the marginal resource cost of inference, and the relevant substitution elasticities.

## 8 Extensions and robustness

The two-asset core explains how hardware longevity, inference cost and substitution elasticities shape growth and factor incomes once AI capacity is part of the production process. This section sketches three modifications that preserve the algebraic backbone of the model while opening avenues for distributional and monetary analysis. In each case the $\{K_t, M_t\}$ block, its steady-state characterisation and its linear general-equilibrium (GE) dynamics remain exactly as in Sections 4–6; the extensions either add margins that are analytically separable or enter through static wedges in wage-setting or price-setting conditions. The eigenvalues of the core two-asset system, and the associated GE and PE dynamics, are therefore unchanged.

### 8.1 Heterogeneous labour and wage polarisation

Let labour consist of routine and non-routine types, $L_t^{\mathrm{R}}$ and $L_t^{\mathrm{N}}$, aggregated with elasticity $\nu > 1$:

$$Y_t = A K_t^{\alpha} M_t^{\zeta} \left[ (1-\omega) L_t^{\mathrm{N}} + \omega \Big( \xi M_t L_t^{\mathrm{R}} \Big)^{(\nu-1)/\nu} \right]^{\frac{\nu}{\nu-1}(1-\alpha-\zeta)}.$$

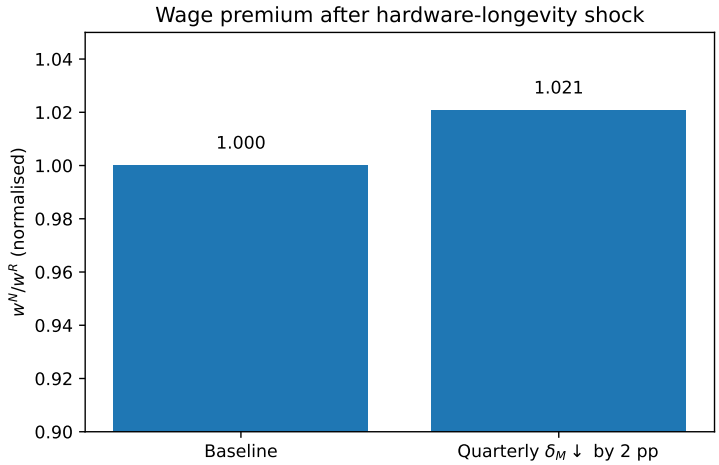
AI capacity is assumed to substitute more strongly for routine tasks; the parameter $\xi > 0$ scales that substitutability. Because each labour type supplies perfect substitutes within its

own class, the wage ratio is pinned down in closed form:

$$\frac{w_t^{\mathrm{N}}}{w_t^{\mathrm{R}}} = (\xi M_t)^{\frac{\nu-1}{\nu}}.$$

Choosing $(\nu, \xi) = (1.08, 1)$ implies that a doubling of $M_t$ widens the non-routine wage premium by $2^{(\nu-1)/\nu} \approx 1.053$, i.e. just over 5%. This magnitude is consistent with the rise in the EU Structure of Earnings Survey between 2018 and 2022 for tasks classified as "information-processing" versus "manual".[13]

**Figure 5: Wage premium after hardware-longevity shock**




*Notes:* Bars display the wage ratio $w^{\mathrm{N}}/w^{\mathrm{R}}$ in the baseline steady state (left, blue) and after a permanent two-percentage-point fall in the quarterly depreciation rate $\delta_M$ (right, amber). For the parameter choice $(\nu, \xi) = (1.08, 1)$, the new steady state exhibits a wage premium of just over 2%, consistent with the log-elasticity $\partial \ln(w^{\mathrm{N}}/w^{\mathrm{R}})/\partial \ln M = (\nu - 1)/\nu \approx 0.074$. Axes are truncated at 0.90–1.05 for clarity.

This extension preserves analytical tractability. The capital block remains two-dimensional, and the wage ratio follows directly from $M_t$ without introducing additional state variables. Shocks that raise steady-state AI capacity, such as the hardware-longevity improvements in Section 7, therefore translate one-for-one into predictable changes in skill premia, with the speed of adjustment governed by the fast eigenvalue of the $\{K_t, M_t\}$ system.

[13] Eurostat SES table EARN_SES_LC, extraction April 2025.

## 8.2 Nominal rigidities and inflation dynamics

Embedding the consumption-side CES aggregator in a Dixit–Stiglitz price-setting framework preserves the log-linearity of the Euler block. Suppose a continuum of differentiated firms sets prices à la Calvo, with probability $1-\rho$ of resetting in any given period. Linearising the New Keynesian Phillips curve around the AI-augmented steady state yields

$$\pi_t = \beta\,\mathbb{E}_t\pi_{t+1} + \kappa\,\hat{u}_t, \qquad \kappa = \frac{(1-\rho)(1-\beta\rho)}{\rho}\,\frac{1-\theta}{\theta\sigma}, \tag{26}$$

where $\hat{u}_t$ is the log deviation of real marginal cost from its flexible-price level, and $\beta$ is the quarterly discount factor. In the present environment, real marginal cost is tied to the tangible-goods price and the CES weight on digital services, so $\hat{u}_t$ becomes a linear function of hardware costs and of the relative price of digital consumption.

A surprise improvement in hardware longevity, modelled as a fall in $\delta_M$, therefore acts as a sequence of negative cost-push shocks. In the binding-capacity regime, the impact on $\hat{u}_t$ is approximately proportional to the change in the effective user cost of AI capacity. With $\rho = 0.75$ (implying an average price duration of one year) and the baseline calibration in Section 5, a four-percentage-point reduction in quarterly $\delta_M$ generates an initial disinflation of roughly 0.3 percentage points on a year-on-year basis, with the effect decaying over about three quarters. The exact paths follow directly from (26) once one specifies a monetary-policy rule and feeds in the $\{K_t, M_t\}$ path from the core GE system.

## 8.3 Endogenous model size

Thus far the stock of AI capacity $M_t$ captures only hardware and associated infrastructure. To allow for endogenous model size, let $S_t$ denote the number of parameters (or an effective complexity index) of the frontier model. The capacity constraint for digital services becomes

$$D_t \;\le\; \psi\, M_t^{\phi} S_t^{\vartheta}, \qquad 0 < \vartheta < 1,$$

so that, under binding capacity, $D_t = \psi\, M_t^{\phi} S_t^{\vartheta}$. The new state variable $S_t$ evolves according to

$$S_{t+1} = (1-\delta_S)S_t + I_t^S,$$

with $\delta_S \in (0,1)$ capturing obsolescence from data staleness and architectural advances, and $I_t^S$ denoting "training investment" (engineering and compute devoted to growing the model).

Because $S_t$ does not enter tangible-goods production $Y_t$, the $\{K_t, M_t\}$ block is determined independently of $S_t$. The system is recursive: once $(K^*, M^*)$ are pinned down, the Euler equation for $S_t$ determines $S^*$ conditional on that block. In the binding-capacity regime the steady-state Euler condition for $S^*$ takes the form

$$1 = \beta\Big[(1-\delta_S) + \Big(MRS^* - \chi\Big)\,\vartheta\,\psi\,(M^*)^{\varphi}(S^*)^{\vartheta-1}\Big], \tag{27}$$

with $MRS^*$ defined as in Section 4. The right-hand side of (27) is strictly decreasing in $S^*$, guaranteeing a unique interior solution for the training stock. A permanent fall in $\delta_S$ raises $S^*$ while leaving $M^*$ and the Jacobian of the $\{K_t, M_t\}$ subsystem essentially unchanged: training and inference respond on separate margins, and the two GE eigenvalues for the hardware block are preserved.

### 8.4 Summary

All three extensions leave the two-asset backbone intact. Skill heterogeneity turns capacity growth into wage dispersion; nominal rigidities turn hardware shocks into measurable inflation episodes; and endogenous model size separates the training and inference margins. Each module can be switched on without re-deriving the core steady-state conditions or the $\{K_t, M_t\}$ dynamics, keeping the framework modular while retaining its analytical tractability. The key objects that continue to govern both the baseline and the extended models are the user-cost system for $(K^*, M^*)$ and the fast–slow pair of eigenvalues that determine how quickly AI booms and conventional-capital adjustments unfold over time.

## 9 Conclusion

This paper has developed a simple two-asset growth model in which AI capacity joins conventional capital as a separate reproducible asset. AI capacity is interpreted as installed, rival hardware and infrastructure that embody non-rival models, deliver low-marginal-cost digital services and raise productivity in the tangible-goods sector. The steady state of this economy is analytically characterised by two user-cost conditions: one pins down a power-law relationship between tangible capital and AI capacity, the other determines the unique interior AI stock as the root of a monotone marginal-benefit map. Linearising around this steady state yields a two-dimensional general-equilibrium (GE) system with two eigenvalues whose stable manifold defines a two-dimensional law of motion for the capital stocks. Simple geometric laws of motion based on the survival rates provide transparent PE benchmarks for policy experiments.

Three quantitative lessons emerge. First, hardware durability is the dominant state variable. It governs the PE passive-decay half-life of AI-investment booms (roughly seven quarters in the baseline calibration), dictates how much capacity firms wish to hold, and controls the size of the digital block. Because replacement flows are large for short-lived assets, the GE planner can close small stock gaps substantially faster than passive decay; conversely, conventional capital adjusts over more than a decade. Second, the elasticity with which digital services substitute for tangible consumption, together with the marginal resource cost of inference, acts as the key wedge between growth and measured inflation: when the marginal cost of inference is very low, modest changes in these parameters can generate large shifts in expenditure shares with only small movements in aggregate output. Third, because the aggregate labour share is pinned down by primitive Cobb–Douglas exponents, most redistribution in the baseline model runs through capital ownership rather than through wages, unless AI is allowed to substitute directly for particular labour types in production.

The analytical structure pays off directly in policy space. In the baseline calibration, a one-percentage-point hardware tax on the quarterly depreciation rate reduces the optimal AI capacity stock by just over ten per cent, while a two-percentage-point tax cuts it by about twenty per cent. Conversely, an equally sized extension of GPU service lives, interpreted as a two-percentage-point fall in the quarterly economic depreciation rate, raises lifetime welfare by around 0.36 per cent in consumption-equivalent terms. The same primitives govern prudential timing: the PE passive-decay benchmark implies that AI-investment booms unwind with a half-life of about seven quarters once investment returns to replacement, and the GE adjustment can be faster still, so a counter-cyclical capital buffer aimed at AI lending can be released within roughly two years of an investment peak without elaborate forecasting. Overall, the framework points to hardware-centric instruments (accelerated depreciation, compute levies, repair mandates) as levers with unusually high marginal bite, and quantifies their effects using transparent steady-state elasticities and simple dynamic adjustments.

Several avenues remain open. Endogenising the training stock allows model size to interact with hardware cycles; layering on heterogeneous labour maps capacity growth into earnings dispersion; and introducing nominal rigidities turns hardware shocks into measurable disinflation episodes. None of these extensions threatens tractability: they either add separable state variables or introduce static wedges, leaving the two-asset core, its user-cost characterisation and its GE eigenvalues intact. What they do require is better empirical measurement of three quantities: effective GPU depreciation, the true marginal resource cost of inference (including energy and cooling), and the substitution elasticities between digital and tangible

uses (in both consumption and production). Improving those measurements is likely to deliver a higher marginal return than rerunning ever more elaborate simulations on poorly pinned-down parameters.

Artificial intelligence therefore confronts macroeconomists with a familiar but neglected fact: assets that depreciate quickly behave very differently from assets that last. When AI hardware depreciates at more than thirty per cent per year, the horizon for policy intervention shrinks from decades to years, and the fiscal cost of poorly timed incentives compounds just as fast. A framework that keeps that arithmetic in the foreground, while remaining simple enough to fit on the back of an envelope, offers a useful starting point for thinking about the macroeconomics of the next wave of capital deepening.

# References

# A Existence and uniqueness of the steady state

This appendix provides a rigorous proof of existence, uniqueness, and interiority of the steady state. The analysis is conducted for a generalised Euler-return map that nests both the slack- and binding-capacity regimes implied by the Kuhn–Tucker conditions and underpins the steady-state characterisation in Section 4. All arguments are purely static and do not depend on the linearised dynamics in Section 6.

## A.1 Feasible domain and primitives

Throughout the appendix we impose the baseline assumptions from Section 3, and we make explicit the parameter restrictions used below:

$$\alpha \in (0,1), \quad \beta \in (0,1), \quad 0 < \zeta < 1-\alpha, \quad \phi \in (0,1],\ \psi > 0, \quad \delta_M \geq 0, \quad \delta \in [0,1), \quad A > 0, \quad \chi \geq 0,$$

and (for the CES subutility) $\theta \in (0,1)$ and $\sigma > 0$. To match the baseline specification in Sections 3–5, we adopt the CES aggregator with the weight on tangible consumption, where $C_T$ corresponds to $C^{\mathrm{T}}$ in the main text:

$$U(C_T, D) = \left[\theta\, C_T^{\rho} + (1-\theta)\, D^{\rho}\right]^{1/\rho}, \qquad \rho \equiv \frac{\sigma - 1}{\sigma},$$

with the case $\sigma = 1$ understood in the CES limit $\rho \to 0$; the expressions below remain valid in that limit (all limits are well-behaved as $\sigma \to 1$). Hence the marginal rate of substitution of $D$ for $C_T$, which is also the shadow price of digital services in units of tangible consumption, is

$$p_D \;\equiv\; \mathrm{MRS}(C_T, D) \;=\; \frac{U_D}{U_{C_T}} \;=\; \frac{1-\theta}{\theta}\left(\frac{C_T}{D}\right)^{1/\sigma}.$$

The CES subutility is strictly increasing and concave for all $\sigma > 0$ and $\theta \in (0,1)$.

**Feasibility.** Feasible allocations require $C_T > 0$ and $D \geq 0$.

**Steady-state ratios and power-law relationships.** The tangible-capital Euler condition implies a steady-state output–capital ratio:

$$\Lambda_K \;\equiv\; \frac{\beta^{-1} - (1-\delta)}{\alpha} \;>\; 0, \qquad \Rightarrow \qquad \frac{Y^*}{K^*} = \Lambda_K.$$

Combining $Y^* = \Lambda_K K^*$ with $Y^* = A(K^*)^{\alpha} M^{\zeta}$ yields power-law relationships in $M$:

$$K^*(M) = \kappa_K\, M^\gamma, \qquad Y^*(M) = \kappa_Y\, M^\gamma, \qquad \gamma \;\equiv\; \frac{\zeta}{1-\alpha} \in (0,1),$$

where $\kappa_K \equiv (A/\Lambda_K)^{1/(1-\alpha)}$ and $\kappa_Y \equiv \Lambda_K\, \kappa_K$. Note that

$$\kappa_Y - \delta\kappa_K = \kappa_K(\Lambda_K - \delta), \qquad \Lambda_K - \delta = \frac{\beta^{-1} - 1 + \delta(1-\alpha)}{\alpha} > 0,$$

so $\kappa_Y - \delta\kappa_K > 0$.

**Capacity and the at-capacity MRS.** Capacity for digital services is

$$D^{\text{cap}}(M) = \psi\, M^\phi.$$

The (counterfactual) MRS of $D$ for $C_T$ evaluated at capacity is

$$p_D^{\text{cap}}(M) = \frac{1-\theta}{\theta}\left(\frac{C_T^{\text{cap}}(M)}{D^{\text{cap}}(M)}\right)^{1/\sigma},$$

where the tangible-goods resource constraint at capacity is

$$C_T^{\text{cap}}(M) \;=\; Y^*(M) - \delta K^*(M) - \delta_M M - \chi D^{\text{cap}}(M) \;=\; \Big(\kappa_Y - \delta\kappa_K\Big) M^\gamma - \delta_M M - \chi\psi\, M^\phi. \quad (3)$$

In the binding regime the optimum saturates capacity, so $D^* = D^{\text{cap}}$ and $p_D^* \geq \chi$ (with strict inequality if and only if the capacity multiplier is strictly positive, $\eta > 0$); in the slack regime the planner chooses $D^* < D^{\text{cap}}$ so that $p_D^* = \chi$.

**Assumption A.1 (Feasibility near the origin).**

$$\phi \;\geq\; \gamma \quad \text{and, if } \phi = \gamma, \text{ then } \;\kappa_Y - \delta\kappa_K \;>\; \chi\psi.$$

Assumption A.1 ensures $C_T^{\text{cap}}(M) > 0$ for sufficiently small $M > 0$: near $M = 0$, $M^\gamma$ dominates $M$ because $M^\gamma/M = M^{\gamma-1} \to \infty$ since $\gamma \in (0,1)$. Requiring $\phi \geq \gamma$ ensures that the resources required for capacity, which scale with $M^\phi$, vanish weakly faster than output near the origin (i.e. $M^\phi/M^\gamma \to 0$ when $\phi > \gamma$); if $\phi = \gamma$ we impose a strictly positive coefficient.

Economic intuition. Assumption A.1 guarantees that, when the stock $M$ is very small, the resource needs for capacity ($\chi\psi M^\phi$) and for maintenance ($\delta_M M$) do not exceed the net output available to tangible consumption $(\kappa_Y - \delta\kappa_K)M^\gamma$. This ensures the economy is viable

near the origin.

**Boundary of the feasible set.** We use the at-capacity resource constraint to define a conservative feasibility domain on which $p_D^{\text{cap}}(M)$ and the generalised map are well-defined on $\mathcal{D}$:

$$M_{\max} \;\equiv\; \sup\{M > 0 :\; C_T^{\text{cap}}(M) > 0\} \;\in\; (0, \infty], \qquad \mathcal{D} \;\equiv\; (0, M_{\max}).$$

On $\mathcal{D}$, $C_T^{\text{cap}}(M) > 0$ and $D^{\text{cap}}(M) > 0$, so $C_T^{\text{cap}}(M)/D^{\text{cap}}(M) > 0$ and hence $p_D^{\text{cap}}(M)$ is well-defined. Since $C_T^{\text{cap}}(\cdot)$ is continuous in $M$, the boundary property below follows. By Assumption A.1 the set $\{M > 0 :\; C_T^{\text{cap}}(M) > 0\}$ is nonempty and open near $M = 0$, so $\mathcal{D} = (0, M_{\max}) \neq \varnothing$. This domain is sufficient for feasibility under binding capacity; if capacity is slack, feasibility may also obtain for larger $M$ with $D^* < D^{\text{cap}}$, but the generalised map below is constructed on $\mathcal{D}$ (and Proposition A.3 shows that no steady state exists with $M \geq M_{\max}$ anyway).

[Characterisation of $M_{\max}$] Under Assumption A.1 there exists a unique $M_{\max} \in (0, \infty]$ such that $C_T^{\text{cap}}(M) > 0$ for all $M \in (0, M_{\max})$ and, if $M_{\max} < \infty$, then $C_T^{\text{cap}}(M_{\max}) = 0$.

*Roadmap.* We study the intersection of $L(M) \equiv (\kappa_Y - \delta\kappa_K) - \chi\psi\, M^{\phi-\gamma}$ and $R(M) \equiv \delta_M M^{1-\gamma}$. The behaviour of $L$ depends on whether $\phi > \gamma$ (decreasing when $\chi > 0$, constant when $\chi = 0$) or $\phi = \gamma$ (constant); $R$ is strictly increasing when $\delta_M > 0$ and identically zero when $\delta_M = 0$. This yields uniqueness (or $M_{\max} = +\infty$) case by case.

*Proof.* Setting $C_T^{\text{cap}}(M) = 0$ in (3) and dividing both sides by $M^\gamma$ (valid for $M > 0$), we obtain

$$(\kappa_Y - \delta\kappa_K) \;-\; \chi\psi\, M^{\phi-\gamma} \;=\; \delta_M M^{1-\gamma}.$$

Let $L(M) \equiv (\kappa_Y - \delta\kappa_K) - \chi\psi\, M^{\phi-\gamma}$ and $R(M) \equiv \delta_M M^{1-\gamma}$. Since $1 - \gamma > 0$, $R(M)$ is continuous and (strictly) increasing on $(0, \infty)$ when $\delta_M > 0$, and $R(M) \equiv 0$ when $\delta_M = 0$.

- If $\phi > \gamma$, then when $\chi > 0$ the function $L(M)$ is strictly decreasing from $L(0^+) = (\kappa_Y - \delta\kappa_K) > 0$ to $-\infty$; if $\chi = 0$, then $L(M) \equiv (\kappa_Y - \delta\kappa_K) > 0$ is constant.
  - If $\delta_M > 0$, $L$ and $R$ intersect once at some finite $M_{\max}$.
  - If $\delta_M = 0$ and $\chi > 0$, the unique finite $M_{\max}$ solves $L(M_{\max}) = 0$ (when $\phi > \gamma$).
  - If $\delta_M = 0$ and $\chi = 0$, then $L(M) \equiv (\kappa_Y - \delta\kappa_K) > 0$ and $R(M) \equiv 0$, so $C_T^{\text{cap}}(M) > 0$ for all $M$ and $M_{\max} = +\infty$.

- If $\phi = \gamma$, then $L(M) \equiv (\kappa_Y - \delta\kappa_K - \chi\psi) > 0$ by Assumption A.1.
  - If $\delta_M > 0$, there is a unique finite solution $M_{\max} > 0$ to $L(M_{\max}) = R(M_{\max})$.
  - If $\delta_M = 0$, the equality has no finite solution and $M_{\max} = +\infty$.

Summary: $M_{\max} = +\infty$ if and only if $\delta_M = 0$ and either **(A)** $\phi = \gamma$ or **(B)** $\phi > \gamma$ with $\chi = 0$. Otherwise $M_{\max} \in (0, \infty)$, unique.

## A.2 Prices, regimes, and the generalised Euler-return map

The Kuhn–Tucker conditions imply two regimes. In the binding regime, capacity saturates so $D^* = \psi M^\phi$ and $p_D^* \geq \chi$ (with strict inequality if and only if $\eta > 0$); in the slack regime the planner sets $p_D^* = \chi$ by choosing $D^* < D^{\text{cap}}$.

We set the Lagrange multiplier on the tangible resource constraint to $\lambda = U_{C_T}$ and measure prices in $C_T$ units by dividing all FOCs by $U_{C_T}$; equivalently, we take $C_T$ as the numeraire. Let $\tilde{\eta}(M) \geq 0$ denote the raw multiplier on the capacity constraint $D \leq \psi M^\phi$, and define the (numeraire-)normalised capacity multiplier $\eta(M) \equiv \tilde{\eta}(M)/U_{C_T}$ so that all multipliers are in $C_T$ units. The $D$-stationarity condition is

$$p_D - \chi - \eta = 0, \qquad \eta \geq 0, \qquad \eta\Big(\psi M^\phi - D\Big) = 0.$$

Evaluating at capacity implies

$$\eta(M) = \max\{0,\, p_D^{\text{cap}}(M) - \chi\}.$$

In the slack regime, $p_D^* = \chi$ at the optimal $D^* < D^{\text{cap}}$. Increasing $D$ toward capacity raises $D$ and lowers $C_T$, thereby reducing the MRS under CES preferences; hence $p_D^{\text{cap}}(M) \leq \chi$ whenever the capacity constraint is slack. Within period $t$, the stocks $K_t$ and $M_t$ are predetermined.

**Generalised Euler return.**

$$\begin{aligned} G(M) &\equiv (1-\delta_M) + \zeta \frac{Y^*(M)}{M} + \eta(M)\phi\psi M^{\phi-1} \\ &= (1-\delta_M) + \zeta \frac{Y^*(M)}{M} \\ &\quad + \max\Big\{0, p_D^{\text{cap}}(M) - \chi\Big\}\phi\psi M^{\phi-1}, \qquad M \in \mathcal{D}. \end{aligned} \tag{4}$$

We refer to $G$ as the generalised Euler return. Economically, the first two terms are the standard gross return on $M$; we refer to the third term in (4) as the services dividend, which is present only when capacity binds.

Equivalently, resolving the max by regime,

$$G(M) = \begin{cases} (1-\delta_M) + \zeta \, \dfrac{Y^*(M)}{M} + \Big(p_D^{\text{cap}}(M) - \chi\Big)\, \phi\psi \, M^{\phi-1}, & \text{if } p_D^{\text{cap}}(M) > \chi, \\ (1-\delta_M) + \zeta \, \dfrac{Y^*(M)}{M}, & \text{if } p_D^{\text{cap}}(M) \le \chi. \end{cases}$$

We analyse $G$ on $\mathcal{D} = (0, M_{\max})$. In steady state, the Euler condition for $M$ is

$$1 \;=\; \beta \, G(M^*). \tag{5}$$

Note on the productivity term: because $K$ is predetermined within the period, the envelope/Euler return uses the partial marginal product holding $K$ fixed, $\partial Y/\partial M = \zeta \, Y/M$ for Cobb–Douglas $Y = AK^{\alpha}M^{\zeta}$. We then evaluate at the steady-state allocation $K^*(M)$ for notational convenience.

[Monotonicity and continuity of $G$] Under Assumption A.1, on $\mathcal{D} = (0, M_{\max})$ the map $G(\cdot)$ is continuous and strictly decreasing. The at-capacity MRS $p_D^{\text{cap}}(M)$ is (weakly) decreasing in $M$ and is constant on $\mathcal{D}$ if and only if $\delta_M = 0$ and $\phi = \gamma$; equivalently, $p_D^{\text{cap}}$ is strictly decreasing unless this knife-edge holds. Moreover, at any $M$ with $p_D^{\text{cap}}(M) = \chi$, the services term is zero on the slack side and tends to zero from the binding side, hence $G$ is continuous (though generally not differentiable) at the endogenous regime boundary. Because $G$ is strictly decreasing on $\mathcal{D}$, it is injective.

Proof: We analyse each component of (4).

(i) *Productivity term.* Since $Y^*(M) = \kappa_Y M^{\gamma}$ with $\gamma \in (0,1)$, $\zeta \, Y^*(M)/M = \zeta\kappa_Y M^{\gamma-1}$ is

strictly decreasing in $M$.

(ii) *At-capacity MRS monotonicity.* Using (3) and $D^{\text{cap}} = \psi M^{\phi}$, first divide by $D^{\text{cap}}$:

$$\frac{C_T^{\text{cap}}}{D^{\text{cap}}} = \frac{\kappa_Y - \delta\kappa_K}{\psi} M^{\gamma-\phi} \; - \; \frac{\delta_M}{\psi} M^{1-\phi} \; - \; \chi = c_1 M^{\gamma-\phi} - c_2 M^{1-\phi} - \chi,$$

with $c_1 = \frac{\kappa_Y - \delta\kappa_K}{\psi} > 0$ and $c_2 = \frac{\delta_M}{\psi} \geq 0$. Differentiating gives

$$\frac{d}{dM}\left(\frac{C_T^{\text{cap}}}{D^{\text{cap}}}\right) = c_1(\gamma-\phi)M^{\gamma-\phi-1} - c_2(1-\phi)M^{-\phi} \; \leq \; 0,$$

because $c_1 > 0$, $c_2 \geq 0$, and under Assumption A.1 we have $\phi \in (0, 1]$ and $\phi \geq \gamma$, so $(\gamma - \phi) \leq 0$ and $(1 - \phi) \geq 0$. Since $C_T^{\text{cap}} > 0$ and $D^{\text{cap}} > 0$ on $\mathcal{D}$, the ratio and hence

$$p_D^{\text{cap}}(M) = \frac{1-\theta}{\theta}\left(\frac{C_T^{\text{cap}}}{D^{\text{cap}}}\right)^{1/\sigma}$$

are continuous and (weakly) decreasing in $M$ on $\mathcal{D}$ (strictly decreasing off the knife-edge).

(iii) *Services dividend.* Let $H(M) \equiv \max\{0,\, p_D^{\text{cap}}(M) - \chi\}\, \phi\psi\, M^{\phi-1}$. In the binding regime ($p_D^{\text{cap}} > \chi$),

$$H'(M) = \phi\psi\Big[p_D^{\text{cap}\prime}(M)M^{\phi-1} + \big(p_D^{\text{cap}}(M) - \chi\big)(\phi-1)M^{\phi-2}\Big] < 0.$$

Indeed, the first term is non-positive since $p_D^{\text{cap}\prime}(M) \leq 0$. If $\phi < 1$, the second term is strictly negative. If $\phi = 1$, then

$$\frac{d}{dM}\left(\frac{C_T^{\text{cap}}}{D^{\text{cap}}}\right) = c_1(\gamma-1)M^{\gamma-2} < 0 \quad (\gamma < 1),$$

so $p_D^{\text{cap}\prime}(M) < 0$ and therefore $H'(M) < 0$. In the slack regime ($p_D^{\text{cap}} \leq \chi$), $H(M) \equiv 0$, and

$$G'(M) \; = \; \zeta\kappa_Y(\gamma-1)M^{\gamma-2} \; < \; 0.$$

Summing (i)–(iii) shows $G$ is continuous and strictly decreasing on $\mathcal{D}$ and continuous (though kinked) at any endogenous regime switch.

## A.3 Existence, uniqueness and interiority

Because $G$ is strictly decreasing, existence by the Intermediate Value Theorem requires that its limit as $M \uparrow M_{\max}$ lies below the inverse discount factor:

$$\beta\, G(M_{\max}^{-}) \;<\; 1.$$

Since $\lim_{M\downarrow 0} G(M) = +\infty$ and $G$ is strictly decreasing on $\mathcal{D}$ (Lemma A.2), this condition is equivalent to existence (and hence uniqueness) of an interior steady state.

**Assumption A.2 (Boundary return below inverse discount factor).**

$$\beta^{-1} \;>\; (1-\delta_M) \;+\; \zeta \lim_{M\uparrow M_{\max}} \frac{Y^*(M)}{M}.$$

When $M_{\max} < \infty$ this reduces to $\beta^{-1} > (1-\delta_M) + \zeta\,\kappa_Y M_{\max}^{\gamma-1}$. When $M_{\max} = +\infty$ (which by Lemma A.1 occurs only if $\delta_M = 0$ and either (A) $\phi = \gamma$ or (B) $\phi > \gamma$ with $\chi = 0$), the limit term is 0 (since $\gamma < 1$), and A.2 becomes $\beta^{-1} > 1 - \delta_M = 1$, automatic for $\beta \in (0,1)$.

[Existence, uniqueness and interiority] Under Assumptions A.1–A.2 there exists a unique $M^* \in (0, M_{\max})$ solving (5). No steady state exists with $M \geq M_{\max}$. The associated steady state is feasible and interior $\Big(\mathrm{C}_T^{\mathrm{cap}}(M^*) > 0$, hence $C_T^*(M^*) \geq C_T^{\mathrm{cap}}(M^*) > 0$ and $D^* > 0\Big)$.

Proof: As $M \downarrow 0$, $\zeta Y^*(M)/M = \zeta\kappa_Y M^{\gamma-1} \to +\infty$ because $0 < \gamma < 1$. Hence $\lim_{M\downarrow 0} G(M) = +\infty$.

If $M_{\max} < \infty$, then as $M \uparrow M_{\max}$ we have $C_T^{\mathrm{cap}}(M) \downarrow 0$ so $\frac{C_T^{\mathrm{cap}}}{D^{\mathrm{cap}}} \to 0$ and thus $p_D^{\mathrm{cap}}(M) \to 0$, implying $H(M) \to 0$. Therefore

$$\lim_{M\uparrow M_{\max}} G(M) = (1-\delta_M) + \zeta\,\frac{Y^*(M_{\max})}{M_{\max}} \;<\; \beta^{-1}$$

by Assumption A.2. If $M_{\max} = +\infty$ (only when $\delta_M = 0$ and either (A) $\phi = \gamma$ or (B) $\phi > \gamma$ with $\chi = 0$), then $\lim_{M\to\infty} \zeta Y^*(M)/M = 0$ and $H(M) \to 0$ (since either $\phi < 1$ or $p_D^{\mathrm{cap}}(M) \to 0$ when $\phi = 1$). Thus $\lim_{M\to\infty} G(M) = (1-\delta_M) = 1 < \beta^{-1}$. By Lemma A.2, $G$ is continuous and strictly decreasing on $(0, M_{\max})$, and the Intermediate Value Theorem yields a unique $M^* \in (0, M_{\max})$ with $1 = \beta G(M^*)$.

No steady state for $M \geq M_{\max}$. If $M_{\max} < \infty$, then for all $M \geq M_{\max}$ we have $C_T^{\mathrm{cap}}(M) \leq 0$.

Feasibility requires $C_T^* > 0$, and using $C_T^* = C_T^{\text{cap}}(M) + \chi\Big(D^{\text{cap}}(M) - D^*\Big)$, this implies

$$\chi\Big(D^{\text{cap}}(M) - D^*\Big) \;>\; -\,C_T^{\text{cap}}(M) \;\geq\; 0.$$

If $\chi > 0$, then necessarily $D^* < D^{\text{cap}}(M)$, i.e., the regime must be slack on $[M_{\max}, \infty)$. (If $\chi = 0$, then $C_T^* \leq 0$, so the allocation is infeasible regardless of regime.) In the slack regime the Euler RHS is $(1-\delta_M) + \zeta\kappa_Y M^{\gamma-1}$, which is strictly decreasing in $M$ and, by A.2,

$$\beta\Big[(1-\delta_M) + \zeta\kappa_Y M_{\max}^{\gamma-1}\Big] \;<\; 1,$$

hence it is $< 1$ for all $M \geq M_{\max}$. Thus no steady state exists with $M \geq M_{\max}$.

Interiority. From $M^* \in \mathcal{D}$ we have $C_T^{\text{cap}}(M^*) > 0$, and since $C_T^* = Y^* - \delta K^* - \delta_M M^* - \chi D^* = C_T^{\text{cap}}(M^*) + \chi\Big(D^{\text{cap}}(M^*) - D^*\Big) \geq C_T^{\text{cap}}(M^*)$, it follows that $C_T^* > 0$. For $D^* > 0$: if binding, $D^* = D^{\text{cap}}(M^*) = \psi M^{*\phi} > 0$; if slack, the CES MRS $p_D = \frac{1-\theta}{\theta}(C_T/D)^{1/\sigma} \to \infty$ as $D \downarrow 0$ for any $C_T > 0$, so the FOC $p_D = \chi < \infty$ cannot be satisfied at $D = 0$.

## A.4 Local comparative statics (binding-regime neighbourhood)

Suppose $M^*$ lies in the binding regime and that a small perturbation keeps the steady state in that regime (Section 5 verifies $p_D^*(M^*) \geq \chi$ with strict inequality in the baseline). Let $g(M;\upsilon)$ denote (4) with the max resolved on the binding set,

$$g(M;\upsilon) \;=\; (1-\delta_M) \;+\; \zeta\,\frac{Y^*(M)}{M} \;+\; \Big(p_D^{\text{cap}}(M) - \chi\Big)\,\phi\psi\,M^{\phi-1},$$

and define $F(M;\upsilon) \equiv \beta\, g(M;\upsilon) - 1$. Then $F(M^*;\upsilon) = 0$, $F_M(M^*;\upsilon) = \beta\, g'(M^*) < 0$ by Lemma A.2, and the Implicit Function Theorem yields

$$\frac{\partial M^*}{\partial \upsilon} \;=\; -\,\frac{F_\upsilon}{F_M} \;=\; \boxed{-\,\frac{\partial g/\partial \upsilon}{g'(M^*)}}.$$

Since $g'(M^*) < 0$, it follows that

$$\operatorname{sign}\left(\frac{\partial M^*}{\partial \upsilon}\right) = \operatorname{sign}\left(\frac{\partial g}{\partial \upsilon}\right).$$

*Compact results (holding $M$ fixed):*

$$\text{(1) Effective depreciation } \delta_M: \quad \frac{\partial g}{\partial \delta_M} = -1 \; + \; \frac{\partial p_D^{\text{cap}}}{\partial \delta_M}\, \phi\psi\, M^{\phi-1},$$
$$< 0 \quad \text{since } \frac{\partial p_D^{\text{cap}}}{\partial \delta_M} < 0.$$

$$\text{(2) Marginal resource cost } \chi: \quad \frac{\partial g}{\partial \chi} = \left(\frac{\partial p_D^{\text{cap}}}{\partial \chi} - 1\right) \phi\psi\, M^{\phi-1},$$
$$< -\,\phi\psi\, M^{\phi-1} < 0 \quad \text{since } \frac{\partial p_D^{\text{cap}}}{\partial \chi} < 0.$$

$$\text{(3) Productivity share } \zeta: \quad \frac{\partial g}{\partial \zeta} = \frac{Y^*(M)}{M} \; + \; \zeta\, \frac{\partial}{\partial \zeta}\!\left(\frac{Y^*(M)}{M}\right) \; + \; \phi\psi\, M^{\phi-1}\, \frac{\partial p_D^{\text{cap}}(M)}{\partial \zeta}.$$

*Derivations (compact).* Where $\ln M$ appears, interpret it as $\ln(M/\bar{M})$ for an arbitrary reference stock $\bar{M} > 0$; this normalisation does not affect the sign of derivatives.

- For $\delta_M$: $\partial C_T^{\text{cap}}/\partial \delta_M = -M < 0$ implies $\frac{\partial}{\partial \delta_M}\left(C_T^{\text{cap}}/D^{\text{cap}}\right) = -(1/\psi)M^{1-\phi} < 0 \Rightarrow \frac{\partial p_D^{\text{cap}}}{\partial \delta_M} < 0$.

- For $\chi$: $\partial C_T^{\text{cap}}/\partial \chi = -D^{\text{cap}} < 0$ implies $\frac{\partial}{\partial \chi}\left(C_T^{\text{cap}}/D^{\text{cap}}\right) = -1 \Rightarrow \frac{\partial p_D^{\text{cap}}}{\partial \chi} = -\frac{1}{\sigma}\frac{1-\theta}{\theta}\left(\frac{C_T^{\text{cap}}}{D^{\text{cap}}}\right)^{1/\sigma-1} < 0$.

- For $\zeta$: write $Y^*(M)/M = \kappa_Y M^{\gamma-1}$ with $\gamma = \zeta/(1-\alpha)$ and note $\kappa_Y$ is independent of $\zeta$ under the user–cost condition.[14] Hence

$$\frac{\partial}{\partial \zeta}\left(\frac{Y^*}{M}\right) = \kappa_Y M^{\gamma-1}\, \frac{\ln M}{1-\alpha}.$$

  Moreover,

$$\frac{C_T^{\text{cap}}}{D^{\text{cap}}} = \frac{\kappa_Y - \delta\kappa_K}{\psi}\, M^{\gamma-\phi} - \frac{\delta_M}{\psi} M^{1-\phi} - \chi, \quad \Rightarrow \quad \frac{\partial}{\partial \zeta}\left(\frac{C_T^{\text{cap}}}{D^{\text{cap}}}\right) = \frac{\kappa_Y - \delta\kappa_K}{\psi}\, M^{\gamma-\phi}\, \frac{\ln M}{1-\alpha},$$

  and therefore

$$\frac{\partial p_D^{\text{cap}}}{\partial \zeta} = \frac{1}{\sigma}\frac{1-\theta}{\theta}\left(\frac{C_T^{\text{cap}}}{D^{\text{cap}}}\right)^{1/\sigma-1} \frac{\kappa_Y - \delta\kappa_K}{\psi}\, M^{\gamma-\phi}\, \frac{\ln M}{1-\alpha}.$$

[14] From $Y^* = \Lambda_K K^*$ and $K^* = (A/\Lambda_K)^{1/(1-\alpha)} M^{\zeta/(1-\alpha)}$, we have $\kappa_Y = \Lambda_K (A/\Lambda_K)^{1/(1-\alpha)}$, independent of $\zeta$.

For $M < 1$ the $\ln M$ terms are negative; for $M > 1$ they are positive, so the sign can flip with scale and regime.

**Remark (Regime identification).** The existence/uniqueness results hold on $\mathcal{D}$ irrespective of regime. Quantitative sections focus on the empirically relevant case where $p_D^*(M^*) \geq \chi$ (binding capacity, with strict inequality whenever the capacity multiplier is strictly positive). Note that $G$ is continuous but has a kink at the endogenous switch $p_D^{\text{cap}}(M) = \chi$; if a perturbation flips the regime, recompute comparative statics with the services term set to zero ($H \equiv 0$) and use one-sided derivatives at the boundary.

**Remark (Limit $\chi \downarrow 0$).** As $\chi \downarrow 0$, the services dividend increases and $M^*$ is (weakly) increasing; nevertheless $M^*$ remains finite because $G(\cdot)$ is strictly decreasing and, by Assumption A.2, $\lim_{M \uparrow M_{\max}} \beta\, G(M) < 1$. When $M_{\max} = +\infty$ (which by Lemma A.1 occurs only if $\delta_M = 0$ and either (A) $\phi = \gamma$ or (B) $\phi > \gamma$ with $\chi = 0$), we have

$$\lim_{M \to \infty} \beta\, G(M) = \beta(1 - \delta_M) = \beta < 1,$$

so the steady state still occurs at a finite $M^*$. In particular, when $\delta_M = 0$ and $\phi > \gamma$, $M_{\max} = \left((\kappa_Y - \delta\kappa_K)/(\chi\psi)\right)^{1/(\phi-\gamma)} \to \infty$ as $\chi \downarrow 0$, but the steady state remains finite.

*Optional closed forms for $M_{\max}$.* In the special cases:

$$\begin{aligned}
&\text{(i) } \phi = \gamma,\ \delta_M > 0: && M_{\max} = \left(\frac{\kappa_Y - \delta\kappa_K - \chi\psi}{\delta_M}\right)^{\frac{1}{1-\gamma}}, \\
&\text{(ii) } \delta_M = 0,\ \phi > \gamma,\ \chi > 0: && M_{\max} = \left(\frac{\kappa_Y - \delta\kappa_K}{\chi\psi}\right)^{\frac{1}{\phi-\gamma}}, \\
&\text{(iii) } \chi = 0,\ \delta_M > 0 \text{ (any } \phi \geq \gamma): && M_{\max} = \left(\frac{\kappa_Y - \delta\kappa_K}{\delta_M}\right)^{\frac{1}{1-\gamma}}.
\end{aligned}$$

These follow directly from the proof of Lemma A.1. Additionally, when $\delta_M = 0$ and $\chi = 0$, $M_{\max} = +\infty$, consistent with Lemma A.1.

# B Jacobian, eigenvalues and local dynamics

This appendix summarises the construction of the local general-equilibrium dynamics reported in Section 6. The key point is that, away from the steady state, gross investment is not equal to depreciation. Accordingly, the linearisation keeps investment explicit in the resource

constraint, and the reduced two-dimensional transition for $(k_t, m_t)$ is obtained by solving the full linearised equilibrium system and imposing saddle-path stability.

Throughout, we focus on the empirically relevant binding-capacity regime in which $D_t = \psi M_t^{\phi}$ and $\mathrm{MRS}_t > \chi$ at the steady state.

## B.1 Equilibrium conditions used for the linearisation

Under binding capacity, the relevant equilibrium conditions are

$$C_t^{\mathrm{T}} + I_t^K + X_t + \chi D_t = Y_t, \tag{B.1}$$

$$K_{t+1} = (1-\delta)K_t + I_t^K, \tag{B.2}$$

$$M_{t+1} = (1-\delta_M)M_t + X_t, \tag{B.3}$$

$$Y_t = AK_t^{\alpha} M_t^{\zeta}, \qquad D_t = \psi M_t^{\phi}, \tag{B.4}$$

$$C_t = \Big[\theta (C_t^{\mathrm{T}})^{\rho} + (1-\theta) D_t^{\rho}\Big]^{1/\rho}, \qquad \rho \equiv \frac{\sigma - 1}{\sigma}, \tag{B.5}$$

$$\mathrm{MRS}_t = \frac{1-\theta}{\theta}\Big(\frac{C_t^{\mathrm{T}}}{D_t}\Big)^{1/\sigma}, \tag{B.6}$$

together with the Euler equations

$$\lambda_t = \beta\, \lambda_{t+1}\Big[(1-\delta) + \alpha\, \frac{Y_{t+1}}{K_{t+1}}\Big], \tag{B.7}$$

$$\lambda_t = \beta\, \lambda_{t+1}\Big[(1-\delta_M) + \zeta\, \frac{Y_{t+1}}{M_{t+1}} + (\mathrm{MRS}_{t+1} - \chi)\, \phi\psi\, M_{t+1}^{\phi-1}\Big], \tag{B.8}$$

where $\lambda_t$ is the shadow value of one unit of tangible goods in utility units, $\lambda_t = U_C(C_t)\, \partial C_t / \partial C_t^{\mathrm{T}}$.

**Why replacement investment cannot be used away from steady state** In the steady state, $I^{K*} = \delta K^*$ and $X^* = \delta_M M^*$, but these identities do not hold along transitions. Substituting them into the non-steady-state feasibility condition would eliminate the channel through which investment moves $t$-dated consumption and therefore the co-state $\lambda_t$. Equations (B.1)–(B.8) retain the correct dependence of $\lambda_t$ on investment, which is essential for the local dynamics.

## B.2 Second-order Euler block and reduced state transition

Let $k_t \equiv \ln K_t - \ln K^*$ and $m_t \equiv \ln M_t - \ln M^*$ denote log deviations from the steady state, and collect them in $x_t \equiv (k_t, m_t)'$. Because $\lambda_t$ depends on $(K_t, M_t, K_{t+1}, M_{t+1})$ via

(B.1)–(B.5), log-linearising the Euler block yields a second-order linear difference equation in $x_t$,

$$A\,x_{t+2} + B\,x_{t+1} + C\,x_t = 0, \tag{B.9}$$

where $A$, $B$ and $C$ are $2 \times 2$ matrices of steady-state elasticities.

A convenient way to obtain the reduced two-state law of motion is to form the $4 \times 4$ companion system for $(x_{t+1}, x_t)$ and select the stable eigenspace. Saddle-path stability implies two eigenvalues inside the unit circle, matching the number of predetermined stocks. The resulting local stable-manifold transition takes the form

$$x_{t+1} = J\,x_t. \tag{B.10}$$

## B.3 Baseline numerical Jacobian and eigenvalues

Under the baseline calibration in Table 1, the computed reduced-form Jacobian is

$$J \approx \begin{pmatrix} 0.947 & 0.061 \\ 0.507 & 0.033 \end{pmatrix}, \tag{B.11}$$

with eigenvalues

$$\lambda_1 \approx 7.7 \times 10^{-6}, \qquad \lambda_2 \approx 0.980. \tag{B.12}$$

The near-zero stable root implies a very rapid adjustment mode that operates through endogenous investment choices for the next-period stocks, while the second root governs the slower component of the local convergence. For comparison, the depreciation-based PE survival rates are $1 - \delta_M \approx 0.905$ and $1 - \delta \approx 0.985$. The GE eigenvalues differ markedly from these PE benchmarks because the corrected linearisation keeps investment endogenous in the feasibility constraint, allowing the planner to re-optimise gross investment rather than holding it at replacement. Figures 1–3 display the resulting GE paths alongside the PE benchmarks.