Artificial Intelligence in the Macroeconomy: Prices, Adoption, and Distribution

Helen Popper* Christoph Schult†

DRAFT — March 12, 2026

Abstract

AI is a produced intermediate input whose current output and allocation affect its future cost. We develop a tractable general equilibrium model in which an AI producer with market power learns by doing and allocates part of its output to self-use in AI production. Because current output lowers future production costs, pricing is forward-looking and adoption evolves endogenously.

In a Cobb–Douglas benchmark, the economy admits a generalized balanced growth path in which AI output grows faster than final output, the relative price of AI falls, and real wages rise with aggregate output. The benchmark yields a closed-form condition for stable, non-explosive growth.

With CES technologies, we show that the effective demand elasticity faced by the AI producer is the sum of input substitution in final-goods production and a general-equilibrium scale effect. Simulations then yield two adoption regimes.

When inputs in final-goods production are complements, the AI producer expands self-use first and scales external adoption later, so aggregate responses remain modest despite rapid AI price declines. When final-goods inputs are substitutes, expansion is scale-first and the two-phase pattern weakens.

The wage premium for specialized labor responds to the same elasticities: it is lowest under complementarity and rises with substitutability in final-goods production. Greater substitutability in AI production amplifies both diffusion and wage-premium dynamics.

*Santa Clara University, Email: hpopper@scu.edu

†Halle Institute for Economic Research, Email: Christoph.Schult@iwh-halle.de

1 Overview

Recent advances in AI combine rapid scale-up, recursive reuse of existing models, and experience-driven improvements. This combination makes the production of AI itself a central determinant of its subsequent adoption and its macroeconomic effects. Standard macroeconomic frameworks typically take the AI technology path as given. We instead incorporate an explicit AI-production sector into a general equilibrium model. The inclusion of an AI sector allows us to explore the dynamics of AI pricing and adoption, and of profits, growth and the relative wage of specialized labor.

In the model, AI is a produced intermediate input supplied by a monopolist that learns by doing and allocates part of its output to internal use in AI production.[1] This creates an intertemporal pricing problem: the producer trades off current revenue against faster future cost reductions and wider adoption. As a result, the AI price path, adoption, and profits are endogenous objects rather than exogenous inputs.

The mechanisms embedded in the model reflect central features of contemporary AI development. Specifically, the combination of learning by doing and AI's use in its own production allows for a rich characterization of AI development cycles. Automated design, coding and data processing—once instantiated—can be reused to refine or extend themselves; and existing models and software libraries are used to generate better subsequent models. While learning by doing reflects the accumulation of knowledge through repeated production, recursive self-use represents a strategic decision: the producer's internal use of AI enhances future capabilities but comes with an opportunity cost in terms of foregone sales.

The dual mechanism underscores both the cumulative nature of experience-driven improvements and the allocation decision inherent in recursive use. The result is that the AI producer internalizes the mechanism via a dynamic term in its pricing rule, which early on pushes the price below a standard, static benchmark price.

We introduce final-goods producers as monopolistic competitors; and both AI and final goods are produced using a constant-returns-to-scale technology in each period. We show that this set-up implies that learning by doing dampens the AI monopolist's markup, the production of AI accelerates, and final-goods production rises. We also distinguish between specialized and nonspecialized labor; and we characterize the behavior of their relative wages and of profits.

After introducing the model and some general results, we focus on the special case of Cobb-Douglas production. This allows us to derive a generalized balanced growth path (GBGP) analytically. With recursive self-use and learning by doing,

[1] Monopoly is a stand-in for strong market power; extension to oligopoly would be natural.

the existence of a GBGP imposes tight restrictions linking AI growth, final output growth, and relative prices. The GBGP characterizes the long-term growth rates of AI, of final production, and of the relative prices of AI, of energy, and of wages. We show that AI production grows faster than final output in this benchmark, and that the price of AI falls continuously toward zero. In this simple case, the wage of specialized labor relative to nonspecialized labor remains constant. The GBGP provides a closed-form long-run benchmark. Specifically, it pins down: the condition for non-explosive growth with constant shares, the relative growth of AI vs. final output, and the trend in the AI price.

Next, we move from Cobb–Douglas to a CES specification and simulate the model over a range of elasticities. Drawing on the classic scale-versus-substitution logic of derived demand (and on Minton and Mulligan's (2024) market-equilibrium application of that logic), we show that in our macro-AI setting the AI producer's effective demand elasticity can be written as the simple linear sum of input substitution in final-goods production and a general-equilibrium scale effect. This decomposition organizes the simulation results.

The AI producer's forward-looking pricing produces a sharp early decline in the AI price, and the distribution of income evolves over time in ways that depend on these elasticities. Most prominently, outcomes depend on input substitutability in the production of final-goods. The simulations reveal two distinct early-adoption patterns: when final-goods inputs are complements, external AI adoption is relatively slow, and AI production is learning-first and capital-light; when final-goods inputs are substitutes, external adoption occurs quickly, and AI production is scale-driven and capital-deepening. Additionally, throughout, greater substitutability in final-goods production raises the specialized–to–nonspecialized wage premium.[2]

These mechanisms and distributional patterns connect to several strands of the macro–AI literature. Jones and Tonetti (2026) use the task framework of Acemoglu and Restrepo (e.g., Acemoglu and Restrepo 2018, 2020) and show that when tasks are complements ("weak links"), rapid machine improvements translate into only gradual aggregate gains. Our model generates a similar phenomenon: the interaction of elasticities governs the effective demand elasticity for produced AI and—through it—the adoption dynamics. Here, the role of complementarity among inputs in final-goods production is comparable to the weak links (or bottlenecks) that can arise in the task-based approach. Our work also draws on Korinek and Stiglitz's (2019) general equilibrium model in which AI can substitute for labor. We complement these approaches by endogenizing

[2]Related ongoing work extends the model to allow for heterogeneous households. The heterogeneous approach allows for the characterization of the distribution of welfare, not just income, while incorporating participation utility from work (alongside disutility from hours) and the role of entrepreneurial contributions.

the dynamics of AI production and pricing, and by characterizing the associated implications for capital, profits, and the wage of specialized labor relative to that of nonspecialized labor.[3]

In separating the two types of labor (specialized and nonspecialized), our work is in the spirit of Benzell and Brynjolfsson's (2019) distinction between scarce 'genius' labor and ordinary labor. Benzell and Brynjolfsson develop a general equilibrium model in which the replicability of digital technologies raises the returns to scarce "genius" labor while compressing ordinary wages and the interest rate. Our contribution extends this insight by embedding the distinction between labor types in a dynamic framework with endogenous AI production; this allows us to characterize not only distributional outcomes but also the long-run evolution of AI production, prices, and profits.[4]

Our treatment of AI as an intermediate input with recursive use connects directly to other recent work differentiating AI from earlier technologies. Aghion, Jones, and Jones (2017) model AI as a factor in idea production that can substitute for researchers; Brynjolfsson and Mitchell (2017) emphasize AI as a general-purpose technology; Cockburn, Henderson, and Stern (2019) describe AI as an "invention of a method of invention" with broad spillovers; and Veldkamp and Chung (2024) model data as a factor of production with increasing returns. Our approach offers a way to capture aspects of these ideas within a unified macroeconomic framework. Taken together, these connections situate the model within the broader macroeconomic analysis of AI.

2 The Model Setup

2.1 AI Production

We treat AI as an intermediate good produced by a monopolist. The monopolist benefits from learning by doing and produces AI (A_t) using inputs of capital ($K_{A,t}$), specialized labor (${L^s}_{A,t}$), energy ($E_{A,t}$), and recursively using AI as an input, ($A_{A,t}$):

$$A_t = (A_{t-1})^{\phi} f(K_{A,t}, L^s_{A,t}, E_{A,t}, A_{A,t}) \tag{2.1}$$

[3]Both Acemoglu–Restrepo and Korinek–Stiglitz allow AI to substitute for human labor, potentially putting downward pressure on wages through displacement channels. We abstract from those channels by treating AI as an additional input to production then analyze how recursive self-use and learning shape AI pricing, adoption, profits, and the skilled–nonskilled wage gap.

[4]Like their work, ours also reflects (though to a lesser extent) the issues raised in Brynjolfsson, Chandar, and Chen's (2025) empirical study of generative AI's labor-market impacts.

where $\phi \in (0,1)$ is a parameter that governs learning by doing based on the producer's prior output, AI is used recursively as an input into AI, and t denotes the time period. We assume that f itself is constant returns to scale in each period.

Note also that while production exhibits constant returns to scale in each period (the distribution parameters sum to one), it exhibits dynamically increasing returns to scale since $\phi > 0$. At the same time, the restriction that $\phi < 1$ prevents explosive dynamics and keeps the model tractable.[5]

2.1.1 AI Producer's Problem

The AI monopolist maximizes the present discounted value of future profits:

$$\max_{\{A_t, K_{A,t}, L^s_{A,t}, E_{A,t}\}_{t=0}^{\infty}} \sum_{t=0}^{\infty} \beta^t \Pi_{A,t}, \tag{2.2}$$

subject to the AI production function, Equation 2.1, and period profits are:

$$\Pi_{A,t} = p_{A,t} A_t - w^s_t L^s_{A,t} - (r_t + \delta) K_{A,t} - p_{E,t} E_{A,t} - p_{A,t} A_{A,t},$$

where $p_{A,t},\ w^s_t,\ p_{E,t}$ are the period-t prices of AI, specialized labor, and energy; and r_t is the interest rate.

The AI monopolist's Bellman equation is:

$$\begin{aligned} V(A_{t-1}) = \max_{A_t} \Big\{ & p_{A,t}\, A_{Y,t} \ - \ C_t(A_t) \ + \ \beta\, V(A_t) \Big\} \\ \text{s.t.} \quad & p_{A,t} = \mathcal{P}_t(A_{Y,t}) \\ & A_{Y,t} = A_t - A_{A,t}, \end{aligned} \tag{2.3}$$

where $\mathcal{P}_t(\cdot)$ is the inverse demand function for AI, and $C_t(A_t)$ is the cost-minimizing expenditure to produce A_t using $K_{A,t}, L^s_{A,t}, E_{A,t}, A_{A,t}$ at factor prices $(r_t, w^s_t, p_{E,t}, p_{A,t})$, with $A_{A,t}$ determined by cost minimization conditional on $A_{Y,t}$.[6]

The first order condition is then:

$$p_{A,t}\Big(1 - \tfrac{1}{\epsilon_t}\Big) \ = \ MC_{A,t} \ - \ \beta\, V'(A_t), \tag{2.4}$$

where marginal cost is $MC_{A,t} \equiv C'_t(A_t)$; and ϵ_t denotes the effective (positive) price elasticity of demand for AI faced by the monopolist.

[5] Nordhaus (2021), among others, explores the possibility of such explosive dynamics.

[6] We also impose a participation constraint: $V_{A,0} = \sum_{t \geq 0} \beta^t \Pi_{A,t} \geq 0$.

Note that, since AI is an intermediate good, this effective elasticity (characterized in the appendix) reflects two distinct margins: input substitution in final-goods production and a general-equilibrium scale effect.

2.1.2 Dynamic Pricing with Self-Use and Learning.

Accounting for both the benefits of learning-by-doing and the recursive use of AI in its own production, the monopolist sets prices dynamically to maximize the present value of profits.

Specifically, solving the first order condition, (2.4), for price gives the dynamic monopoly pricing rule:

$$p_{A,t} = \frac{MC_{A,t} - \beta\, V'(A_t)}{1 - \frac{1}{\epsilon_t}}. \tag{2.5}$$

The term $\beta\, V'(A_t)$ captures the value of turning today's output into future capability: part from experience (learning-by-doing) and part from self-use (own output used as an input). Because $\beta\, V'(A_t) > 0$, the optimal price is strictly below the standard static monopoly price: $\frac{MC_{A,t} - \beta\, V'(A_t)}{1-\frac{1}{\epsilon_t}} < \frac{MC_{A,t}}{1-\frac{1}{\epsilon_t}}$.

That is, Equation 2.5 implies that to maximize long-run profits, the AI monopolist lowers prices today to expand production, accelerate learning, and reduce future costs.

2.2 Final-Goods Production

We assume a continuum of differentiated final goods indexed by $i \in [0, 1]$. Each variety $Y_t(i)$ is produced by a firm with constant returns to scale:

$$Y_t(i) = Z_t h(K_{Y,t}(i), L^s_{Y,t}(i), L^n_{Y,t}(i), E_{Y,t}(i), A_{Y,t}(i)), \tag{2.6}$$

where Z_t denotes total factor productivity in the final-goods sector and is assumed to grow exogenously at the rate g_z.[7]

Aggregate final output is a CES aggregate over the differentiated varieties:

$$Y_t = \left(\int_0^1 Y_t(i)^{\frac{\varepsilon-1}{\varepsilon}}\, di \right)^{\frac{\varepsilon}{\varepsilon-1}}, \tag{2.7}$$

where $\varepsilon > 1$ is the household's elasticity of substitution between the differentiated final goods.

Note that, in general equilibrium, this elasticity, ε is distinct from—and absent from—the effective elasticity faced by the AI monopolist, ϵ_t, which (as

[7]To ensure that an elasticity of substitution, σ_Y, is well defined, we assume that the production function is twice continuously differentiable and quasi-concave with positive input use.

described in the appendix) includes a general-equilibrium scale effect that varies over time.

Demand for each variety is then:

$$Y_t(i) = \left(\frac{p_t(i)}{P_t}\right)^{-\varepsilon} Y_t, \tag{2.8}$$

where $p_t(i)$ is the price of variety i, and P_t is the final-goods price index:

$$P_t = \left(\int_0^1 p_t(i)^{1-\varepsilon}\, di\right)^{\frac{1}{1-\varepsilon}} \tag{2.9}$$

We define the final good as the numeraire: $P_t = 1$. Under symmetry all firms choose the same price in equilibrium, so $p_t(i) = 1\ \forall i$, and the aggregate price index remains constant.

Each final-goods firm chooses its price and input quantities to maximize profits (subject to the production function and CES demand), where profits for variety i in period t are given by:

$$\begin{aligned}\Pi_{Y,t}(i) = p_t(i)Y_t(i) \\ - w_t^s L_{Y,t}^s(i) - w_t^n L_{Y,t}^n(i) \\ - (r_t + \delta)K_{Y,t}(i) - p_{E,t}E_{Y,t}(i) - p_{A,t}A_{Y,t}(i) - f,\end{aligned} \tag{2.10}$$

where f is a fixed cost; and we assume free entry.

The pricing first-order condition implies:

$$p_t(i) = \mu_Y \cdot MC_t(i), \quad \mu_Y = \frac{\varepsilon}{\varepsilon - 1}. \tag{2.11}$$

That is, final goods are priced at a constant markup over marginal cost; and the markup, as is standard in such models, depends on the household's elasticity of substitution between the differentiated final goods.

2.3 The Household

The representative household consumes a continuum of differentiated final-goods varieties $C_t(i)$ for $i \in [0, 1]$, which aggregate into a composite good C_t according to a CES aggregator:

$$C_t = \left(\int_0^1 C_t(i)^{\frac{\varepsilon-1}{\varepsilon}}\, di\right)^{\frac{\varepsilon}{\varepsilon-1}}, \quad \varepsilon > 1. \tag{2.12}$$

The household derives utility from this composite good and supplies labor inelastically, and lifetime utility is given by a standard CRRA function:

$$\max \sum_{t=0}^{\infty} \beta^t \frac{C_t^{1-\sigma_c} - 1}{1 - \sigma_c}, \quad \sigma_c > 0, \sigma_c \neq 1. \tag{2.13}$$

The household earns income from supplying both types of labor, renting capital, and receiving monopoly profits from the AI firm.[8] It allocates income between consumption of final goods and saving:[9]

$$\int_0^1 p_t(i)[C_t(i) + S_t(i)]di = w_t^s L_t^s + w_t^n L_t^n + (r_t + \delta)K_t + \Pi_{A,t}.$$

Since we assume symmetry across goods, we can write: $P_t = p_t(i)$, $C_t = C_t(i)$, and $S_t = S_t(i)$; and, this simplifies to:

$$C_t + S_t = w_t^s L_t^s + w_t^n L_t^n + (r_t + \delta)K_t + \Pi_{A,t} \tag{2.14}$$

The household chooses consumption and saving subject to this constraint and the intertemporal accumulation of capital. This yields the standard Euler equation:

$$\frac{C_{t+1}}{C_t} = [\beta(1 + r_{t+1})]^{\frac{1}{\sigma_c}}. \tag{2.15}$$

2.4 Factor Markets

We assume that specialized labor, nonspecialized labor, and energy are all fixed, that capital accumulates endogenously, and that markets clear. So, we have the following conditions:

Labor Markets

$$L_{Y,t}^s + L_{A,t}^s = \bar{L}^s$$
$$L_{Y,t}^n = \bar{L}^n$$

AI Market

$$A_{Y,t} + A_{A,t} = A_t$$

Energy Market

$$E_{Y,t} + E_{A,t} = \bar{E}$$

[8]Note that we assume symmetry and free entry in the final-goods sector, so final-goods firms make zero profits in equilibrium, as in Dixit and Stiglitz (1977).

[9]We assume the household does not buy energy directly.

Interest Rate

$$r_t = MPK_Y - \delta = MPK_A - \delta.$$

Capital Accumulation[10]

$$\begin{aligned} K_{Y,t+1} &= (1-\delta)K_{Y,t} + I_{Y,t} \\ K_{A,t+1} &= (1-\delta)K_{A,t} + I_{A,t} \\ S_t &= I_{Y,t} + I_{A,t} \end{aligned}$$

Note that S_t reflects household saving out of income, which includes labor income, capital income, and monopoly profits.

3 A Generalized Balanced Growth Path

We derive a generalized balanced growth path (GBGP) that characterizes the long-run equilibrium of an economy with recursively produced AI and learning by doing. The GBGP pins down when self-use and learning jointly sustain stable, non-explosive growth with constant factor shares; and it delivers closed-form expressions for long-run growth rates and relative prices. In what follows, a generalized balanced growth path refers to a path with constant growth rates and constant cost shares in both sectors.

In order to analytically characterize this path, we specify both production of AI and of final goods to be Cobb-Douglas at the firm level.

So, AI production becomes:

$$A_t = A_{t-1}^{\phi} K_{A,t}^{\alpha_K} (L_{A,t}^s)^{\alpha_{L^s}} E_{A,t}^{\alpha_E} A_{A,t}^{\alpha_A}, \tag{3.1}$$

where the αs represent the distribution parameters, with α_A capturing the recursive use of AI; and, as before, ϕ represents the extent of learning by doing.

Final-goods production (setting $Z_t = 1 \forall t$) becomes:

$$Y_t = K_{Y,t}^{\omega_K} (L_{Y,t}^s)^{\omega_{L^s}} (L_{Y,t}^n)^{\omega_{L^n}} E_{Y,t}^{\omega_E} A_{Y,t}^{\omega_A} \tag{3.2}$$

where the ωs represent the distribution parameters in final production.

In each case, we assume the distribution parameters sum to one: $\alpha_K + \alpha_{L^s} + \alpha_E + \alpha_A = 1$ and $\omega_K + \omega_{L^s} + \omega_{L^n} + \omega_E + \omega_A = 1$. As in Section 2.2, we have symmetry in equilibrium; So we can treat one representative variety as

[10] We assume that the unconsumed final goods can be converted into capital for use in the production either of AI or of any of the varieties of final goods.

determining aggregate output. Hence, the growth path of the economy can be characterized using a representative firm's Cobb-Douglas production function, Equation 3.2.

3.1 Consumption Growth, the Interest Rate, and Capital–Output Ratios in the GBGP

We begin with the household's consumption growth rate. Since the GBGP requires constant consumption growth, the Euler equation implies a constant real interest rate.

Specifically, the Euler equation (Equation 2.15, reproduced here) relates the growth rate of consumption to the interest rate:

$$\frac{C_{t+1}}{C_t} = [\beta(1 + r_{t+1})]^{\frac{1}{\sigma_c}} .$$

The GBGP requires constant consumption growth. Since the representative household satisfies a standard Euler equation, the growth rate of consumption is determined by the intertemporal preferences of the household. This pins down the real interest rate. So the Euler equation becomes:

$$\frac{C_{t+1}}{C_t} = [\beta(1 + r^*)]^{\frac{1}{\sigma_c}} , \tag{3.3}$$

where r^* is the constant interest rate in the GBGP.

Taking logs of both sides and rearranging gives the constant growth rate of consumption:

$$c^* = \frac{1}{\sigma_c} \log (\beta(1 + r^*)) \tag{3.4}$$

This expression determines the growth rate of consumption as a function of preference parameters and the equilibrium real interest rate.

Turning to the interest rate, we note that each final good variety is produced by a monopolistically competitive firm using Cobb-Douglas technology, and although firms set prices with a markup, they minimize cost conditional on output. As a result, input demands satisfy standard marginal productivity conditions. Because final-goods firms price at a constant markup μ_Y, and the final good is the numeraire ($P_t = 1$), marginal cost satisfies $MC_t = 1/\mu_Y$. Hence factor payments equal marginal products scaled by $1/\mu_Y$. Specifically, cost minimization implies:

$$r^* + \delta = \frac{1}{\mu_Y} \omega_K \frac{Y_t(i)}{K_{Y,t}(i)} = \alpha_K \frac{A_t}{K_{A,t}}.$$

Under symmetry, $Y_t(i) = Y_t$ and $K_{Y,t}(i) = K_{Y,t}$, and this simplifies to:

$$r^* + \delta = \frac{1}{\mu_Y}\,\omega_K \frac{Y_t}{K_{Y,t}} = \alpha_K \frac{A_t}{K_{A,t}}.$$

So the marginal product of capital is constant and equal across the two sectors. This implies constant capital–output ratios, and the following growth rate equalities:

$$g^* \equiv g_Y = g_{K_Y}, \tag{3.5}$$
$$g_A = g_{K_A}, \tag{3.6}$$

where g^* is the growth rate of final-goods output, g_{K_Y} is the growth rate of capital in the final-goods sector, and g_A and g_{K_A} are the growth rates of AI output and of capital used in AI production, respectively.

3.2 AI Production in GBGP

The growth rate of AI production is determined by learning-by-doing, recursive AI use, and capital accumulation. Taking logs and growth rates of the AI production function (Equation 3.1) gives:

$$g_A = \phi g_A + \alpha_K g_{K_A} + \alpha_{L^s} g_{L^s} + \alpha_E g_{E_A} + \alpha_A g_{A_A}, \tag{3.7}$$

where, as above, $g_A = g_{K_A}$ denote the growth rates of AI production and of the capital used in its production; and g_{L^s}, g_{E_A}, and g_{A_A} denote the growth rates of the inputs of specialized labor, energy and AI itself into AI production.

We first pare this down by noting that under the Cobb-Douglas structure, cost shares remain constant (and in this case, equal to the distribution parameters).[11] Because the total supplies of labor and of energy are fixed, their prices must rise at the rate of output to preserve these cost shares. So, the wage of specialized and nonspecialized labor both grow at the rate of output in the GBGP:

$$g_{w^s} = g_{w^n} = g^* \tag{3.8}$$
$$g_{p_E} = g^*. \tag{3.9}$$

And, the wage of specialized labor relative to nonspecialized labor remains constant.

[11] Here, and throughout, the constancy of shares refers to the constant cost (value) shares implied by Cobb-Douglas cost minimization: with changing prices, constant shares need not imply that input quantities grow at the same rate as output.

Correspondingly, the quantities allocated to each sector remain constant over time. That is, their growth rates are zero:

$$g_{L_A^s} = g_{L_Y^s} = 0, \tag{3.10}$$

$$g_{E_A} = g_{E_Y} = 0. \tag{3.11}$$

Then, with $g_{K_A} = g_A$ (following Equation 3.6), Equation 3.7 becomes: $g_A = \phi g_A + \alpha_K g_A + \alpha_A g_{A_A}$. This gives us the following expression for the growth of AI production:

$$g_A = \kappa g_{A_A}. \tag{3.12}$$

where:

$$\kappa = \frac{\alpha_A}{1 - \phi - \alpha_K}.$$

This expression relates the growth rate of AI production, g_A, to the growth rate of the recursive AI input, g_{A_A}. And the coefficient, κ, captures the extent to which recursive use amplifies AI output growth through accumulated learning by doing and accumulated capital. The larger the coefficient, the more recursive AI use is amplified in AI production over time.

Note that if $\kappa < 1$, the recursive AI share rises over time and would eventually exhaust total AI output. Further, an interior constant-share generalized balanced growth path requires that the recursive share remain constant. So $\kappa = 1$, which implies:

$$\phi = 1 - \alpha_A - \alpha_K.$$

And, correspondingly:

$$g_{A_A} = g_A. \tag{3.13}$$

That is, in the GBGP, the recursive use of AI grows at the same rate as the production of AI itself. We adopt this restriction in the Cobb–Douglas benchmark to obtain closed-form long-run expressions; the CES economy relaxes this property and is analyzed numerically.

With $\kappa = 1$, we can read off the long-run objects in closed form: AI grows faster than final output, the AI price falls every period, real wages grow with output, and the specialized/nonspecialized wage ratio is flat in this benchmark.

3.3 AI Prices and Profits in the GBGP

As above, each final good variety is produced by a monopolistically competitive firm that sets price as a constant markup over marginal cost. Cost minimization implies that the marginal product of AI input equals its marginal cost divided by the final-goods producer's markup. Thus, the price of AI is:

$$p_{A,t} = \frac{\omega_A}{\mu_Y} \cdot \frac{Y_t}{A_{Y,t}}. \tag{3.14}$$

Taking logs and differentiating with respect to time yields the growth rate of the price of AI:

$$g_{p_A} = g_Y - g_{A_Y}$$

Combining this with Equation 3.16, we obtain

$$g_{p_A} = \left(\frac{\omega_A}{1 - \omega_K} - 1 \right) g_{A_Y}. \tag{3.15}$$

Since constant returns to scale implies $\omega_A < 1 - \omega_K$, we have $g_{p_A} < 0$: that is, the price of AI falls continuously in the GBGP.

Because factor shares and markups are constant on the GBGP, per-period AI profits grow at the same rate as final output. Hence the AI profit share is constant and is pinned down by the AI markup, the AI share in final-goods costs, and the steady self-use share.

3.4 Final-Goods Production Growth in GBGP

Since each final good variety is produced using a Cobb-Douglas technology, and—as above—symmetry allows us to treat the growth rate of a representative firm's output as determining the growth rate of aggregate output, its growth rate is:

$$g_Y = \omega_K g_{K_Y} + \omega_{L^s} g_{L^s} + \omega_{L^n} g_{L^n} + \omega_E g_E + \omega_A g_{A_Y}, \tag{3.16}$$

As above, capital grows at the same rate as output under the constant GBGP interest rate. And since labor and energy are fixed in supply, cost minimization requires their allocation across sectors to remain constant over time. So, $g_{L^s} = g_{L^n} = g_E = 0$; and Equation 3.16 becomes $g^* = \omega_K g^* + \omega_A g_{A_Y}$, or:

$$g^* \equiv g_Y = \frac{\omega_A}{1 - \omega_K} \, g_{A_Y} \tag{3.17}$$

The growth rate of final output is proportional to the growth of AI used in its production, where the proportionality factor depends simply on the distribution parameters for capital and AI, ω_A and ω_K; and—since constant returns to scale impose that their sum is less than one—the proportionality factor is less than one. That is, final output grows more slowly than the AI input.

To summarize: in this GBGP, AI production grows faster than final output, so the price of AI falls continuously. Real wages of both specialized and nonspecialized labor rise in step with overall output, while their relative wage remains constant. Finally, the real interest rate is pinned down by household preferences.

In contrast to the standard neoclassical steady state with fixed factors and no endogenous technological amplification, here recursive self-use together with learning-by-doing sustains positive balanced growth even when labor and energy are fixed. The real interest rate is pinned down by preferences through the Euler equation, while growth is driven by the endogenous dynamics of AI production.

This interplay makes the Cobb–Douglas case a useful benchmark, clarifying how learning and recursion sustain balanced growth, and setting the stage for the more flexible CES setting.

4 CES Production: Simulation

In this section, we relax the unit elasticity assumption and specify both AI and final-goods production as CES. Keeping period-by-period constant returns and the market structure intact yields a simple expression for the effective demand elasticity faced by the AI producer: it is the sum of input substitution in final-goods production and a general-equilibrium scale effect. We first lay out the CES setup, then we turn to the simulations.

4.1 CES Specification

We begin by specifying the CES production function for AI.

$$A_t = (A_{t-1})^{\phi} \left[\alpha_K^{\frac{1}{\sigma_A}} K_{A,t}^{\rho_A} + \alpha_{L^s}^{\frac{1}{\sigma_A}} (L_{A,t}^s)^{\rho_A} + \alpha_E^{\frac{1}{\sigma_A}} E_{A,t}^{\rho_A} + \alpha_A^{\frac{1}{\sigma_A}} (A_{A,t})^{\rho_A} \right]^{\frac{1}{\rho_A}}. \tag{4.1}$$

As before: $\phi \in (0,1)$ is a parameter that governs learning by doing based on the producer's prior output; $\alpha_A \in (0,1)$ is the distribution parameter of AI used recursively as an input into AI; from constant returns to scale, $\alpha_K + \alpha_{L^s} + \alpha_E + \alpha_A = 1$; and $\sigma_A = \frac{1}{1-\rho_A}$ governs the degree of input substitutability.[12]

[12] In the numerical work below, we generalize the production function to allow for the learning by doing to change over time.

In final-goods production, we again assume a continuum of differentiated final goods indexed by $i \in [0, 1]$. Each variety $Y_t(i)$ is now produced by a firm using a CES technology, with a fixed cost (f) and free entry:

$$Y_t(i) = Z_t \left[\omega_K^{\frac{1}{\sigma_Y}} K_{Y,t}^{\rho_Y} + \omega_{L^s}^{\frac{1}{\sigma_Y}} (L_{Y,t}^s)^{\rho_Y} + \omega_{L^n}^{\frac{1}{\sigma_Y}} (L_{Y,t}^n)^{\rho_Y} + \omega_E^{\frac{1}{\sigma_Y}} E_{Y,t}^{\rho_Y} + \omega_A^{\frac{1}{\sigma_Y}} (A_{Y,t})^{\rho_Y} \right]^{\frac{1}{\rho_Y}} \tag{4.2}$$

where $\omega_K + \omega_{L^s} + \omega_{L^n} + \omega_E + \omega_A = 1$; Z_t denotes total factor productivity in the final-goods sector and is assumed to grow exogenously at the rate g_z; and ρ_Y governs the degree of input substitutability.

Profits for variety i in period t are then given by:

$$\Pi_{Y,t}(i) = p_t(i)Y_t(i) \tag{4.3}$$
$$- w_t^s L_{Y,t}^s(i) - w_t^n L_{Y,t}^n(i) \tag{4.4}$$
$$- (r_t + \delta)K_{Y,t}(i) - p_{E,t}E_{Y,t}(i) - p_{A,t}A_{Y,t}(i) - f \tag{4.5}$$

We can now turn to the parameterization and numerical simulations.

4.2 Simulation

The simulations start from a low-AI initial condition; from that point on, all prices, quantities, and adoption dynamics evolve endogenously according to the model's equilibrium conditions.

We begin with the special case of the CES formulation, Cobb–Douglas case with $\sigma_A = \sigma_Y = 1$, and we simulate the special case to examine its dynamics. Then, we depart from unit elasticity to allow us to examine how input substitutability affects the AI production, pricing, and income distribution.

4.2.1 Parameterization

The parameter values for all the simulations are chosen to be broadly consistent with standard macroeconomic practice and with the stylized facts of the AI sector. Preferences are set to yield plausible consumption–savings behavior; production parameters are drawn from factor income shares; and elasticities are varied over a range to capture cases of complementarity and substitutability. Key parameter values are given in Table 1. Full parameter values are reported in Appendix B.

The distribution parameters are chosen to reflect suggestive evidence regarding the shares of specialized labor. The share of specialized labor (with the kind of qualifications at the major AI labs in the U.S. and elsewhere) in

Table 1: Key Parameters Used in Simulations

Symbol	Description	Values
ϕ	Learning	0.150
α_A	Recursive AI distribution parameter	0.425
α_K	Capital distribution parameter in AI production	0.425
σ_A	CES elasticity, AI production	0.900, 1.000, 1.100
ω_A	AI distribution parameter in final-goods production	0.025
ω_K	Capital distribution parameter in final-goods production	0.350
σ_Y	CES elasticity, final-goods production	0.100–2.000
β	Discount factor	0.980
σ_c	CRRA coefficient (risk aversion)	2.000
ε	Elasticity of substitution across varieties	6.000

the workforce appears to be only a few percent; and workers with specialized AI development knowledge currently earn roughly five times what educated non-specialist workers earn, and about ten times that of nonspecialists overall.[13]

4.2.2 Dynamic Simulation Results

We begin the simulations with the Cobb-Douglas specification, $\sigma_A = \sigma_Y = 1$. These are shown in figures 1-4, at the end of the paper.

Figure 1 illustrates the results for the AI production. As shown in the top row of Figure 1, AI production accelerates early then tapers off, and the use of energy and specialized labor spikes before reverting toward its initial starting level. In the second row, we can see that the price of AI drops and the recursive use of AI rises early before tapering off as production tapers off.

Figure 2 gives the corresponding results for final-goods production. As shown in the top row, final production rises more gradually than AI production, and—since total energy and specialized labor are fixed–their use in final-goods production counterbalances their use in AI production: first falling sharply before recovering substantially. As shown in the second row, unspecialized labor use is constant (since it is in fixed supply and not used in AI production), but the use of AI rises rapidly–with the production of AI seen in Figure 1.

Figure 3 illustrates how prices behave. As shown in the top row, real wages for both specialized and nonspecialized labor rise; and the relative specialized-to-nonspecialized labor spikes before returning toward its original level. The second row shows a gradual, but permanent rise in the price of energy. We also see a temporary rise in the interest rate that corresponds to the buildup of capital in both sectors. Profit, like wages also rise.

[13]See, for example, OECD (2023); U.S. Bureau of Labor Statistics (2024); Insider (2025); H1B Data (2025); PYMNTS (2025).

Figures 5-8 depict the results when inputs are complements in AI production ($\sigma_A = 0.9$), with a range of values for σ_Y. As shown in the top row of figure 5, AI production initially rises rapidly, then tapers off; however it rises more when inputs are highly substitutable in final-goods production. The use of energy and specialized labor follow a corresponding pattern. As shown in the second row (and as in the Cobb-Douglas case), the price of AI drops initially, then levels off, and AI's recursive use rises with AI production.

Figure 6 gives the results for final-goods production when inputs are complements in AI production. While final-goods production rises more when its inputs are highly substitutable, the use of energy and specialized labor again counterbalance their use in AI: they decline before tapering off— all the more so as σ_Y rises.

Figures 7 and 8 gives the results for prices. As shown in the top row, real wages rise, as expected; but the specialized-to-nonspecialized wage premium rises with the degree of input substitutability in final goods. The second row, shows that energy prices, the interest rate, and AI profits, like the wage premium, also rise with final goods input substitutability. The bottom row shows that the AI markup and the corresponding price move inversely with the final goods' input substitutability.

Figures 9 and 12 give the results when inputs in AI production are substitutes, with figure 9 giving the results for AI production. As shown in the top row and in the first panel and last panels in the bottom row, all aspects of AI production are higher when final goods inputs are substitutable, while the price–shown in the middle panel of the bottom row–falls correspondingly more.

Final goods-production, shown in Figure 10, rises most when its inputs are substitutable. Its use of AI rises correspondingly; and, again, the use of energy and specialized labor counterbalance their use in AI.

Figures 11 and 12 give the pricing results. As in the case of AI input complements, real wages–shown in the top row, both rise. And, while the specialized-to-nonspecialized wage premium rises when σ_Y is greater than one, it falls when the elasticity equals one or less, and it is lowest when the inputs in final-goods production are complements. As shown in the second row, profits follow the same pattern as specialized labor: notably, profits fall when final goods inputs are complements. The price of AI, shown in the last row, falls in all cases, but it falls most when final good inputs are substitutes.

Across cases, higher substitutability in final-goods production strengthens pass-through, lowers the AI markup and price, expands quantity, and raises AI profits. And, when final-goods inputs are complements, the opposite holds—markups are higher but quantity and external sales respond weakly, so AI profits are lower. When AI inputs are more substitutable, these patterns are amplified but not reversed. The specialized–to–nonspecialized wage premium

rises when $\sigma_Y > 1$ and is lowest when $\sigma_Y < 1$.

With complements in final-goods production, adoption is learning-first and capital-light: the AI producer still lowers price to scale and learn, but a larger share of output is allocated to self-use, so external sales and profits grow more slowly.

5 Conclusion

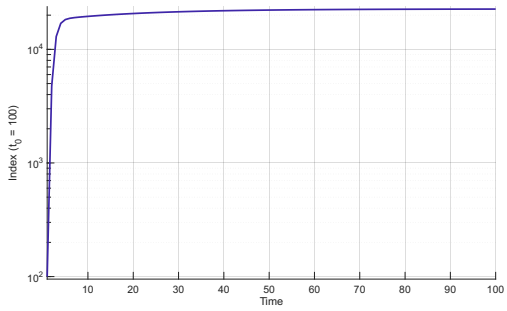
This paper establishes three results. First, recursive self-use of AI, together with learning-by-doing, reshapes monopoly pricing and scale: early price cuts raise future productivity, and self-use amplifies the returns to learning. Second, under Cobb–Douglas we characterize a generalized balanced growth path in closed form. The characterization pins down the conditions for stable, non-explosive growth, yields long-run growth rates as functions of primitives, and provides a long-run analytical benchmark.

Finally, the CES simulations connect adoption and distribution to elasticities: when final-goods inputs are complements, AI adoption is learning-first and capital-light before scaling; when final-goods inputs are substitutes, adoption is scale-first and the two-phase pattern is attenuated. On the distribution side, the specialized–to–nonspecialized wage premium is lowest when final-goods inputs are complements and rises when final-goods inputs are substitutes. Greater substitutability in AI production amplifies these patterns without changing their sign.

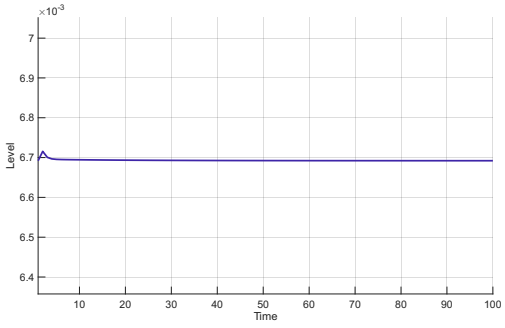
The model also provides a measurement agenda: it maps observed AI prices and cost shares into an implied effective demand elasticity and separates input substitution from the general-equilibrium scale effect. These objects provide targets for future empirical work using cost shares, sectoral adoption measures, and wage premia.

Figure 1: AI Sector: $\sigma_A = 1$ and $\sigma_Y = 1$

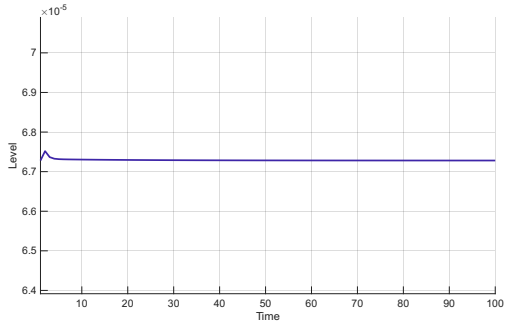
AI Production



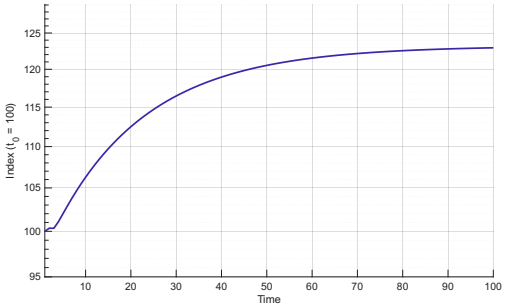
AI Energy use



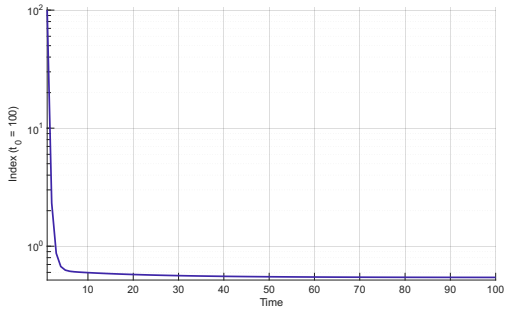
L^S



AI Capital



AI Price



AI in AI

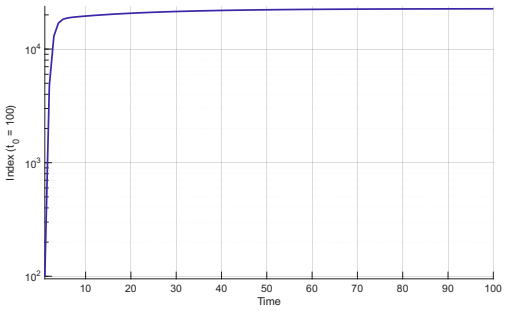
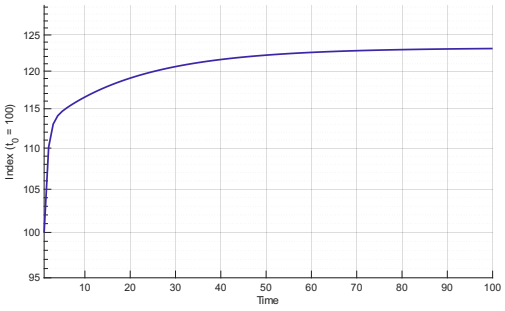
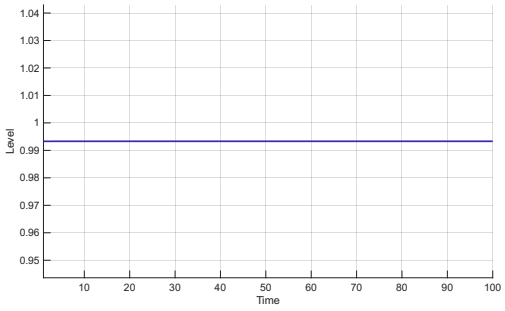


Figure 2: Final goods: $\sigma_A = 1$ and $\sigma_Y = 1$

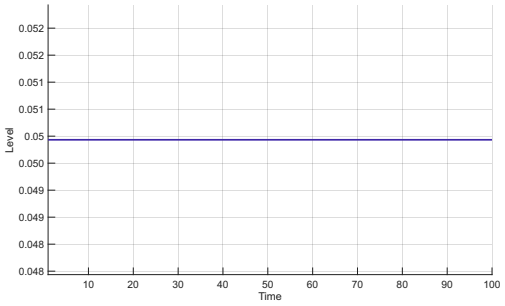
Final Production



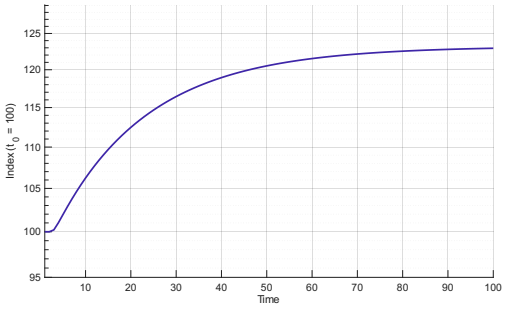
Energy use



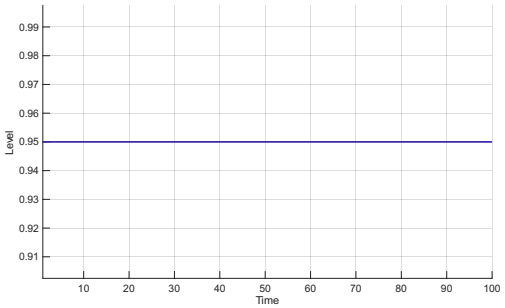
L^S



Capital



L^N



AI in final

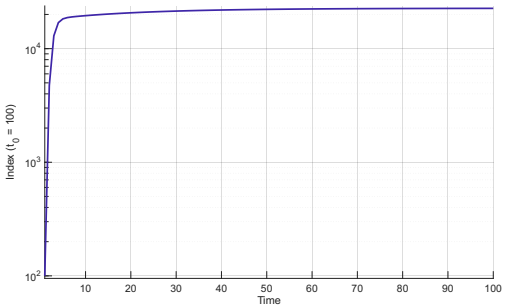
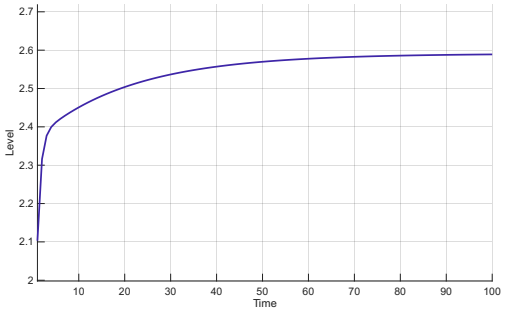
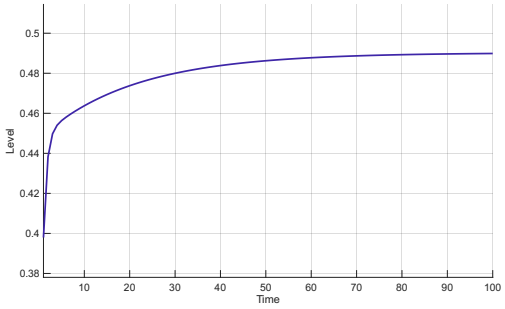


Figure 3: Relative prices: $\sigma_A = 1$ and $\sigma_Y = 1$

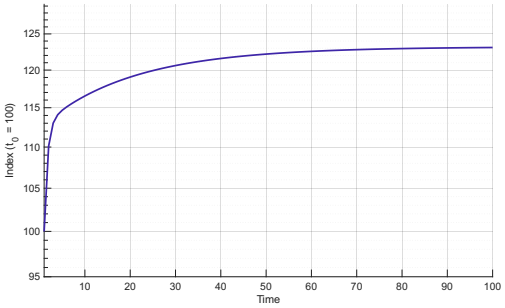
real wage L^S



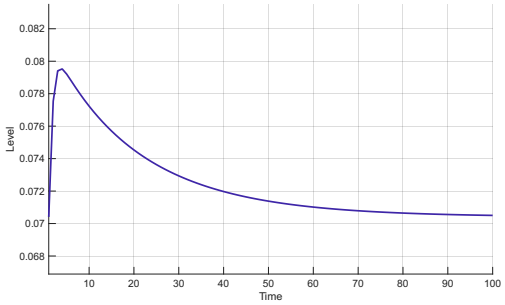
real wage L^N



price for energy



interest rate



wage premium

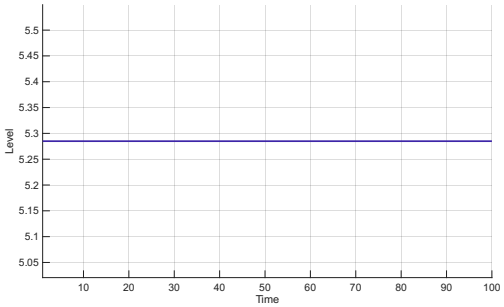
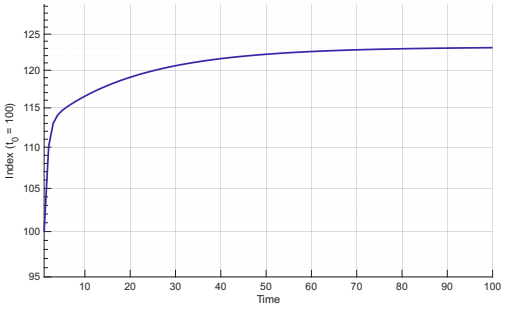
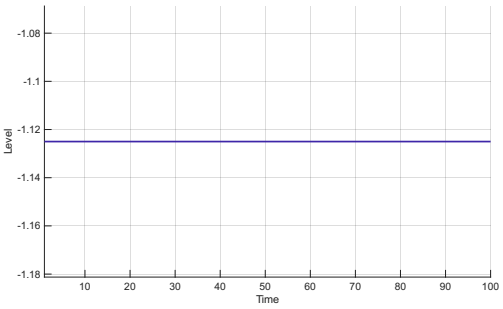


Figure 4: AI profitability: $\sigma_A = 1$ and $\sigma_Y = 1$

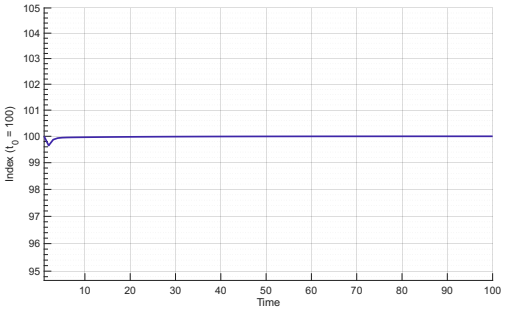
profits AI



AI price elasticity



AI mark-up



Price of AI

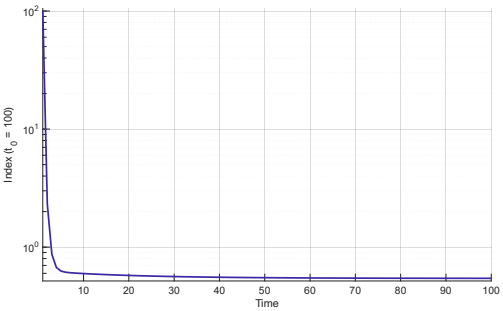
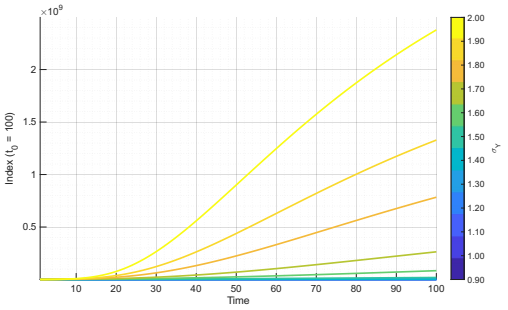
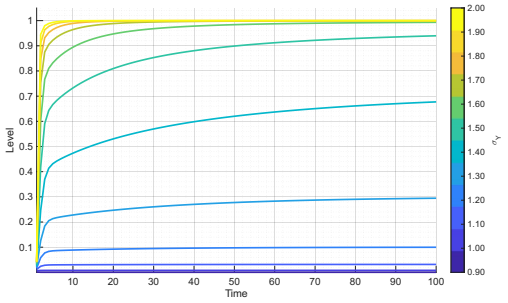


Figure 5: AI Sector: $\sigma_A = 0.9$ and $\sigma_Y \in [0.9, 2]$

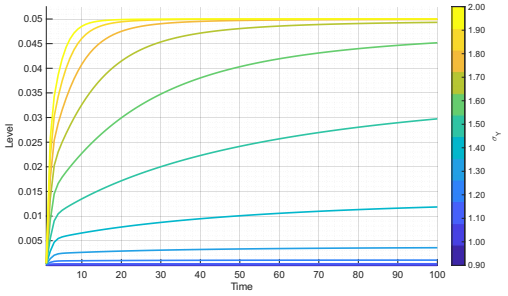
AI Production



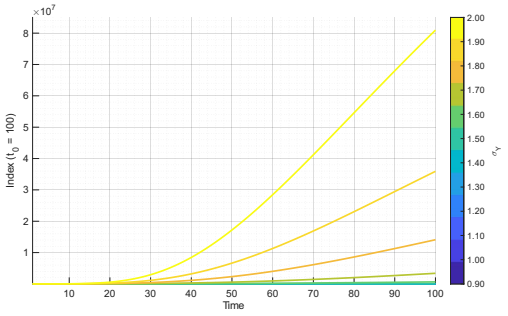
AI Energy use



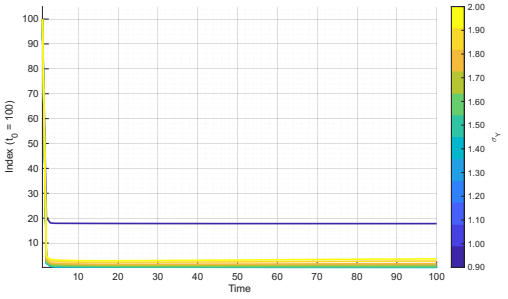
L^S



AI Capital



AI Price



AI in AI

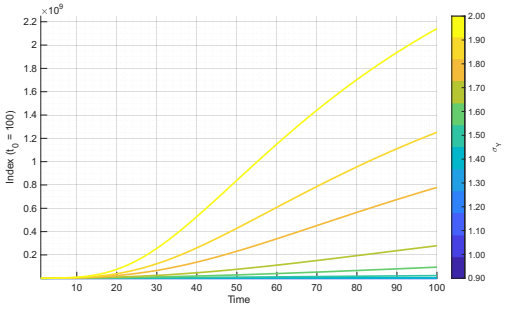
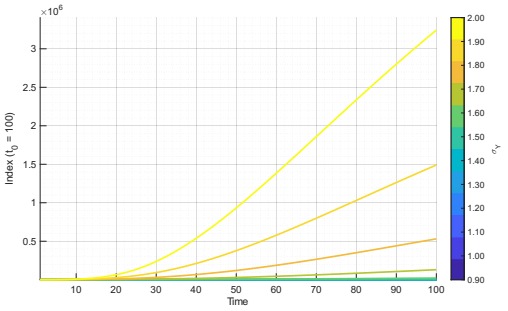
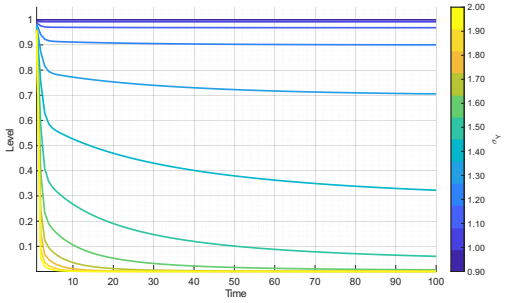


Figure 6: Final goods: $\sigma_A = 0.9$ and $\sigma_Y \in [0.9, 2]$

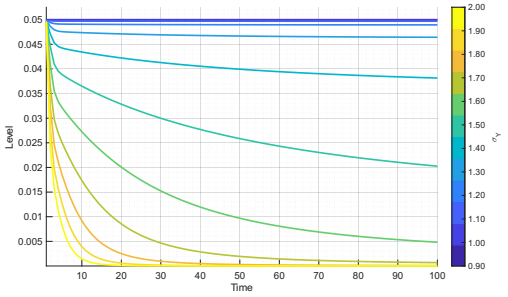
Final Production



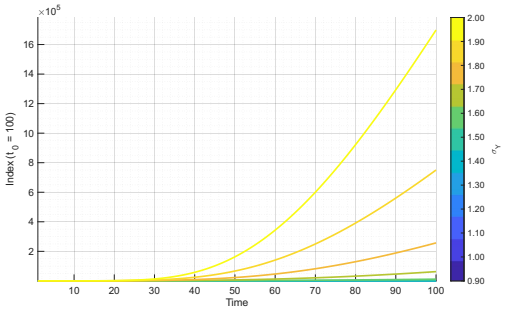
Energy use



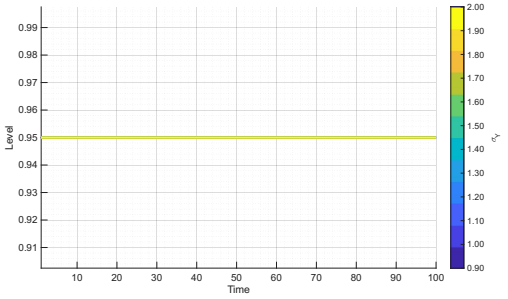
L^S



Capital



L^N



AI in final

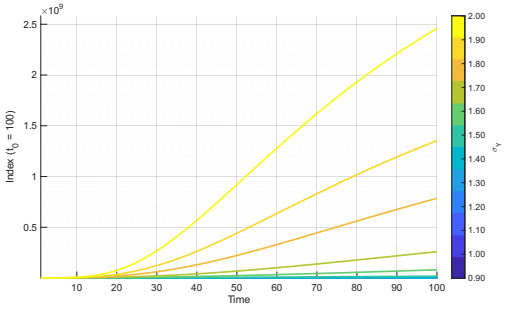
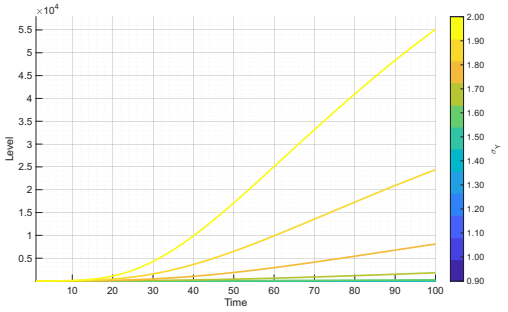
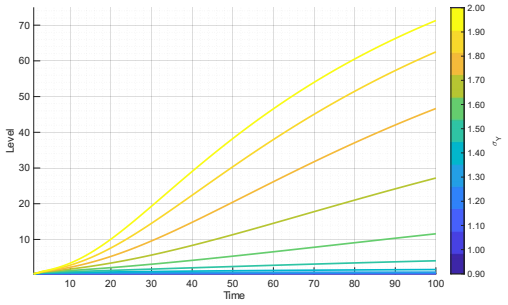


Figure 7: Relative prices: $\sigma_A = 0.9$ and $\sigma_Y \in [0.9, 2]$

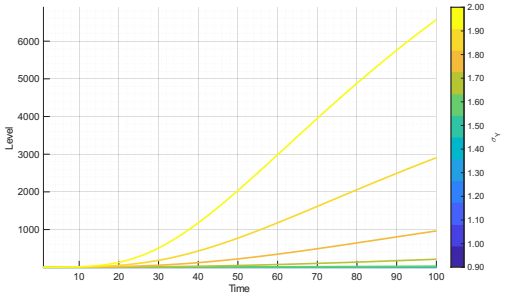
real wage L^S



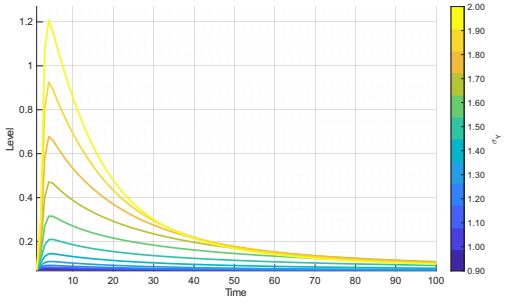
real wage L^N



price for energy



interest rate



wage premium

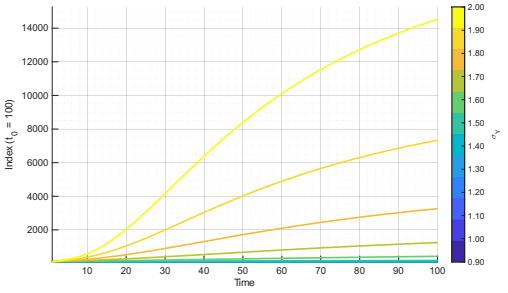
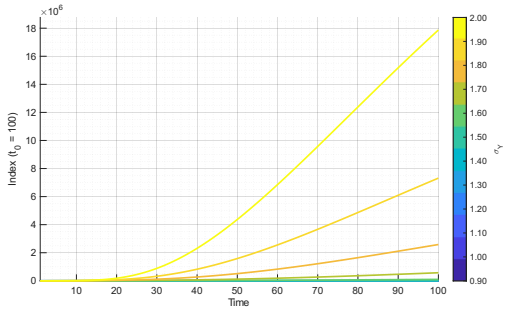
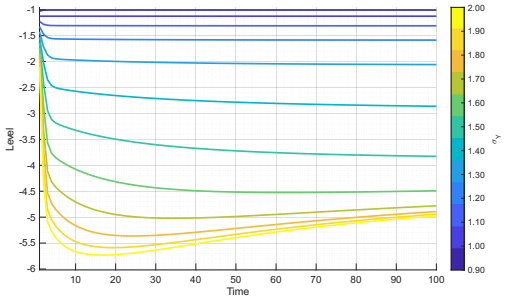


Figure 8: AI profitability: $\sigma_A = 0.9$ and$\sigma_Y \in [0.9, 2]$

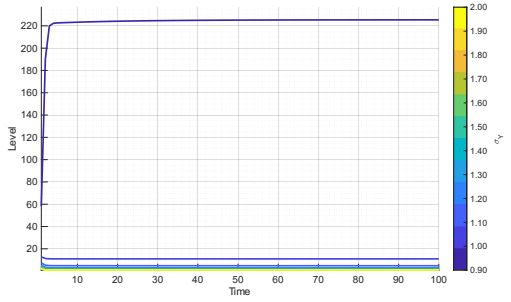
profits AI



AI price elasticity



AI mark-up



Price of AI

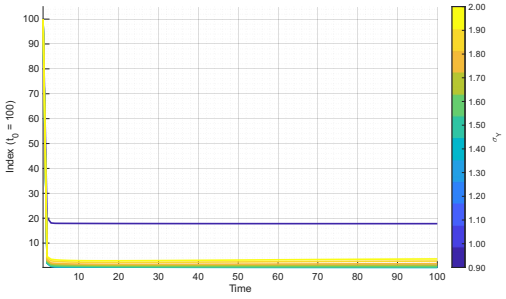
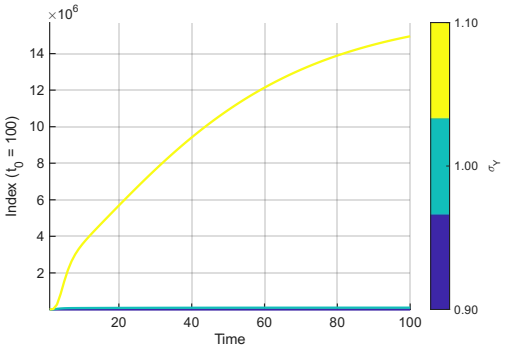
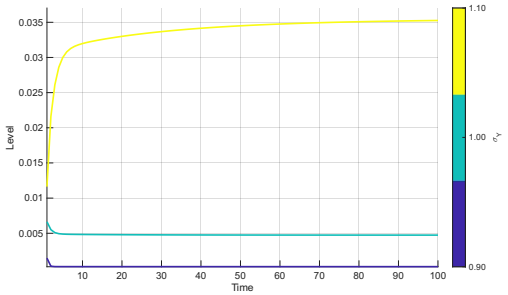


Figure 9: AI Sector: $\sigma_A = 1.1$ and $\sigma_Y \in [0.9, 1.0, 1.1]$

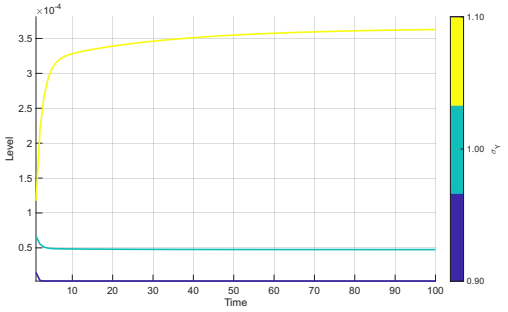
AI Production



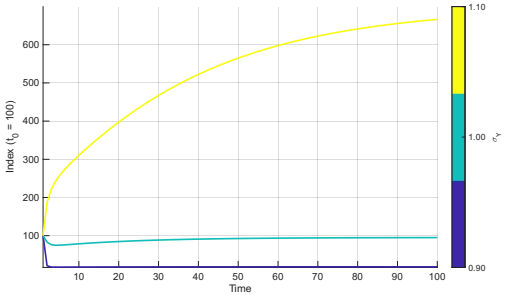
AI Energy use



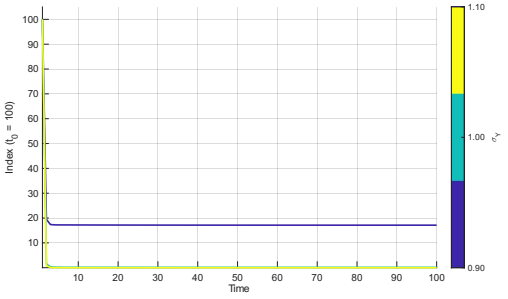
L^S



AI Capital



AI Price



AI in AI

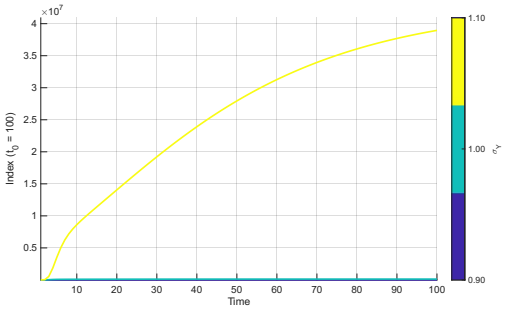


Figure 10: Final goods: $\sigma_A = 1.1$ and $\sigma_Y \in [0.9, 1.0, 1.1]$

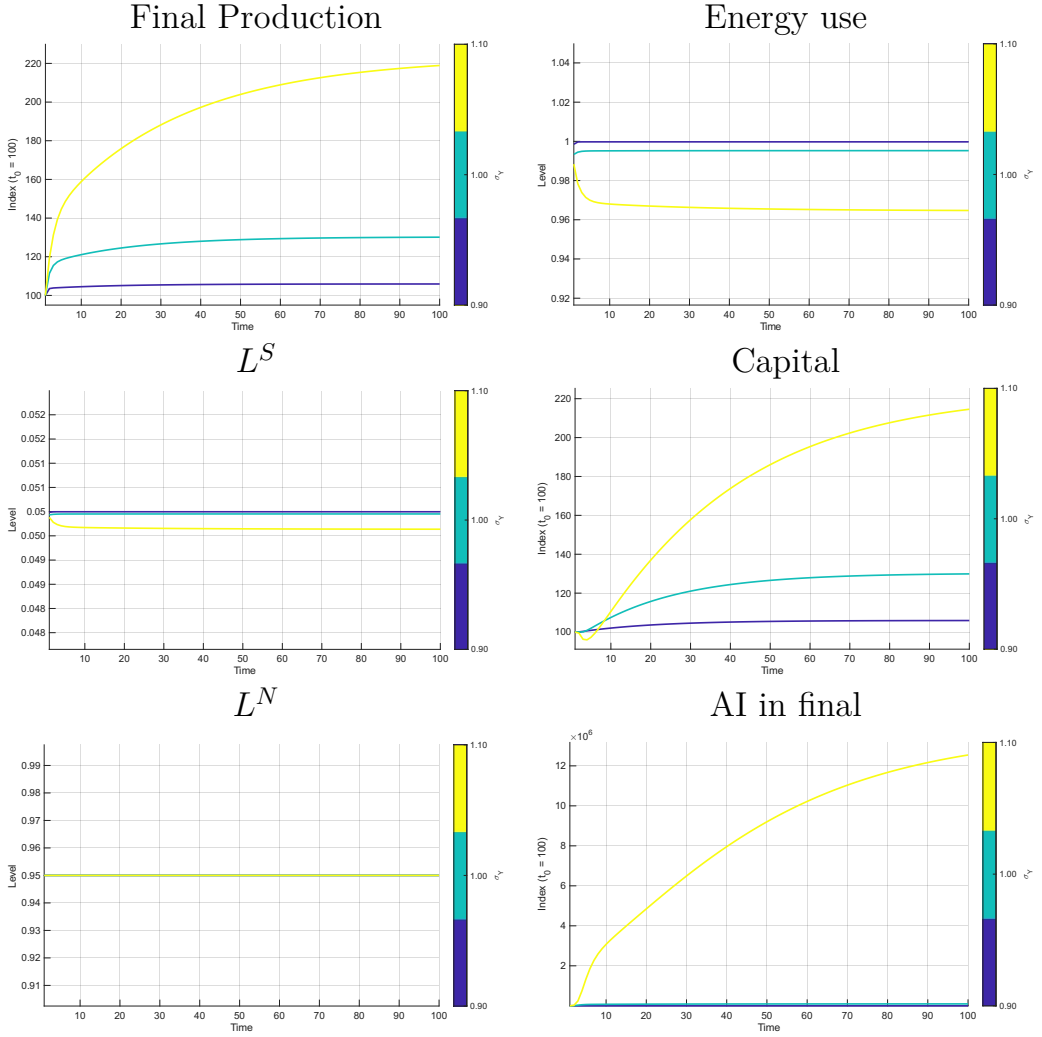
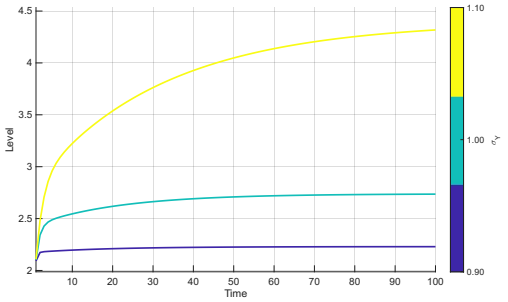
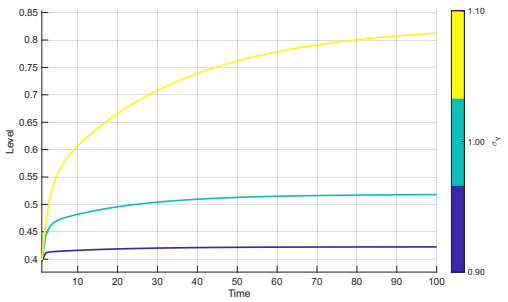


Figure 11: Relative prices: $\sigma_A = 1.1$ and $\sigma_Y \in [0.9, 1.0, 1.1]$

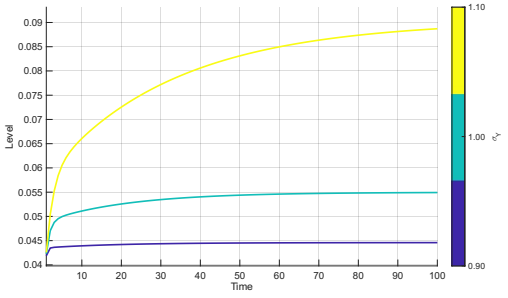
real wage L^S



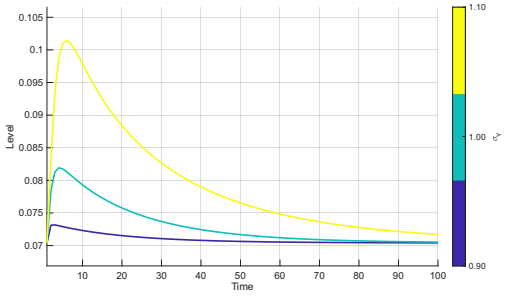
real wage L^N



price for energy



interest rate



wage premium

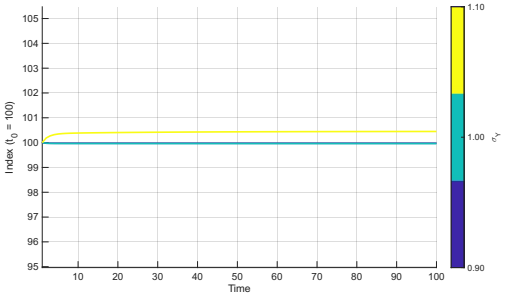
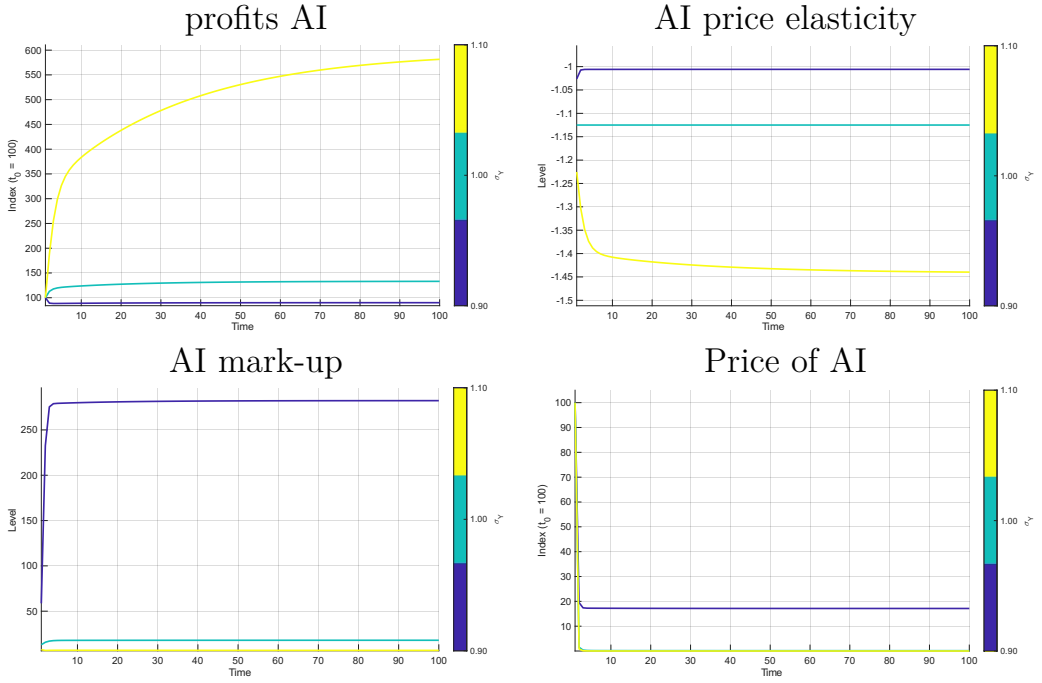


Figure 12: AI profitability: $\sigma_A = 1.1$ and $\sigma_Y \in [0.9, 1.0, 1.1]$



Appendices

A Effective Elasticity of AI Demand

Setting $Z_t = 1$, each final-goods variety i is produced with a constant returns to scale CES technology with elasticity of substitution σ_Y, and the unit cost index is:

$$\mathcal{U}_t \equiv \left(\sum_{j \in \mathcal{J}} \omega_j \, p_{j,t}^{1-\sigma_Y} \right)^{\frac{1}{1-\sigma_Y}}. \tag{A.1}$$

Cost minimization gives CES conditional demands. Applying this to $j = A$ and aggregating with symmetry gives:

$$A_{Y,t} = \omega_A \left(\frac{p_{A,t}}{\mathcal{U}_t} \right)^{-\sigma_Y} Y_t. \tag{A.2}$$

Taking logs of (A.2), defining the (time-varying) scale effect as $\eta_t \equiv -\frac{d \ln Y_t}{d \ln p_{A,t}}$, and totally differentiating with respect to $\ln p_{A,t}$ yields:

$$\frac{d \ln A_{Y,t}}{d \ln p_{A,t}} = -\sigma_Y + \sigma_Y \frac{d \ln \mathcal{U}_t}{d \ln p_{A,t}} - \eta_t. \tag{A.3}$$

Under monopolistic competition in final goods, $p_t(i) = \mu_Y MC_t(i)$; from symmetry, $p_t(i) = P_t$, and from the normalization $P_t = 1$; so, $MC_t(i) = 1/\mu_Y$.

From our constant returns to scale CES technology (and $Z_t = 1$), $MC_t(i) = \mathcal{U}_t$. Thus, $\mathcal{U}_t = 1/\mu_Y$ and $\frac{d \ln \mathcal{U}_t}{d \ln p_{A,t}} = 0$ along the symmetric equilibrium.[14]

So, defining the (positive) effective demand elasticity as $\epsilon_t \equiv -\frac{d \ln A_{Y,t}}{d \ln p_{A,t}}$, (A.3) becomes:

$$\epsilon_t = \sigma_Y + \eta_t. \tag{A.4}$$

This decomposition provides the key link between substitution in final-goods production and the AI producer's pricing and adoption dynamics. In this model, the effective elasticity faced by the AI producer is the simple sum of input substitution in final-goods production and the general-equilibrium scale effect.

[14] In partial equilibrium (holding other input prices fixed), $\frac{\partial \ln \mathcal{U}_t}{\partial \ln p_{A,t}} > 0$; however, here we use the total (general-equilibrium) derivative $\frac{d \ln \mathcal{U}_t}{d \ln p_{A,t}}$ along the symmetric equilibrium.

B Parameter Values

Table 2: Parameter Values for Simulations

Symbol	Description	Value
Household / Intertemporal		
β	Discount factor	0.980
σ_c	Coefficient of relative risk aversion	2.000
ε	Elasticity of substitution across varieties	6.000
Capital		
δ	Depreciation rate	0.050
Technology: AI production (A_t)		
ϕ	Learning parameter	0.150
α_A	Recursive AI distribution parameter	0.425
α_K	Capital distribution parameter	0.425
α_{L^s}	Specialized labor distribution parameter	0.050
α_E	Energy distribution parameter	0.100
σ_A	Elasticity of substitution	0.900, 1.000, 1.100
Technology: final-goods production (Y_t)		
ω_K	Capital distribution parameter	0.350
ω_A	AI distribution parameter	0.025
ω_{L^s}	Specialized labor distribution parameter	0.125
ω_{L^n}	Nonspecialized labor distribution parameter	0.450
ω_E	Energy distribution parameter	0.050
σ_Y	Elasticity of substitution	0.100-2.000
Endowments / Initial conditions		
$\bar{L}^s$	Specialized labor supply	0.050
$\bar{L}^n$	Nonspecialized labor supply	0.950
$\bar{E}$	Energy endowment	1.000
$A_{Y,0}$	Initial A_Y	1.187
$A_{A,0}$	Initial A_A	0.001

References

Daron Acemoglu and Pascual Restrepo. The race between man and machine: Implications of technology for growth, factor shares, and employment. *American Economic Review*, 108(6):1488–1542, 2018.

Daron Acemoglu and Pascual Restrepo. Automation and new tasks: How technology displaces and reinstates labor. *Journal of Political Economy*, 128(6): 2188–2244, 2020.

Philippe Aghion, Benjamin F. Jones, and Charles I. Jones. Artificial intelligence and economic growth. *Brookings Papers on Economic Activity*, 48(2 (Fall)): 1–45, 2017.

Seth G Benzell and Erik Brynjolfsson. Digital abundance and scarce genius: Implications for wages, interest rates, and growth. Technical report, National Bureau of Economic Research, 2019.

Erik Brynjolfsson and Tom Mitchell. What can machine learning do? workforce implications. *Science*, 358(6370):1530–1534, 2017. doi: 10.1126/science.aap8062.

Erik Brynjolfsson, Bharat Chandar, and Ruyu Chen. Canaries in the coal mine? six facts about the recent employment effects of artificial intelligence. Technical report, Stanford Digital Economy Lab, Stanford University, August 2025.

Iain M Cockburn, Rebecca Henderson, and Scott Stern. 4. the impact of artificial intelligence on innovation: An exploratory analysis. In *The economics of artificial intelligence*, pages 115–148. University of Chicago Press, 2019.

Avinash K Dixit and Joseph E Stiglitz. Monopolistic competition and optimum product diversity. *The American economic review*, 67(3):297–308, 1977.

H1B Data. Anthropic pbc: Research engineer salaries in san francisco, 2025. URL https://h1bdata.info/index.php?city=SAN+FRANCISCO&em=ANTHROPIC+PBC&job=RESEARCH+ENGINEER. Accessed September 5, 2025.

Business Insider. Top ai startup companies' salaries: Openai, anthropic, perplexity, July 2025. URL https://www.businessinsider.com/top-ai-startup-companies-salaries-pay-data-openai-anthropic-perplexity-2025-7.

Charles I. Jones and Christopher Tonetti. Past automation and future a.i.: How weak links tame the growth explosion. January 2026. Version 0.3, very preliminary.

Anton Korinek and Joseph E. Stiglitz. Artificial intelligence and its implications for income distribution and unemployment. In Ajay Agrawal, Joshua Gans, and Avi Goldfarb, editors, *The Economics of Artificial Intelligence: An Agenda*, pages 349–390. University of Chicago Press, 2019.

William D Nordhaus. Are we approaching an economic singularity? information technology and the future of economic growth. *American Economic Journal: Macroeconomics*, 13(1):299–332, 2021.

OECD. The supply, demand and characteristics of the ai workforce across oecd countries. Technical report, OECD Publishing, 2023. URL https://www.oecd.org/content/dam/oecd/en/publications/reports/2023/02/the-supply-demand-and-characteristics-of-the-ai-workforce-across-oecd-countries_fcfe53fb/bb17314a-en.pdf.

PYMNTS. Ai startups: What openai, anthropic pay their technical staff, July 2025. URL https://www.pymnts.com/artificial-intelligence-2/2025/ai-startups-what-openai-anthropic-pay-their-technical-staff.

U.S. Bureau of Labor Statistics. Median weekly earnings of full-time workers with only a bachelor's degree: $1541 in q2 2024. The Economics Daily, July 2024. URL https://www.bls.gov/opub/ted/2024/median-weekly-earnings-of-full-time-workers-with-only-a-bachelors-degree-1541-in-q2-2024.htm.

Laura Veldkamp and Jinhee Chung. Data as a factor of production. *Journal of Political Economy*, 2024. Forthcoming.