Is Automating AI Research Enough for a Growth Explosion?

Tom Davidson [1] **Basil Halperin** [2] **Thomas Houlden** [3] **Anton Korinek** [4]

[1]Forethought, [2]Stanford, [3]London School of Economics, [4]University of Virginia

April 2025

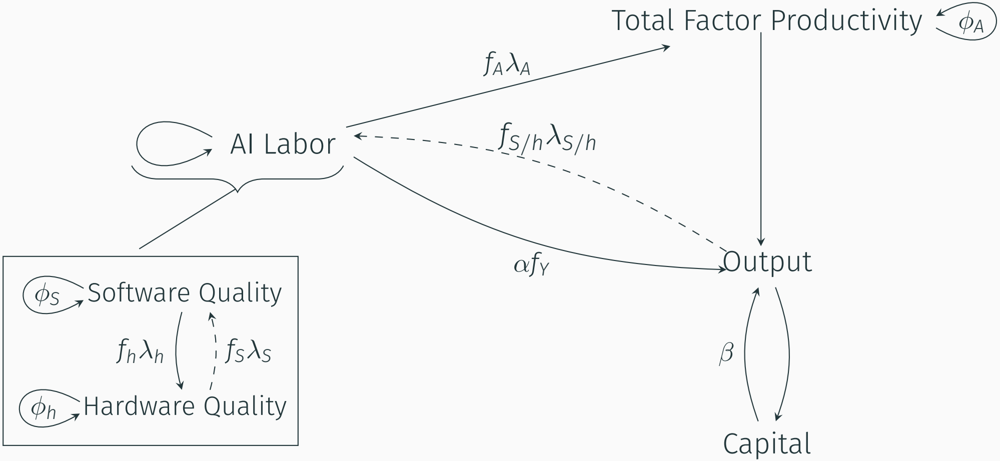
Motivation

> "Perhaps some areas, like robotics, might take longer to figure out by default. And the societal rollout, e.g. in medical or legal professions, could easily be slowed by societal choices or regulation. But **once models can automate AI research itself, that's enough—enough to kick off intense feedback loops—and we could very quickly make further progress, the automated AI engineers themselves solving all the remaining bottlenecks to fully automating everything.** In particular, millions of automated researchers could very plausibly compress a decade of further algorithmic progress into a year or less."

Situational Awareness, Aschenbrenner (2024)

The software-hardware model of AI

The software-hardware model of AI



Roadmap ▸ Lit review

1. Building blocks of the model

2. The software-hardware model

3. Scope of claims

Building blocks of the model

Building blocks of the model

The software-hardware model

Scope of claims

The intelligence explosion

> **“Let an ultraintelligent machine be defined as a machine that can far surpass all the intellectual activities of any man.**
>
> — I.J. Good (1965)

The intelligence explosion

> “**Let an ultraintelligent machine be defined as a machine that can far surpass all the intellectual activities of any man.** Since the design of machines is one of these intellectual activities, an ultraintelligent machine could design even better machines;
>
> — I.J. Good (1965)

The intelligence explosion

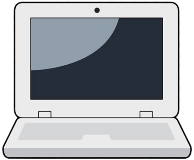
> "**Let an ultraintelligent machine be defined as a machine that can far surpass all the intellectual activities of any man.** Since the design of machines is one of these intellectual activities, an ultraintelligent machine could design even better machines; **there would then unquestionably be an 'intelligence explosion'**, and the intelligence of man would be left far behind."
>
> — I.J. Good (1965)

The intelligence explosion?

"**Let an ultraintelligent machine be defined as a machine that can far surpass all the intellectual activities of any man.** Since the design of machines is one of these intellectual activities, an ultraintelligent machine could design even better machines; **there would then unquestionably be an 'intelligence explosion'**, and the intelligence of man would be left far behind."

— I.J. Good (1965)

$\dot{A}_t$

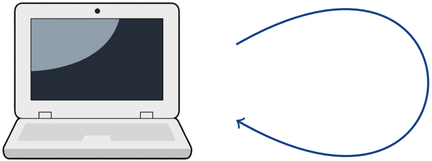


The intelligence explosion?

"**Let an ultraintelligent machine be defined as a machine that can far surpass all the intellectual activities of any man.** Since the design of machines is one of these intellectual activities, an ultraintelligent machine could design even better machines; **there would then unquestionably be an 'intelligence explosion'**, and the intelligence of man would be left far behind."

— I.J. Good (1965)

$$\dot{A}_t = A_t$$

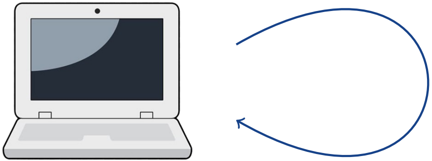


The intelligence explosion?

"**Let an ultraintelligent machine be defined as a machine that can far surpass all the intellectual activities of any man.** Since the design of machines is one of these intellectual activities, an ultraintelligent machine could design even better machines; **there would then unquestionably be an 'intelligence explosion'**, and the intelligence of man would be left far behind."

— I.J. Good (1965)

$$\dot{A}_t = A_t^{1+\phi}$$

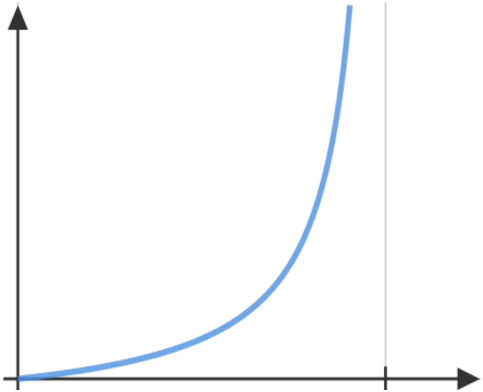


The intelligence explosion?

> “**Let an ultraintelligent machine be defined as a machine that can far surpass all the intellectual activities of any man.** Since the design of machines is one of these intellectual activities, an ultraintelligent machine could design even better machines; **there would then unquestionably be an ‘intelligence explosion’**, and the intelligence of man would be left far behind.”
>
> — I.J. Good (1965)

$$\dot{A}_t = A_t^{1+\phi}$$

$$\phi > 0$$

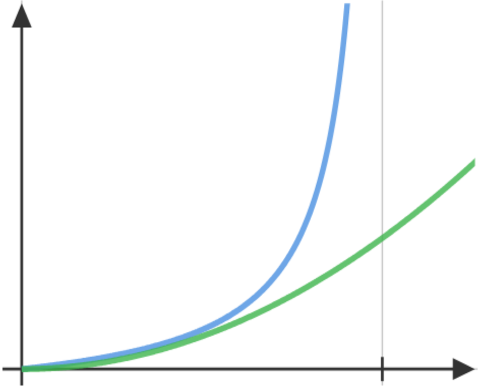


The intelligence explosion?

> “**Let an ultraintelligent machine be defined as a machine that can far surpass all the intellectual activities of any man.** Since the design of machines is one of these intellectual activities, an ultraintelligent machine could design even better machines; **there would then unquestionably be an ‘intelligence explosion’**, and the intelligence of man would be left far behind.”
>
> — I.J. Good (1965)

$$\dot{A}_t = A_t^{1+\phi}$$

$$\phi \in (-1, 0)$$



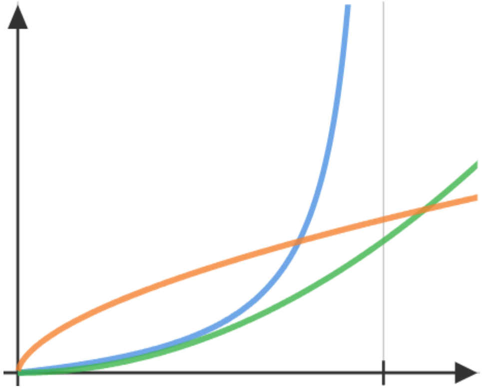
The intelligence explosion?

"**Let an ultraintelligent machine be defined as a machine that can far surpass all the intellectual activities of any man.** Since the design of machines is one of these intellectual activities, an ultraintelligent machine could design even better machines; **there would then unquestionably be an 'intelligence explosion'**, and the intelligence of man would be left far behind."

— I.J. Good (1965)

$$\dot{A}_t = A_t^{1+\phi}$$

$$\phi < -1$$

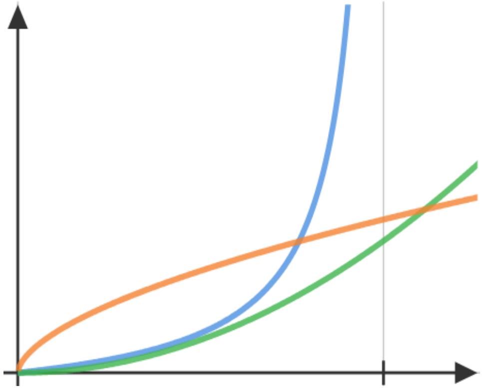


The intelligence explosion? The role of diminishing returns

> “**Let an ultraintelligent machine be defined as a machine that can far surpass all the intellectual activities of any man.** Since the design of machines is one of these intellectual activities, an ultraintelligent machine could design even better machines; **there would then unquestionably be an ‘intelligence explosion’**, and the intelligence of man would be left far behind.”
>
> — I.J. Good (1965)

$$\dot{A}_t = A_t^{1+\phi}$$

$$\phi < -1$$



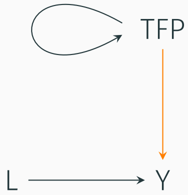
Will an *intelligence* explosion cause an *economic* explosion?

$$\dot{A} = A_t^{1+\phi}$$

Will an *intelligence* explosion cause an *economic* explosion?

$$\dot{A} = A_t^{1+\phi}$$

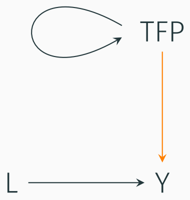
$$Y_t = A_t^{\gamma} L_t^{\alpha}$$



Will an *intelligence* explosion cause an *economic* explosion?

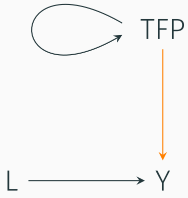
$$\dot{A} = A_t^{1+\phi}$$
$$Y_t = A_t^{\gamma} L_t^{\alpha}$$

If $\gamma > 0$, any form of intelligence explosion causes the same form of economic explosion



Will an *intelligence* explosion cause an *economic* explosion?

$$\dot{A} = A_t^{1+\phi}$$
$$Y_t = A_t L_t^{\alpha}$$



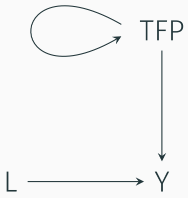
If $\gamma > 0$, any form of intelligence explosion causes the same form of economic explosion

- **Economic singularity** condition:

$$\phi > 0$$

Other feedback loops matter:

$$\dot{A}_t = A_t^{1+\phi}$$
$$Y_t = A_t L_t^{\alpha} K_t^{\beta}$$

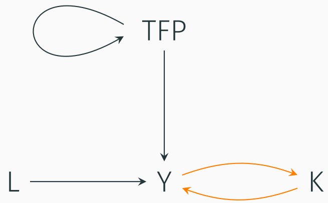


Other feedback loops matter:

$$\dot{A}_t = A_t^{1+\phi}$$

$$Y_t = A_t L_t^{\alpha} K_t^{\beta}$$

$$\dot{K}_t = s_K Y_t - \delta K_t$$



Other feedback loops matter:

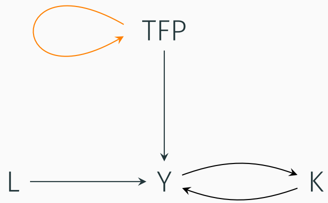
$$\dot{A}_t = A_t^{1+\phi}$$

$$Y_t = A_t L_t^{\alpha} K_t^{\beta}$$

$$\dot{K}_t = s_K Y_t - \delta K_t$$

Economic singularity condition:

$$\phi > 0$$



Other feedback loops matter: the role of *accumulable factors*

$$\dot{A}_t = A_t^{1+\phi}$$

$$Y_t = A_t L_t^{\alpha} K_t^{\beta}$$

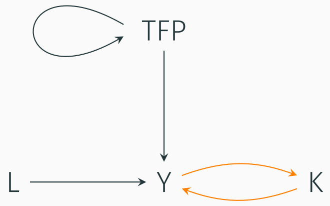
$$\dot{K}_t = s_K Y_t - \delta K_t$$

Economic singularity condition:

$$\phi > 0$$

or

$$\beta > 1$$

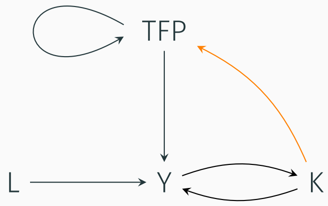


Other feedback loops matter: the role of *accumulable factors*

$$\dot{A}_t = A_t^{1+\phi} (\kappa K_t)^{\lambda}$$

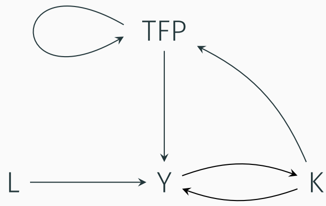
$$Y_t = A_t L_t^{\alpha} \left((1-\kappa) K_t \right)^{\beta}$$

$$\dot{K}_t = s_K Y_t - \delta K_t$$



Other feedback loops matter: the role of *accumulable factors*

$$\dot{A}_t = A_t^{1+\phi}(\kappa K_t)^{\lambda}$$
$$Y_t = A_t L_t^{\alpha}((1-\kappa)K_t)^{\beta}$$
$$\dot{K}_t = s_K Y_t - \delta K_t$$



Focusing on accumulable factors:

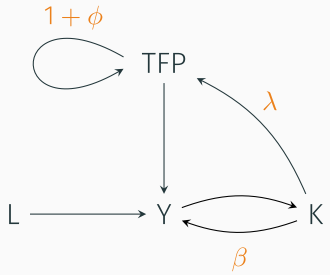
$$\dot{A}_t = \text{stuff} \cdot A_t^{1+\phi} K_t^{\lambda}$$
$$\dot{K}_t = \text{stuff} \cdot A_t K_t^{\beta}$$

Other feedback loops matter: the role of *accumulable factors*

$$\dot{A}_t = A_t^{1+\phi}(\kappa K_t)^{\lambda}$$
$$Y_t = A_t L_t^{\alpha}\left((1-\kappa)K_t\right)^{\beta}$$
$$\dot{K}_t = s_K Y_t - \delta K_t$$

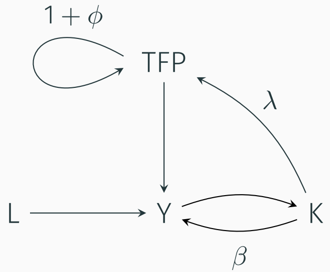
Focusing on accumulable factors:

$$\dot{A}_t = \text{stuff} \cdot A_t^{1+\phi} K_t^{\lambda}$$
$$\dot{K}_t = \text{stuff} \cdot A_t K_t^{\beta}$$



Other feedback loops matter: the role of *accumulable factors*

$$\dot{A}_t = A_t^{1+\phi}(\kappa K_t)^{\lambda}$$
$$Y_t = A_t L_t^{\alpha}\left((1-\kappa)K_t\right)^{\beta}$$
$$\dot{K}_t = s_K Y_t - \delta K_t$$



Focusing on accumulable factors:

$$\dot{A}_t = \text{stuff} \cdot A_t^{1+\phi} K_t^{\lambda}$$
$$\dot{K}_t = \text{stuff} \cdot A_t K_t^{\beta}$$

Proposition (explosive systems).
System explodes in finite time if the exponent matrix, $\begin{bmatrix} 1+\phi & \lambda \\ 1 & \beta \end{bmatrix}$, has an eigenvalue > 1.

Explosion conditions:

$\phi > 0$ or $\beta > 1$

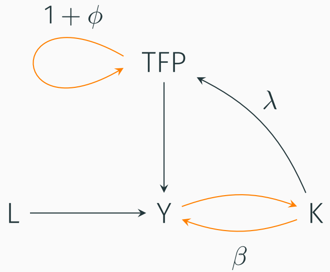
Other feedback loops matter: the role of *accumulable factors*

$$\dot{A}_t = A_t^{1+\phi}(\kappa K_t)^{\lambda}$$
$$Y_t = A_t L_t^{\alpha}\left((1-\kappa)K_t\right)^{\beta}$$
$$\dot{K}_t = s_K Y_t - \delta K_t$$

Focusing on accumulable factors:

$$\dot{A}_t = \text{stuff} \cdot A_t^{1+\phi} K_t^{\lambda}$$
$$\dot{K}_t = \text{stuff} \cdot A_t K_t^{\beta}$$

Proposition (explosive systems).
System explodes in finite time if the exponent matrix, $\begin{bmatrix} 1+\phi & \lambda \\ 1 & \beta \end{bmatrix}$, has an eigenvalue > 1.



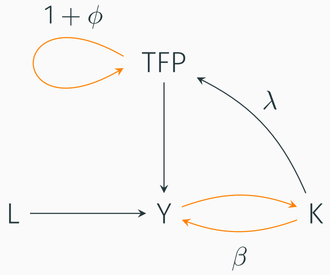
Explosion conditions:

$$\phi > 0 \text{ or } \beta > 1$$

$$\underbrace{(1+\phi)+\beta}_{\text{direct effects}} \qquad > 1$$

Other feedback loops matter: the role of *accumulable factors*

$$\dot{A}_t = A_t^{1+\phi}(\kappa K_t)^\lambda$$
$$Y_t = A_t L_t^\alpha \left((1-\kappa)K_t\right)^\beta$$
$$\dot{K}_t = s_K Y_t - \delta K_t$$



Focusing on accumulable factors:

$$\dot{A}_t = \text{stuff} \cdot A_t^{1+\phi} K_t^\lambda$$
$$\dot{K}_t = \text{stuff} \cdot A_t K_t^\beta$$

Proposition (explosive systems).
System explodes in finite time if the exponent matrix, $\begin{bmatrix} 1+\phi & \lambda \\ 1 & \beta \end{bmatrix}$, has an eigenvalue > 1.

Explosion conditions:

$$\phi > 0 \text{ or } \beta > 1$$
$$\underbrace{(1+\phi)+\beta}_{\text{direct effects}} - (1+\phi)\beta > 1$$

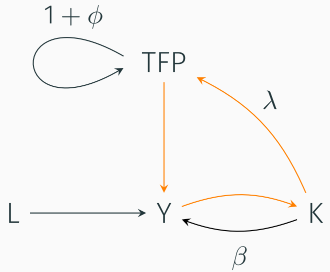
Other feedback loops matter: the role of *accumulable factors*

$$\dot{A}_t = A_t^{1+\phi}(\kappa K_t)^{\lambda}$$
$$Y_t = A_t L_t^{\alpha}\left((1-\kappa)K_t\right)^{\beta}$$
$$\dot{K}_t = s_K Y_t - \delta K_t$$

Focusing on accumulable factors:

$$\dot{A}_t = \text{stuff} \cdot A_t^{1+\phi} K_t^{\lambda}$$
$$\dot{K}_t = \text{stuff} \cdot A_t K_t^{\beta}$$

Proposition (explosive systems).
System explodes in finite time if the exponent matrix, $\begin{bmatrix} 1+\phi & \lambda \\ 1 & \beta \end{bmatrix}$, has an eigenvalue > 1.



Explosion conditions:

$$\phi > 0 \text{ or } \beta > 1$$

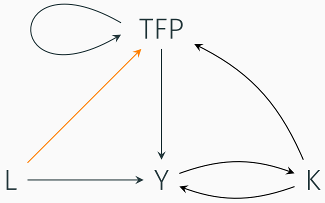
$$\underbrace{(1+\phi)+\beta}_{\text{direct effects}} - (1+\phi)\beta + \underbrace{\lambda \cdot 1}_{\text{indirect effects}} > 1$$

The canonical semi-endogenous growth model

$$\dot{A}_t = A_t^{1+\phi}(\ell L_t)^{\lambda}(\kappa K_t)^{\lambda}$$

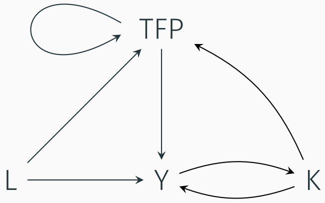
$$Y_t = A_t\left((1-\ell)L_t\right)^{\alpha}\left((1-\kappa)K_t\right)^{\beta}$$

$$\dot{K}_t = s_K Y_t - \delta K_t$$



The canonical semi-endogenous growth model

$$\dot{A}_t = A_t^{1+\phi}(\ell L_t)^\lambda(\kappa K_t)^\lambda$$
$$Y_t = A_t\left((1-\ell)L_t\right)^\alpha\left((1-\kappa)K_t\right)^\beta$$
$$\dot{K}_t = s_K Y_t - \delta K_t$$

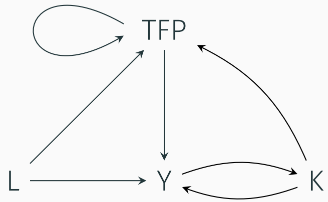


Best guess calibration:

- $\phi = -3.4$ (Bloom et al 2020) — $\phi > 0$ ✗
- $\beta = 0.4$ (capital share in production) — $\beta > 1$ ✗
- $\lambda = 0.1$ (capital share in R&D) — $(1+\phi)+\beta-(1+\phi)\beta+\lambda > 1$ ✗

The canonical semi-endogenous growth model

$$\dot{A}_t = A_t^{1+\phi}(\ell L_t)^\lambda(\kappa K_t)^\lambda$$
$$Y_t = A_t\left((1-\ell)L_t\right)^\alpha\left((1-\kappa)K_t\right)^\beta$$
$$\dot{K}_t = s_K Y_t - \delta K_t$$

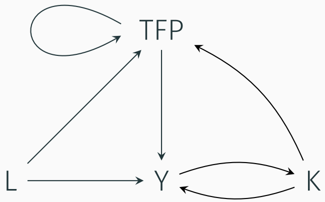


Takeaways:

1. Where are the feedback loops?

The canonical semi-endogenous growth model

$$\dot{A}_t = A_t^{1+\phi}(\ell L_t)^{\lambda}(\kappa K_t)^{\lambda}$$
$$Y_t = A_t\left((1-\ell)L_t\right)^{\alpha}\left((1-\kappa)K_t\right)^{\beta}$$
$$\dot{K}_t = s_K Y_t - \delta K_t$$

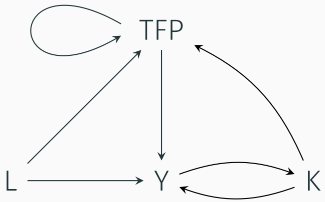


Takeaways:

1. Where are the feedback loops?
2. How strong are the diminishing returns?

The canonical semi-endogenous growth model

$$\dot{A}_t = A_t^{1+\phi}(\ell L_t)^{\lambda}(\kappa K_t)^{\lambda}$$
$$Y_t = A_t\left((1-\ell)L_t\right)^{\alpha}\left((1-\kappa)K_t\right)^{\beta}$$
$$\dot{K}_t = s_K Y_t - \delta K_t$$

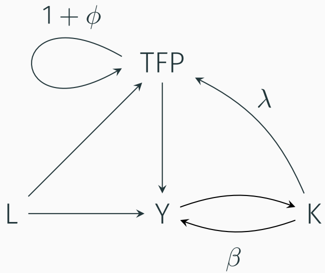


Takeaways:

1. Where are the feedback loops?
2. How strong are the diminishing returns?
3. What are the accumulative factors?

Introducing automation

$$\dot{A}_t = A_t^{1+\phi}(\ell_A L_t)^{\lambda}(\kappa_A K_t)^{\lambda}$$
$$Y_t = A_t\,(\ell_Y L_t)^{\alpha}\,(\kappa_Y K_t)^{\beta}$$
$$\dot{K}_t = s_K Y_t - \delta K_t$$

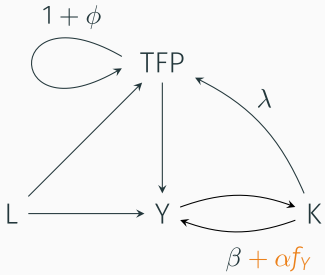


Explosion conditions:

$$\phi > 0$$
$$\beta > 1$$
$$(1+\phi) + \beta - (1+\phi)\beta + \lambda > 1$$

Introducing automation

$$\dot{A}_t = A_t^{1+\phi}(\ell_A L_t)^{\lambda}(\kappa_A K_t)^{\lambda}$$
$$Y_t = A_t\,(\ell_Y L_t)^{\alpha(1-f_Y)}\,(\kappa_Y K_t)^{\beta+f_Y\alpha}$$
$$\dot{K}_t = s_K Y_t - \delta K_t$$



Explosion conditions:

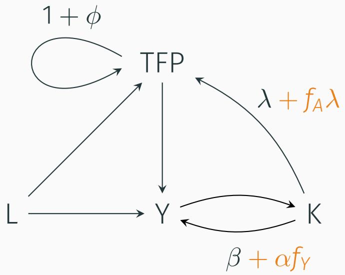
$$\phi > 0$$
$$\beta + \alpha f_Y > 1$$
$$(1+\phi) + \beta + \alpha f_Y - (1+\phi)(\beta + \alpha f_Y) + \lambda > 1$$

Introducing automation

$$\dot{A}_t = A_t^{1+\phi}(\ell_A L_t)^{\lambda(1-f_A)}(\kappa_A K_t)^{\lambda+f_A\lambda}$$

$$Y_t = A_t\,(\ell_Y L_t)^{\alpha(1-f_Y)}\,(\kappa_Y K_t)^{\beta+f_Y\alpha}$$

$$\dot{K}_t = s_K Y_t - \delta K_t$$



Explosion conditions:

$$\phi > 0$$

$$\beta + \alpha f_Y > 1$$

$$(1+\phi) + \beta + \alpha f_Y - (1+\phi)(\beta + \alpha f_Y) + \lambda + f_A\lambda > 1$$

The software-hardware model

Building blocks of the model

The software-hardware model

Scope of claims

Software-hardware model: overview

Canonical semi-endogenous growth model, plus:

Software-hardware model: overview

Canonical semi-endogenous growth model, plus:

1. Automation of labor with “AI”

Software-hardware model: overview

Canonical semi-endogenous growth model, plus:

1. Automation of labor with “AI”
2. AI = software · hardware · hardware quality

The software-hardware model of AI

The software-hardware model of AI

AI substituting for labor:

$$\text{AI} \equiv Z = \underbrace{S}_{\text{software}} \cdot \underbrace{C}_{\text{hardware}}$$

The software-hardware model of AI

AI substituting for labor:

$$\text{AI} \equiv Z = \underbrace{S}_{\text{software}} \cdot \underbrace{C}_{\text{hardware}}$$

- **Software:** "algorithmic efficiency"

The software-hardware model of AI

AI substituting for labor:

$$\text{AI} \equiv Z = \underbrace{S}_{\text{software}} \cdot \underbrace{C}_{\text{hardware}}$$

- **Software:** "algorithmic efficiency"
- **Hardware:** computer hardware ("compute")

The software-hardware model of AI

AI substituting for labor:

$$\text{AI} \equiv Z = \underbrace{S}_{\text{software}} \cdot \underbrace{C}_{\text{hardware}} = \underbrace{S}_{\text{software}} \cdot \underbrace{c \cdot h}_{\text{hardware}}$$

- **Software:** "algorithmic efficiency"
- **Hardware:** computer hardware ("compute")
  - Hardware quantity: c, "number of computer chips"
  - Hardware quality: h, "how many calculations (FLOPs) per chip"

Software and hardware evolution

Hardware accumulates: just another form of capital

$$C_t = s_C Y_t - \delta_C C_t$$

Software and hardware evolution

Hardware accumulates: just another form of capital

$$\dot{C}_t = s_C Y_t - \delta_C C_t$$

Software is like ideas: better software allows for *faster software progress*

$$\dot{S}_t = (\ell_S L_t)^{\lambda_S} S_t^{1+\phi_S}$$

Software and hardware evolution

Hardware accumulates: just another form of capital

$$C_t = s_C Y_t - \delta_C C_t$$

Software is like ideas: better software allows for *faster software progress*

$$\dot{S}_t = (\ell_S L_t)^{\lambda_S} S_t^{1+\phi_S}$$

Remember ideas production function:

$$\dot{A}_t = (\ell_A L_t)^{\lambda_A} A_t^{1+\phi_A}$$

Software and hardware evolution

Hardware accumulates: just another form of capital

$$\dot{C}_t = h_t s_C Y_t - \delta_C C_t$$

Software is like ideas: better software allows for *faster software progress*

$$\dot{S}_t = (\ell_S L_t)^{\lambda_S} S_t^{1+\phi_S}$$

Hardware quality is like ideas and investment-specific technical change: better hardware quality allows for *faster accumulation of effective hardware*

[a la Greenwood-Hercowitz-Krusell]

$$\dot{h} = (\ell_h L_t)^{\lambda_h} h_t^{1+\phi_h}$$

Software and hardware evolution

Hardware accumulates: just another form of capital

$$\dot{C}_t = h_t s_C Y_t - \delta_C C_t$$

Software is like ideas: better software allows for *faster software progress*

$$\dot{S}_t = (\ell_S L_t)^{\lambda_S} S_t^{1+\phi_S}$$

Hardware quality is like ideas and investment-specific technical change: better hardware quality allows for *faster accumulation of effective hardware*

[a la Greenwood-Hercowitz-Krusell]

$$\dot{h} = (\ell_h L_t)^{\lambda_h} h_t^{1+\phi_h}$$

AI: $\mathrm{AI} = \underbrace{S}_{\text{software}} \cdot \underbrace{c \cdot h}_{\text{hardware}}$

AI replaces human labor in tasks

AI replaces human labor in tasks

Automation by AI: AI AI replaces human labor in some fraction of economic tasks, f_x, in sector x.

AI replaces human labor in tasks

Automation by AI: AI AI replaces human labor in some fraction of economic tasks, f_x, in sector x.

Labor in sector X: *(without automation)*

$$L_{x,t} = \ell_x L_t$$

Effective labor in sector X: *(with automation)*

$$\hat{L}_{x,t} = (\ell_x L_t)^{1-f_x} \cdot Z_{x,t}{}^{f_x}$$

AI replaces human labor in tasks

Automation by AI: AI AI replaces human labor in some fraction of economic tasks, f_x, in sector x.

Labor in sector X: *(without automation)*

$$L_{x,t} = \ell_x L_t$$

Effective labor in sector X: *(with automation)*

$$\hat{L}_{x,t} = (\ell_x L_t)^{1-f_x} \cdot Z_{x,t}{}^{f_x}$$

$$= (\ell_x L_t)^{1-f_x} \cdot \left(\underbrace{S_t}_{\text{software}} \cdot \underbrace{c_{x,t} \cdot h_t}_{\text{hardware}} \right)^{f_x}$$

Note: effective labor accumulates

The software-hardware model: equations

Output: $Y_t = A_t \hat{L}_{Y,t}^{\alpha} K_t^{\beta}$

The software-hardware model: equations

Output: $Y_t = A_t \hat{L}_{Y,t}^{\alpha} K_t^{\beta}$

Accumulable factors: $\dot{K}_t = s_K Y_t - \delta_K K_t$

$$\dot{c}_t = h_t s_c Y_t - \delta_c c_t$$

The software-hardware model: equations

Output:

$$Y_t = A_t \hat{L}_{Y,t}^{\alpha} K_t^{\beta}$$

Accumulable factors:

$$\dot{K}_t = s_K Y_t - \delta_K K_t$$

$$\dot{c}_t = h_t s_c Y_t - \delta_c c_t$$

Ideas:

$$\dot{A}_t = \hat{L}_{A,t}^{\lambda_A} A_t^{1+\phi_A}$$

$$\dot{S}_t = \hat{L}_{S,t}^{\lambda_S} S_t^{1+\phi_S}$$

$$\dot{h}_t = \hat{L}_{h,t}^{\lambda_S} h_t^{1+\phi_h}$$

The software-hardware model: equations

Output: $Y_t = A_t \hat{L}_{Y,t}^{\alpha} K_t^{\beta}$

Accumulable factors:

$$\dot{K}_t = s_K Y_t - \delta_K K_t$$

$$\dot{c}_t = h_t s_c Y_t - \delta_c c_t$$

Ideas:

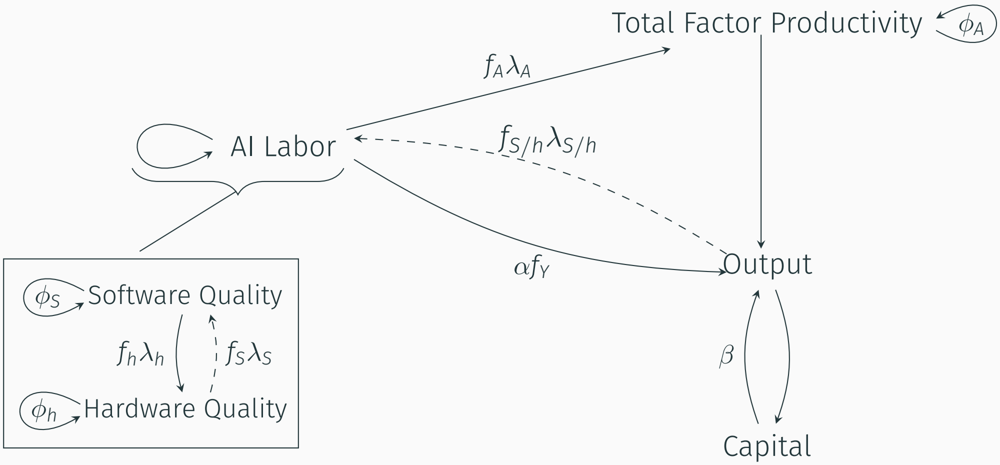
$$\dot{A}_t = \hat{L}_{A,t}^{\lambda_A} A_t^{1+\phi_A}$$

$$\dot{S}_t = \hat{L}_{S,t}^{\lambda_S} S_t^{1+\phi_S}$$

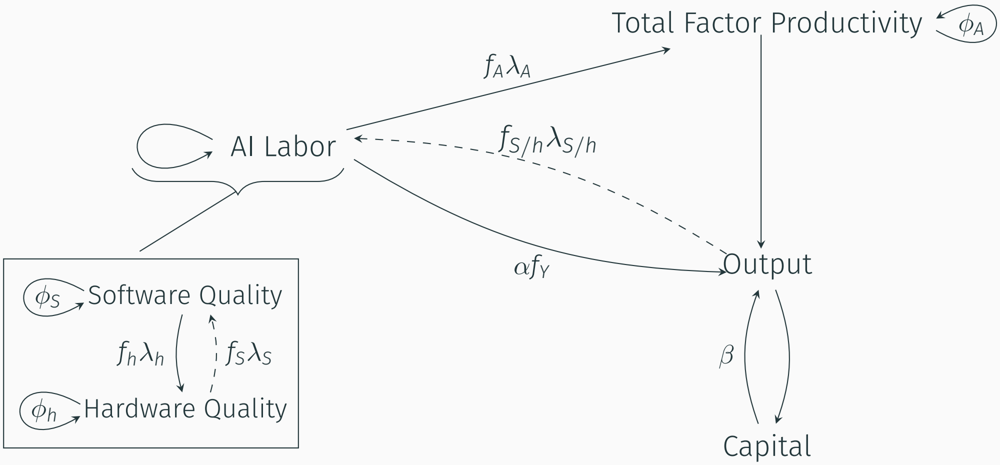
$$\dot{h}_t = \hat{L}_{h,t}^{\lambda_S} h_t^{1+\phi_h}$$

AI automation: $\hat{L}_{x,t} = L_{x,t}^{1-f_x} \cdot (S_t \cdot c_{x,t} \cdot h_t)^{f_x}$

The software-hardware model: diagram



The software-hardware model: diagram



Strength of feedback increasing with all exponents

Explosion condition

Simplify the problem by assuming complete depreciation. Substituting in effective labor expressions and removing non-accumulable factors

$$\dot{S}_t \propto S_t^{f_S\lambda_S\frac{1-\beta}{1-f_Y\alpha-\beta}+1+\phi_S} h_t^{f_S\lambda_S\frac{1-\beta}{1-f_Y\alpha-\beta}} A_t^{\frac{f_S\lambda_S}{1-f_Y\alpha-\beta}}$$

$$\dot{h}_t \propto S_t^{f_h\lambda_h\frac{1-\beta}{1-f_Y\alpha-\beta}} h_t^{f_h\lambda_h\frac{1-\beta}{1-f_Y\alpha-\beta}+1+\phi_h} A_t^{\frac{f_h\lambda_h}{1-f_Y\alpha-\beta}}$$

$$\dot{A}_t \propto S_t^{f_A\lambda_A\frac{1-\beta}{1-f_Y\alpha-\beta}} h_t^{f_A\lambda_A\frac{1-\beta}{1-f_Y\alpha-\beta}} A_t^{\frac{f_A\lambda_A}{1-f_Y\alpha-\beta}+1+\phi_A}$$

Explosion condition

Simplify the problem by assuming complete depreciation. Substituting in effective labor expressions and removing non-accumulable factors

$$\dot{S}_t \propto S_t^{f_S\lambda_S \frac{1-\beta}{1-f_Y\alpha-\beta}+1+\phi_S} h_t^{f_S\lambda_S \frac{1-\beta}{1-f_Y\alpha-\beta}} A_t^{\frac{f_S\lambda_S}{1-f_Y\alpha-\beta}}$$

$$\dot{h}_t \propto S_t^{f_h\lambda_h \frac{1-\beta}{1-f_Y\alpha-\beta}} h_t^{f_h\lambda_h \frac{1-\beta}{1-f_Y\alpha-\beta}+1+\phi_h} A_t^{\frac{f_h\lambda_h}{1-f_Y\alpha-\beta}}$$

$$\dot{A}_t \propto S_t^{f_A\lambda_A \frac{1-\beta}{1-f_Y\alpha-\beta}} h_t^{f_A\lambda_A \frac{1-\beta}{1-f_Y\alpha-\beta}} A_t^{\frac{f_A\lambda_A}{1-f_Y\alpha-\beta}+1+\phi_A}$$

Applying explosion proposition yields **explosion threshold**:

$$\frac{1}{1-\beta} f_A r_A + \frac{\alpha}{1-\beta} f_Y + f_S r_S + f_h r_h > 1$$

Explosion condition

Applying explosion proposition yields **explosion threshold**:

$$\frac{1}{1-\beta} f_A r_A + \frac{\alpha}{1-\beta} f_Y + f_S r_S + f_h r_h > 1$$

Explosion condition

Applying explosion proposition yields **explosion threshold**:

$$\frac{1}{1-\beta} f_A r_A + \frac{\alpha}{1-\beta} f_Y + f_S r_S + f_h r_h > 1$$

r **factor:** for $x \in \{A, S, h\}$,

$$r_x \equiv \frac{\lambda_x}{-\phi_x}$$

- Intuition: in canonical model, $g_A = r_A \cdot$ population growth

Calibrating parameters

Explosion condition: $\frac{1}{1-\beta} f_A r_A + \frac{\alpha}{1-\beta} f_Y + f_S r_S + f_h r_h > 1$

Term	Parameter	Estimate	Source
Labor share	$\alpha \equiv 1 - \beta$	0.6	–

Calibrating parameters

Explosion condition: $\frac{1}{1-\beta} f_A r_A + \frac{\alpha}{1-\beta} f_Y + f_S r_S + f_h r_h > 1$

Term	**Parameter**	**Estimate**	**Source**
Labor share	$\alpha \equiv 1-\beta$	0.6	–
TFP *r*-factor	r_A	0.32	Bloom et al (2024)

Calibrating parameters

Explosion condition: $\frac{1}{1-\beta} f_A r_A + \frac{\alpha}{1-\beta} f_Y + f_S r_S + f_h r_h > 1$

Term	Parameter	Estimate	Source
Labor share	$\alpha \equiv 1 - \beta$	0.6	–
TFP r-factor	r_A	0.32	Bloom et al (2024)
Hardware r-factor	r_h (Moore's law)	5	Bloom et al (2024)

Calibrating parameters

Explosion condition: $\frac{1}{1-\beta} f_A r_A + \frac{\alpha}{1-\beta} f_Y + f_S r_S + f_h r_h > 1$

Term	Parameter	Estimate	Source
Labor share	$\alpha \equiv 1 - \beta$	0.6	–
TFP r-factor	r_A	0.32	Bloom et al (2024)
Hardware r-factor	r_h (Moore's law)	5	Bloom et al (2024)
Software r-factor	r_S (Stockfish)	0.825	Erdil et al (2024)
	r_S (Other software)	≈1.3	Erdil et al (2024)

Calibrating parameters

Explosion condition: $0.5f_A + f_Y + f_S + 5f_h > 1$

Term	Parameter	Estimate	Source
Labor share	$\alpha \equiv 1 - \beta$	0.6	–
TFP *r*-factor	r_A	0.32	Bloom et al (2024)
Hardware *r*-factor	r_h (Moore's law)	5	Bloom et al (2024)
Software *r*-factor	r_S (Stockfish)	0.825	Erdil et al (2024)
	r_S (Other software)	≈1.3	Erdil et al (2024)

Calibrating parameters

Explosion condition: $0.5f_A + f_Y + f_S + 5f_h > 1$

Term	Parameter	Estimate	Source
Labor share	$\alpha \equiv 1 - \beta$	0.6	–
TFP *r*-factor	r_A	0.32	Bloom et al (2024)
Hardware *r*-factor	r_h (Moore's law)	5	Bloom et al (2024)
Software *r*-factor	r_S (Stockfish)	0.825	Erdil et al (2024)
	r_S (Other software)	≈1.3	Erdil et al (2024)

Interpretation: Software and hardware have **much lower** diminishing returns to research than the rest of the economy

Calibrating parameters

Explosion condition: $0.5f_A + f_Y + f_S + 5f_h > 1$

Term	Parameter	Estimate	Source
Labor share	$\alpha \equiv 1 - \beta$	0.6	–
TFP *r*-factor	r_A	0.32	Bloom et al (2024)
Hardware *r*-factor	r_h (Moore's law)	5	Bloom et al (2024)
Software *r*-factor	r_S (Stockfish)	0.825	Erdil et al (2024)
	r_S (Other software)	≈1.3	Erdil et al (2024)

Interpretation: Software and hardware have **much lower** diminishing returns to research than the rest of the economy $\implies$ if software/hardware grow as share of economy, large growth effects

Scope of claims

Building blocks of the model

The software-hardware model

Scope of claims

What we do *not* speak to

1. **No bottlenecks** (e.g. compute, data)
   - Cobb-Douglas technology $\implies$ one thing can always substitute for another

What we do *not* speak to

1. **No bottlenecks** (e.g. compute, data)
   - Cobb-Douglas technology $\implies$ one thing can always substitute for another
2. **What about "non-explosive" growth accelerations?**

What we do *not* speak to

1. **No bottlenecks** (e.g. compute, data)
    - Cobb-Douglas technology $\implies$ one thing can always substitute for another
2. **What about "non-explosive" growth accelerations?**
3. **Quality of parameter value estimates**

What we do *not* speak to

1. **No bottlenecks** (e.g. compute, data)
   - Cobb-Douglas technology $\implies$ one thing can always substitute for another
2. **What about “non-explosive” growth accelerations?**
3. **Quality of parameter value estimates**
4. **Endogenous automation**

What we do *not* speak to

1. **No bottlenecks** (e.g. compute, data)
   - Cobb-Douglas technology $\implies$ one thing can always substitute for another
2. **What about "non-explosive" growth accelerations?**
3. **Quality of parameter value estimates**
4. **Endogenous automation**
5. **More:**
   - Endogenous allocation rules
   - Decentralized allocation: roles of industrial organization + externalities
   - Learning by doing
   - Capital adjustment costs
   - Time to build

Conclusion: “Why wasn’t automating agriculture enough for a growth explosion?”

Tractors $\rightarrow$ more food $\rightarrow$ more people $\rightarrow$ better tractors $\rightarrow \cdots$

Conclusion: “Why wasn’t automating agriculture enough for a growth explosion?”

Tractors $\rightarrow$ more food $\rightarrow$ more people $\rightarrow$ better tractors $\rightarrow \cdots$

1. Maybe it was? Our condition speaks to ‘are we *on track*’ for a growth explosion

Conclusion: "Why wasn't automating agriculture enough for a growth explosion?"

Tractors $\rightarrow$ more food $\rightarrow$ more people $\rightarrow$ better tractors $\rightarrow \cdots$

1. **Maybe it was?** Our condition speaks to 'are we *on track*' for a growth explosion

2. **'Diminishing returns' is one reason**; diminishing returns are *less* strong in hardware and software

Conclusion: "Why wasn't automating agriculture enough for a growth explosion?"

Tractors $\rightarrow$ more food $\rightarrow$ more people $\rightarrow$ better tractors $\rightarrow \cdots$

1. Maybe it was? Our condition speaks to 'are we *on track*' for a growth explosion

2. 'Diminishing returns' is one reason; diminishing returns are *less* strong in hardware and software

3. Bottlenecks or other limits: we do not speak to *all* limits

Thank you!

Appendix

Appendix

On bottlenecks

Cobb-Douglas: with $\alpha > 0$

$$Y = L^{\alpha} K^{1-\alpha}$$

Fix L, send $K \to \infty \implies Y \to \infty$.

Potential bottlenecks:

- **Compute** bottlenecking algorithmic progress
- **Algorithmic progress** bottlenecking compute
- **Energy** bottlenecking everything
- **Data** bottlenecking everything

CES with complements: with $\phi < 0$

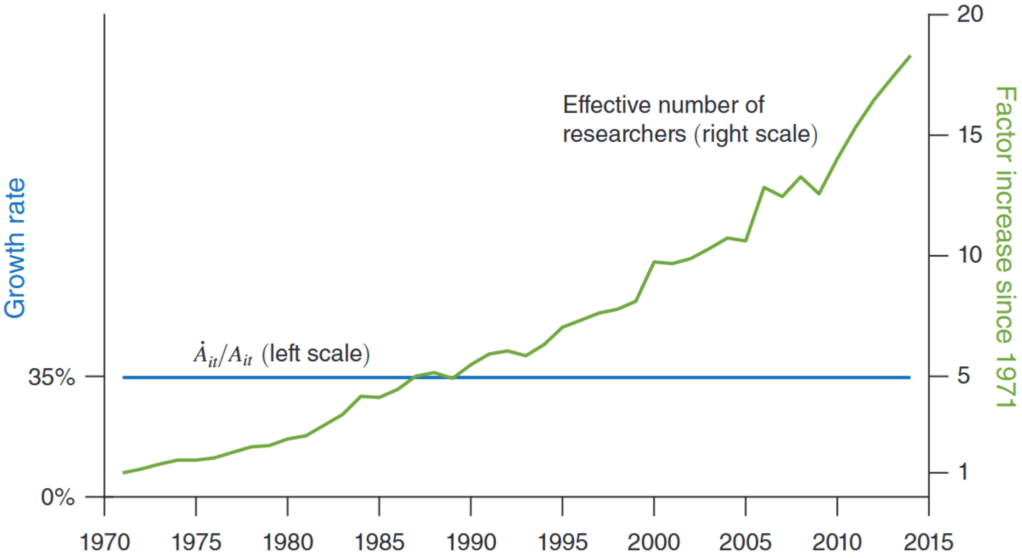
$$Y = \left[L^{\phi} + K^{\phi}\right]^{1/\phi}$$

Fix L, send $K \to \infty \implies Y = L$

Potential reasons to think bottlenecks will be less of an issue:

- 2x efficient algorithims $\implies$ 2x as many experiments
- Aum and Shin (2024): software and labor are substitutes not complements

Could ϕ be falling over time? Doesn't appear to be for Moore's Law



Multisector semi-endogenous growth model

Standard one-sector model:

- Idea production functions:

$$\dot{A}_t = L_t^{\lambda} A_t^{1+\phi}$$

- BGP:

$$\frac{\dot{A}}{A} = \frac{\lambda}{-\phi} n$$

Multisector semi-endogenous growth model

Standard one-sector model:

- Idea production functions:

$$\dot{A}_t = L_t^{\lambda} A_t^{1+\phi}$$

- BGP:

$$\frac{\dot{A}}{A} = \frac{\lambda}{-\phi} n$$

Two-sector model:

- Aggregate TFP: $A_t = A_{1t}^{\sigma_1} A_{2t}^{\sigma_2}$
- Idea production functions:*

$$\dot{A}_{it} = (s_i L_t)^{\lambda_i} A_{it}^{1+\phi_i}$$

* s_i exogenous and constant ("Solow-style"). It can be shown, though, that optimally s_1/s_2 is constant under Cobb-Douglas aggregation.

Multisector semi-endogenous growth model

Standard one-sector model:

- Idea production functions:

$$\dot{A}_t = L_t^{\lambda} A_t^{1+\phi}$$

- BGP:

$$\frac{\dot{A}}{A} = \frac{\lambda}{-\phi} n$$

Two-sector model:

- Aggregate TFP: $A_t = A_{1t}^{\sigma_1} A_{2t}^{\sigma_2}$
- Idea production functions:*

$$\dot{A}_{it} = (s_i L_t)^{\lambda_i} A_{it}^{1+\phi_i}$$

- BGP:

$$\frac{\dot{A}}{A} = \sum_i \left[\sigma_i \frac{\lambda_i}{-\phi_i} n \right]$$

* s_i exogenous and constant ("Solow-style"). It can be shown, though, that optimally s_1/s_2 is constant under Cobb-Douglas aggregation.

Multisector semi-endogenous growth model

Standard one-sector model:

- Idea production functions:

$$\dot{A}_t = L_t^{\lambda} A_t^{1+\phi}$$

- BGP:

$$\frac{\dot{A}}{A} = \frac{\lambda}{-\phi} n$$

Two-sector model:

- Aggregate TFP: $A_t = A_{1t}^{\sigma_1} A_{2t}^{\sigma_2}$
- Idea production functions:*

$$\dot{A}_{it} = (s_i L_t)^{\lambda_i} A_{it}^{1+\phi_i}$$

- BGP:

$$\frac{\dot{A}}{A} = \sum_i \left[\sigma_i \frac{\lambda_i}{-\phi_i} n \right]$$

Comparative static: Suppose $-\phi_1 > -\phi_2$. Increase σ_2. Obviously $g_A \uparrow$

* s_i exogenous and constant ("Solow-style"). It can be shown, though, that optimally s_1/s_2 is constant under Cobb-Douglas aggregation.