Artificial Intelligence and Economic Growth

Chad Jones
Stanford GSB

NBER Innovation Bootcamp – July 2025

What are the implications of A.I. for economic growth?

- Build some growth models with A.I.
  - A.I. helps to make goods
  - A.I. helps to make ideas
- Implications
  - Long-run growth
  - Share of GDP paid to labor vs capital
- Catastrophic risks from A.I.?

Talk based on material from several papers

- Aghion, B. Jones, and C. Jones (2019) “Artificial Intelligence and Economic Growth”
- Jones (2024 AER Insights) “The A.I. Dilemma: Growth versus Existential Risk”
- Jones (2025) “How much should we spend to reduce A.I.’s existential risk?”

Two Main Themes (Aghion, B. Jones, and C. Jones, 2019)

- A.I. modeled as a continuation of automation
  - Automation = replace labor in particular tasks with machines and algorithms
  - *Past:* textile looms, steam engines, electric power, computers
  - *Future:* driverless cars, paralegals, pathologists, maybe researchers, maybe everyone?

- A.I. may be limited by Baumol's cost disease
  - *Baumol:* growth constrained not by what we do well but rather by what is essential and yet hard to improve

The Zeira 1998 Model

Simple Model of Automation (Zeira 1998)

- Production uses n tasks/goods:

$$Y = AX_1^{\alpha_1} X_2^{\alpha_2} \cdot \ldots \cdot X_n^{\alpha_n},$$

where $\sum_{i=1}^{n} \alpha_i = 1$ and

$$X_{it} = \begin{cases} L_{it} & \text{if not automated} \\ K_{it} & \text{if automated} \end{cases}$$

- Substituting gives

$$Y_t = A_t K_t^{\alpha} L_t^{1-\alpha}$$

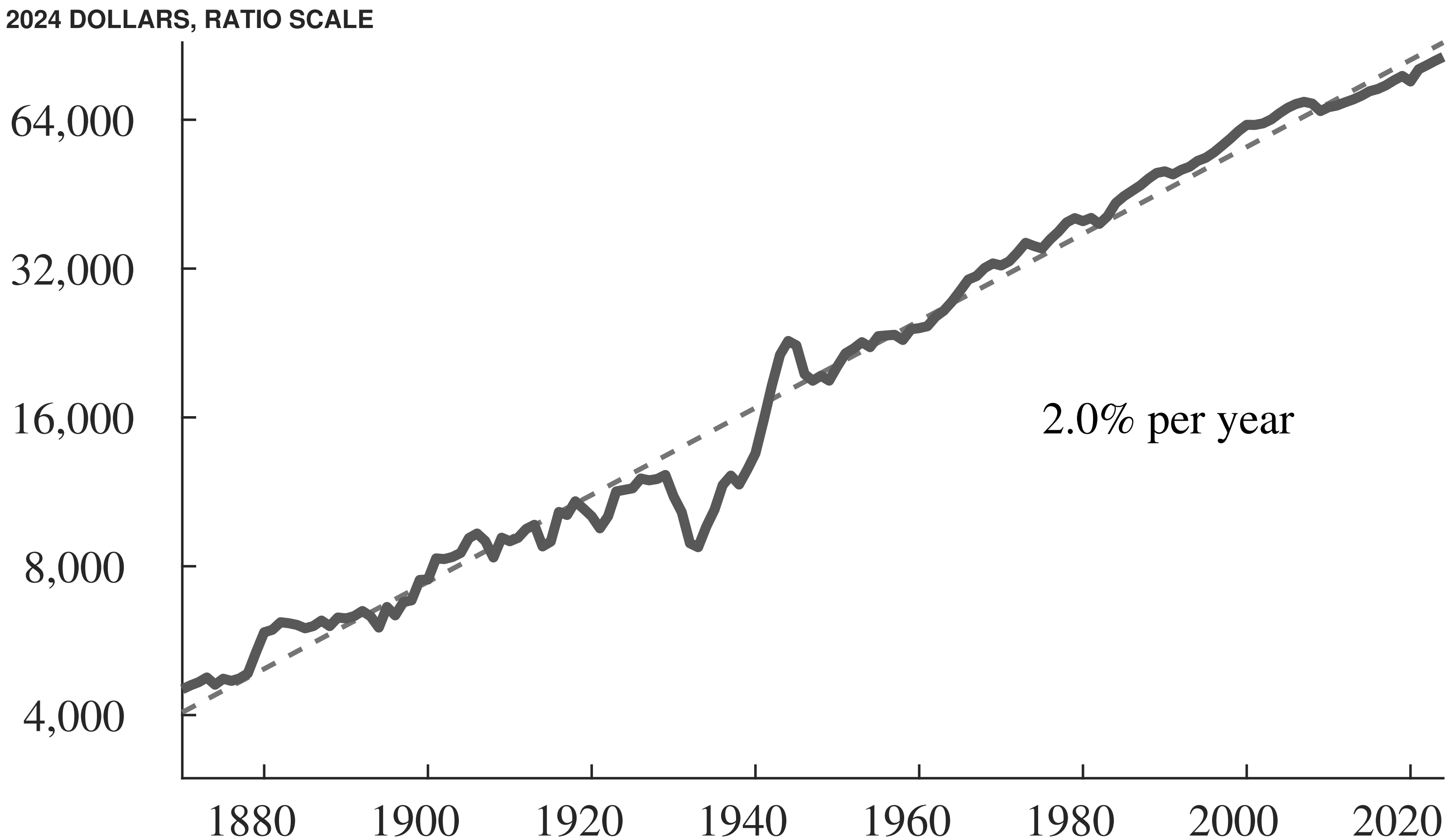
$$Y_t = A_t K_t^{\alpha} L_t^{1-\alpha}$$

- Comments:
  - α reflects the *fraction* of tasks that are automated
  - Embed in neoclassical growth model $\Rightarrow$

$$g_y = \frac{g_A}{1-\alpha} \quad \text{where} \quad y_t \equiv Y_t / L_t$$

- Automation: $\uparrow \alpha$ raises both capital share and LR growth
  - Hard to reconcile with 20th century
  - Substantial automation but stable growth and capital shares

Average income per person in the U.S.



Recent papers

- Acemoglu and Restrepo (2017, 2018, 2019, 2020, 2021, 2022, 2023)
  - Foundational work in this literature
  - Old tasks are gradually automated as new (labor) tasks are created
  - Fraction automated can then be steady
  - Rich framework, with endogenous innovation and automation
  - Acemoglu-Restrepo (2022 ECMA): Rising automation can explain 60% of changes in the U.S. wage distribution since 1980
- Hemous and Olson (2016, 2025)
- B. Jones and Liu (2024)

Automation and Baumol's Cost Disease

AJJ Economic Environment

Final good	$Y_t = \left(\int_0^1 y_{it}^{\frac{\sigma-1}{\sigma}} di\right)^{\frac{\sigma}{\sigma-1}}$ where $\sigma < 1$ **(Baumol effect)**
Tasks	$y_{it} = \begin{cases} K_{it} & \text{if automated} \quad i \in [0, \beta_t] \\ L_{it} & \text{if not automated} \quad i \in [\beta_t, 1] \end{cases}$
Capital accumulation	$\dot{K}_t = I_t - \delta K_t$
Resource constraint (K)	$\int_0^1 K_{it} di = K_t$
Resource constraint (L)	$\int_0^1 L_{it} di = L$
Resource constraint (Y)	$Y_t = C_t + I_t$
Allocation	$I = \bar{s}_K Y$

Automation and growth

- Combining equations

$$Y_t = \left[\beta_t \left(\frac{K_t}{\beta_t}\right)^{\frac{\sigma-1}{\sigma}} + (1-\beta_t)\left(\frac{L}{1-\beta_t}\right)^{\frac{\sigma-1}{\sigma}}\right]^{\frac{\sigma}{\sigma-1}}$$

- How β interacts with K: two effects
    - β: what fraction of tasks have been automated
    - β: Dilution as $K/\beta \Rightarrow K$ spread over more tasks
- Same for labor: $L/(1-\beta_t)$ means given L concentrated on fewer tasks, raising "effective labor"

Rewriting in classic CES form

- Collecting the β terms into factor-augmenting form:

$$Y_t = F(B_t K_t, A_t L_t)$$

where

$$B_t = \left(\frac{1}{\beta_t}\right)^{\frac{1}{1-\sigma}} \text{ and } A_t = \left(\frac{1}{1-\beta_t}\right)^{\frac{1}{1-\sigma}}$$

- Effect of automation: $\uparrow \beta_t \;\Rightarrow\; \downarrow B_t$ and $\uparrow A_t$

Intuition: dilution effects just get magnified since $\sigma < 1$

Automation

- Suppose a constant fraction of non-automated tasks get automated every period:

$$\dot{\beta}_t = \theta(1 - \beta_t)$$

$$\Rightarrow \quad \beta_t \to 1$$

- What happens to $1 - \beta_t =: m_t$?

$$\frac{\dot{m}_t}{m_t} = -\theta$$

The fraction of labor-tasks falls at a constant exponential rate

Putting it all together

$$Y_t = F(B_t K_t, A_t L_t) \text{ where } B_t = \left(\frac{1}{\beta_t}\right)^{\frac{1}{1-\sigma}} \text{ and } A_t = \left(\frac{1}{1-\beta_t}\right)^{\frac{1}{1-\sigma}}$$

- $\beta_t \to 1 \Rightarrow B_t \to 1$
- But A_t grows at a constant exponential rate!

$$\frac{\dot{A}_t}{A_t} = -\frac{1}{1-\sigma}\frac{\dot{m}_t}{m_t} = \frac{\theta}{1-\sigma}$$

- When a constant fraction of remaining goods get automated and $\sigma < 1$, the automation model features an asymptotic BGP that satisfies Uzawa

$$\alpha_{Kt} \equiv \frac{F_K K}{Y} = \beta_t^{\frac{1}{\sigma}} \left(\frac{K_t}{Y_t}\right)^{\frac{\sigma-1}{\sigma}} \to \left(\frac{\bar{s}_K}{g_Y + \delta}\right)^{\frac{\sigma-1}{\sigma}} < 1$$

Intuition for AJJ result

- Why does automation lead to balanced growth and satisfy Uzawa?
  - $\beta_t \to 1$ so the KATC piece "ends" eventually
  - Labor per task: $L/(1-\beta_t)$ rises exponentially over time!
  - Constant population, but concentrated on an exponentially shrinking set of goods $\Rightarrow$ exponential growth in "effective" labor
- Labor earns 2/3 of GDP even though labor tasks are vanishing
  - Baumol: these are the tasks that are scarce and essential, so they demand a high share of GDP
- Limitation
  - An asymptotic result
  - Only occurs as $\beta_t \to 1$, so unclear if relevant for U.S. or other modern economies

B. Jones and Liu (AER 2024)

- BGP can occur "today" with $\beta_t < 1$, not asymptotically
  - Adds capital-augmenting technical change ("faster computers") = Z_t
  - Capital share is $\alpha_{Kt} = \beta_t \,/\, Z_t$
  - Might describe modern economies

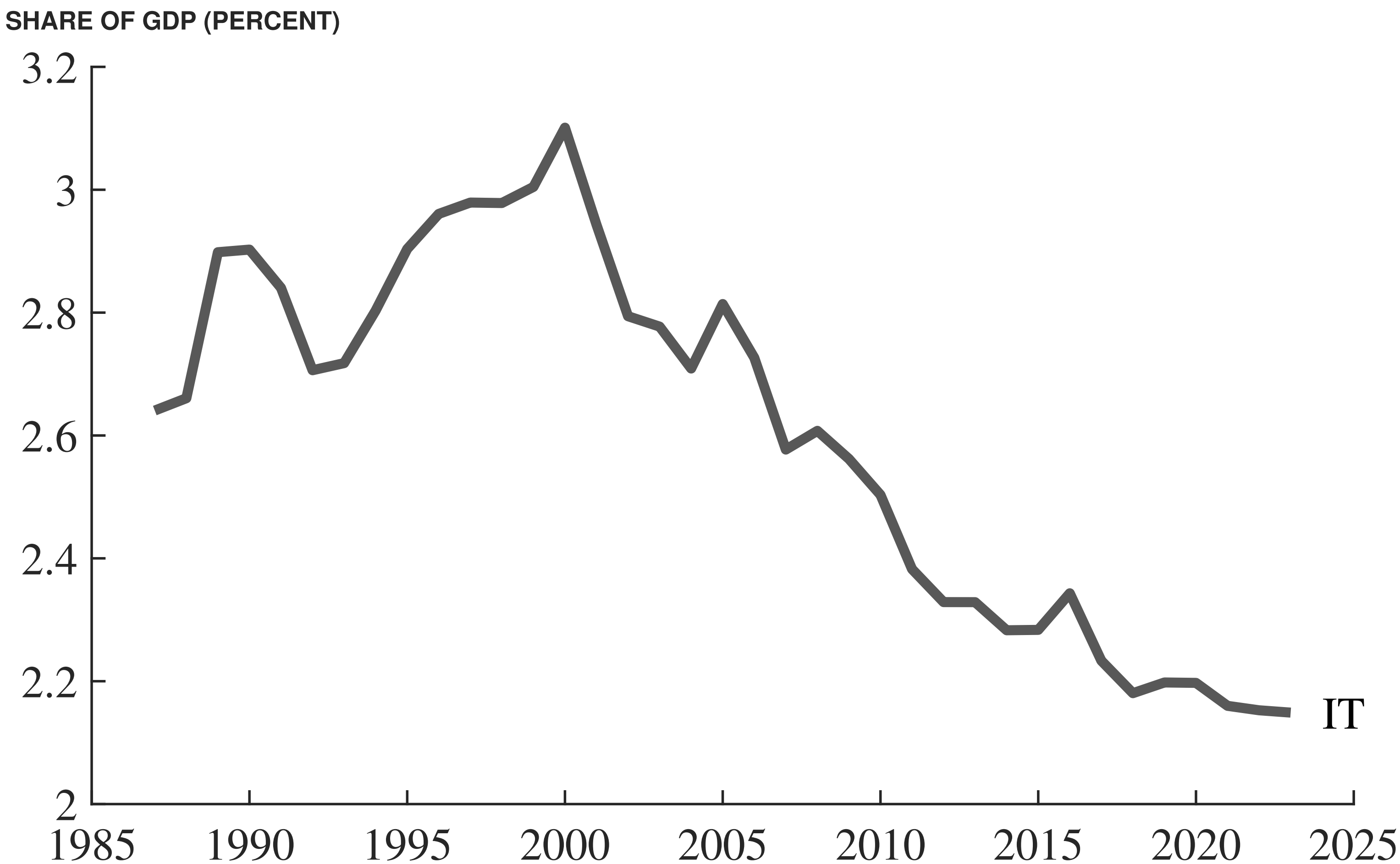
- Automation and KATC coexist along the BGP with stable factor shares
  - If β_t and Z_t rise at the same rate.

- But notice that as $\beta_t \to 1$, if $\uparrow Z_t$ continues, then the capital share falls to zero!
  - With $\sigma < 1$, the ever declining price of computers drives its factor share to zero

New project with Chris Tonetti (in progress)

- Generalize the basic model shown so far and quantify it
  - How much of historical growth in Agriculture, Motor Vehicles, and other key sectors is due to automation?
- Idea production functions?
  - How much of growth in software is due to automation?
  - Other idea PFs (harder since need to measure output of ideas)
- Speculate on what growth over the next decade due to A.I. might look like using the previous quantifications as a guide

Share of Factor Payments: Information Technology (Jones and Tonetti)



A.I. and Ideas

A.I. in the Idea Production Function

- Let production of goods and services be $Y_t = A_t L_t$
- Let idea production be:

$$\dot{A}_t = A_t^{\phi} \left(\int_0^1 X_{it}^{\frac{\sigma-1}{\sigma}} di \right)^{\frac{\sigma}{\sigma-1}} , \ \sigma < 1$$

- Assume fraction β_t of tasks are automated by date t. Then:

$$\dot{A}_t = A_t^{\phi} F(B_t K_t, C_t S_t) \ \text{ where } \ B_t = \left(\frac{1}{\beta_t} \right)^{\frac{1}{1-\sigma}} \ \text{ and } \ C_t = \left(\frac{1}{1-\beta_t} \right)^{\frac{1}{1-\sigma}}$$

- This is like before...

A.I. in the Idea Production Function

- Intuition: with $\sigma < 1$ the scarce factor comes to dominate

$$F(B_tK_t, C_tS_t) = C_tS_t\, F\left(\frac{B_tK_t}{C_tS_t}, 1\right) \to \text{Constant} \cdot C_tS_t$$

- So, with continuous automation

$$\dot{A}_t \to A_t^{\phi} C_t S_t$$

- And asymptotic balanced growth path becomes

$$g_A = \frac{g_C + g_S}{1 - \phi}$$

- We get a “boost” from continued automation (g_C)

Theory: A.I. can raise growth

- Automation (computers, internet, etc.) has been ongoing for decades
  - Recall $g_C = \frac{1}{1-\sigma} \cdot \theta$
  - where θ is the fraction of remaining labor tasks that get automated each year

    $\Rightarrow$ continued automation by itself may not raise growth

- However, an increase in the rate of automation via A.I. $\uparrow \theta$ could raise growth
  - Rapid advances in reasoning models (OpenAI's o1-pro, o3) suggest possible!

- Extreme version: If all research tasks are automated, then

$$\dot{A}_t = K_t A_t^{\phi}$$

and a growth explosion is possible (e.g. if $\phi > 0$)

What would A.I. accelerating economic growth look like?

- Near-term productivity boosts from A.I.
  - Software: 25% productivity improvements already
  - In the next decade: A.I. agents that can automate most coding?
  - Virtuous circle: code up even better A.I. agents
- With Moore's Law price decreases $\Rightarrow$ millions($\uparrow$) of virtual research assistants
  - Automate cognitive tasks $\Rightarrow$ invent new ideas
  - E.g. better chips, better robots, medical technologies, etc.
  - A.I. + robots for physical tasks
- Potential to raise growth rates substantially over the next two decades?

Bottlenecks and Baumol Effects

- Economic history $\Rightarrow$ may take longer than we expect
  - Electricity and computers changed the economy over 50 years
- Automation has been going on for 150 years with no speed up in growth
  - Electricity, engines, semiconductors, the internet, smartphones
  - Yet growth always 2% per year
- Maybe those great ideas are what *kept* growth from slowing
  - Perhaps A.I. = latest great idea letting us maintain 2% growth for a while longer. (pessimistic view, but possible)

The Labor Market, Jobs, and Meaningful Work

- The world where A.I. “changes everything” is a world where GDP is incredibly high
  - The size of the pie available for redistribution is enormous
  - Transition could be hard
- As we get richer, we naturally work less
  - Rising leisure, lower retirement ages. This is a good thing!
  - “Work” is a bad in most of our models
- But there is also good work, meaningful work
  - Chess more popular than ever despite iPhone $>$ Magnus Carlsen
  - We may choose to value experiences involving people (arts, music, sports)
    Keeps labor share high?

Catastrophic Risks?

Can we use economic analysis to think about the serious risks?

Two Versions of Existential Risk

- Bad actors:
  - Could use Claude/GPT-6 to cause harm
  - E.g. design a new virus that is extremely lethal and takes 3 weeks for symptoms
  - Nuclear weapons mangeable because so rare; if every person had them...
- Alien intelligence:
  - How would we react to a spaceship near Saturn on the way to Earth?
  - “How do we have power over entities more powerful than us, forever?” (Stuart Russell)

A Thought Experiment (Jones, 2024 AERI)

- AGI more important than electricity, but more dangerous than nuclear weapons?

- The Oppenheimer Question:

  - If nothing goes wrong, AGI accelerates growth to 10% per year

  - But a one-time small chance that A.I. kills everyone

  - Develop or not? What risk are you willing to take: 1%? 10%?

What does standard economic analysis imply?

Findings:

- Log utility: Willing to take a 33% risk!

  (Maybe entrepreneurs are not very risk averse?)

- More risk averse ($\gamma = 2$ or 3), risk cutoff plummets to 2% or less
  - Diminishing returns to consumption
  - We do not need a 4th flat screen TV or a 3rd iphone.
    Need more years of life to enjoy already high living standards.

- But 10% growth $\Rightarrow$ cure cancer, heart disease
  - Even $\gamma = 3$ willing to take large risks (25%) to cut mortality rates in half
  - Each person dies from cancer or dies from A.I. Just total risk that matters. . .
  - True even if the social discount rate falls to zero

How much should we spend to reduce A.I.'s catastrophic risk? (Jones 2025)

How much should we spend to reduce A.I.'s catastrophic risk? (Jones 2025)

- Covid pandemic: "spent" 4% of GDP to mitigate a mortality risk of 0.3%
  - A.I. risk is at least this large — survey of experts: 5% median

    $\Rightarrow$ spend at least this much?
  - Are we massively underinvesting in mitigating this risk?

How much should we spend to reduce A.I.'s catastrophic risk? (Jones 2025)

- Covid pandemic: "spent" 4% of GDP to mitigate a mortality risk of 0.3%
  - A.I. risk is at least this large — survey of experts: 5% median

    $\Rightarrow$ spend at least this much?
  - Are we massively underinvesting in mitigating this risk?
- Better intuition
  - VSL = $10 million
  - To avoid a mortality risk of 1% $\Rightarrow$ WTP = $1\% \times \$10$ million $= \$100,000$
  - This is more than 100% of a year's per capita GDP
  - Xrisk over two decades $\Rightarrow$ **annual investment of 5% of GDP**
- Large investments worthwhile, even with no value on future generations

How much should we spend to reduce A.I.'s catastrophic risk? (Jones 2025)

- Covid pandemic: "spent" 4% of GDP to mitigate a mortality risk of 0.3%
  - A.I. risk is at least this large — survey of experts: 5% median

    $\Rightarrow$ spend at least this much?
  - Are we massively underinvesting in mitigating this risk?

- Better intuition
  - VSL = $10 million
  - To avoid a mortality risk of 1% $\Rightarrow$ WTP = $1\% \times \$10$ million $= \$100,000$
  - This is more than 100% of a year's per capita GDP
  - Xrisk over two decades $\Rightarrow$ **annual investment of 5% of GDP**

- Large investments worthwhile, even with no value on future generations

Incomplete so far: how effective is mitigation?

Model

- Setup
  - One-time existential risk at probability $\delta(x)$
  - One-time investment x to mitigate the risk ($\delta'(x) < 0$)
  - Exogenous endowment y_t (grows rapidly via A.I.)

- Optimal mitigation:

$$\max_{x_t} u(c_t) + (1 - \delta(x_t))\, \beta\, V_{t+1}$$

$$s.t.\ \ c_t + x_t = y_t$$

$$V_{t+1} = \sum_{\tau=0}^{\infty} \beta^{\tau} u(y_{t+1+\tau})$$

(consume y_t in future)

Optimal Mitigation

- FOC:

$$u'(c_t) = -\delta'(x_t)\beta V_{t+1}$$

- Let $\eta_{\delta,x} \equiv -\frac{\delta'(x_t)x_t}{\delta(x_t)}$ and $s_t \equiv x_t/y_t$

$$\frac{s_t}{1-s_t} = \underbrace{\eta_{\delta,x}}_{\substack{\text{effectiveness}\\\text{of spending}\\ >0.01?}} \times \underbrace{\delta(x_t)}_{\substack{\text{risk to be}\\\text{mitigated}\\ 1\%?}} \times \underbrace{\beta\frac{V_{t+1}}{u'(c_t)\,c_t}}_{\substack{\text{value of}\\\text{life}\\ >180}}$$

- Taking the smallest numbers:

$$\frac{s}{1-s} \geq 0.01 \times 1\% \times 180 = 1.8\%$$

Functional forms

- Existential risk: $$\delta(x) = (1-\phi)\delta_0 + \phi\delta_0 e^{-\alpha N x}$$
  - δ_0 is the risk without mitigation
  - ϕ is the share of the risk that can be eliminated by spending
  - α is the effectiveness of spending
  - N is the number of people each spending x
  - With infinite spending, risk falls to $(1-\phi)\delta_0$
- To calibrate α:

$$\alpha N = -T \log(1-\xi) \approx \xi T$$

ξ is the share of the risk that can be eliminated by spending 100% of GDP for one year

T is "time of perils" = years until risk gets realized (period length)

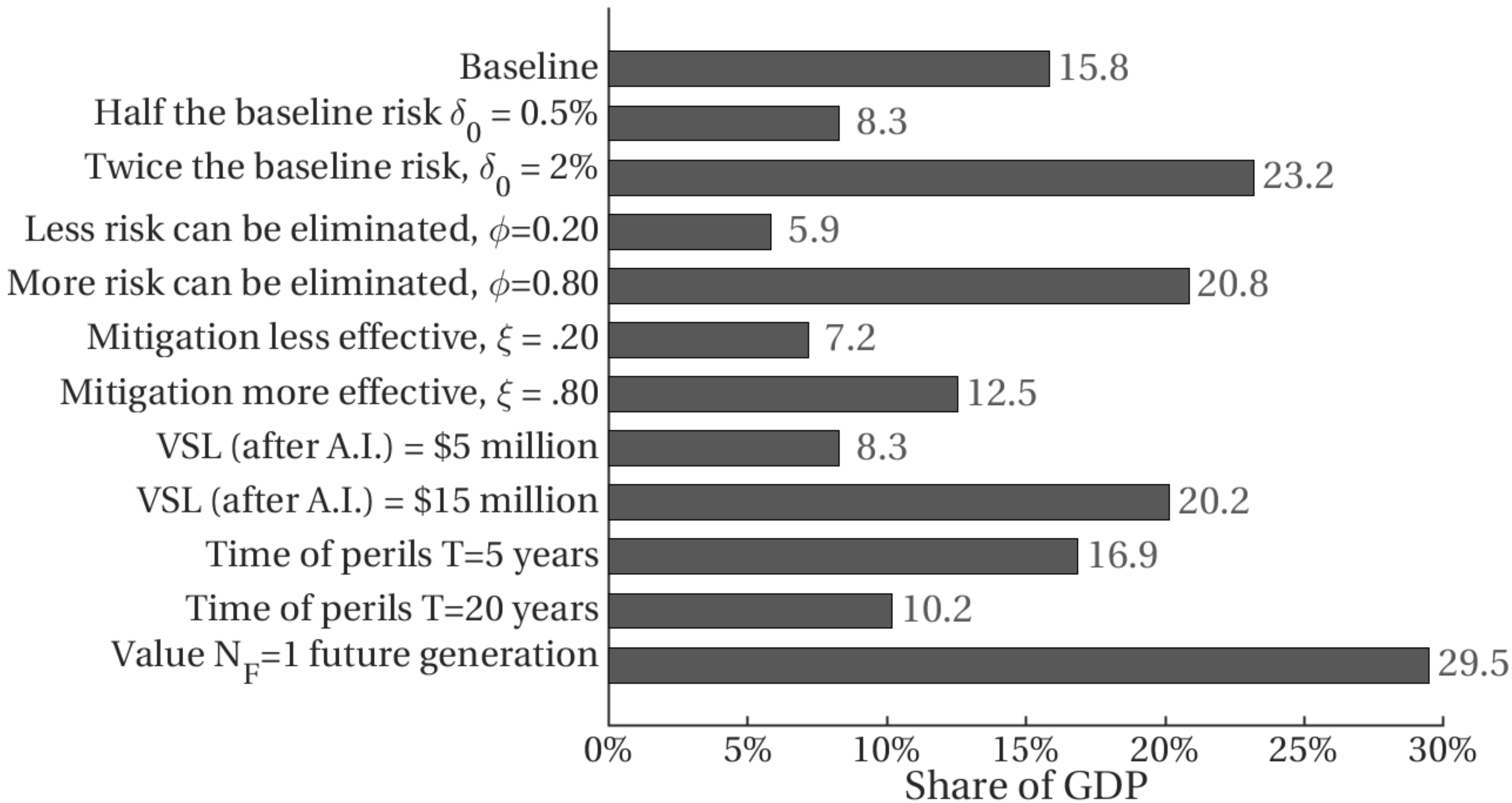
Calibration

$$\delta(x) = (1-\phi)\delta_0 + \phi\delta_0 e^{-\alpha N x}$$

	Parameter	Value	Distribution
Extinction risk, no mitigation	δ_0	1%	Uniform (0%, 2%)
Share that can be eliminated	ϕ	0.5	Uniform (0, 1)
Effectiveness of spending	ξ	0.5	Uniform (0, 0.99)
Value of life	$V_{t+1}/u'(y_t)$	180	Uniform (0.5*180, 1.5*180)
Time of perils (period length)	T	10 years	Uniform (5, 20)
CRRA	θ	2	...
Discount factor	β	0.99^T	...
Value of future generations		**0**	purely selfish for now

Baseline case: Spending a year's GDP reduces risk from 1% to 0.75%

Optimal Spending to Reduce Existential Risk



When should we not invest in mitigation?

- From FOC: Do not invest if $u'(y_0) > -\delta'(0)\beta V_{t+1}$
- Using functional forms and approximations:

$$1 \; > \; \alpha N \; \cdot \; \phi\delta_0\beta\frac{V_{t+1}}{u'(y_0)} \;\; \approx \;\; \underbrace{\xi\, T}_{\text{effectiveness of spending}} \;\; \cdot \;\; \underbrace{\phi\delta_0\beta\tfrac{V_{t+1}}{u'(y_0)}}_{\text{WTP = EV of lives lost to x-risk}}$$

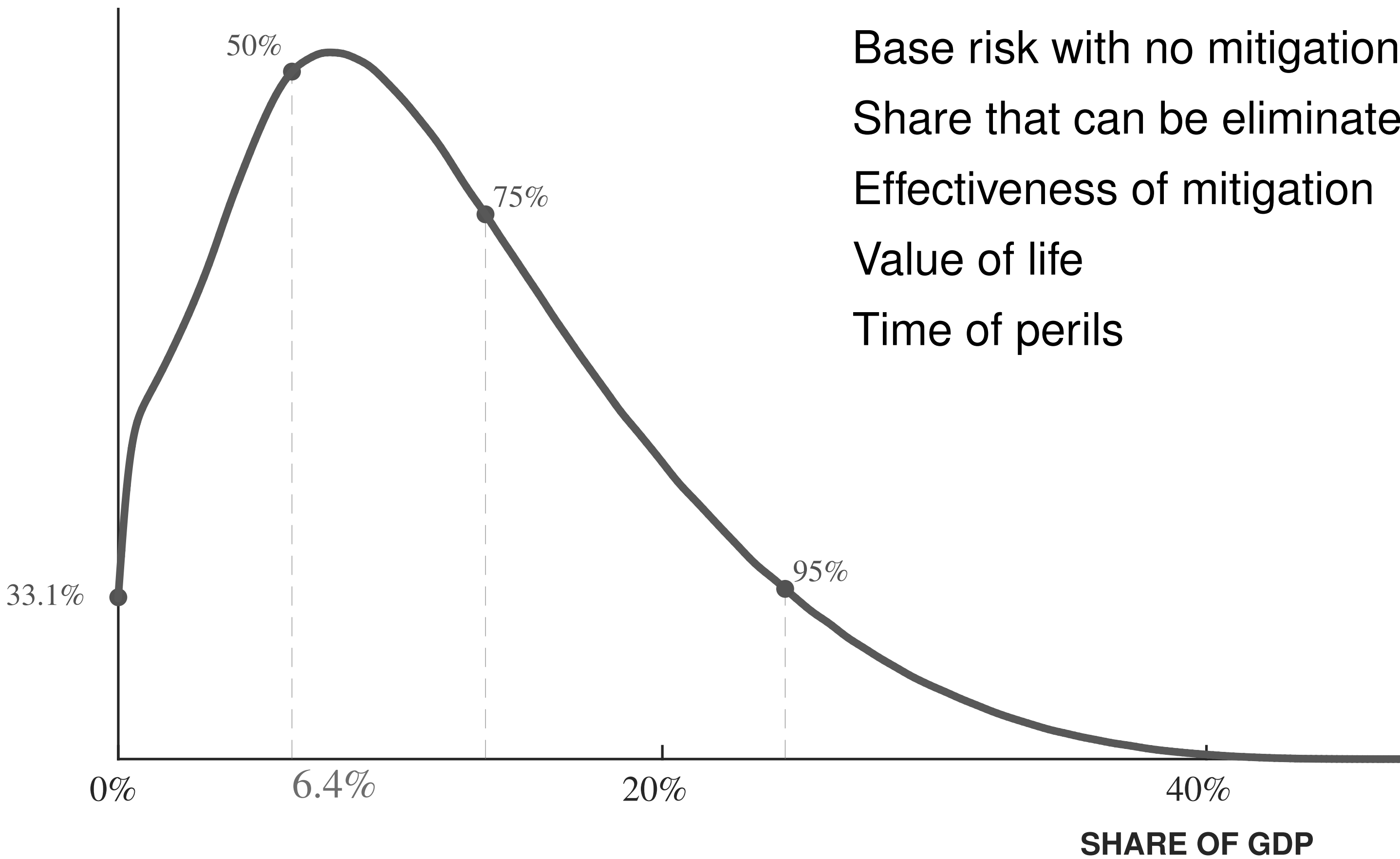
$$\Longrightarrow \;\; \xi\, T \;\cdot\; \text{WTP} < \; 1$$

- $\xi = 1/2$, $T = 10$, and WTP $= 60\%$ of GDP, LHS = 3
  - But ϕ or ξ or δ_0 $\Rightarrow$ 5x smaller $\Rightarrow$ invest zero   (Little risk, or not much can be done)

Monte Carlo Results

10 million simulations

Optimal Mitigation: Monte Carlo Simulation



Uniform distributions over:

Base risk with no mitigation	0 – 2%
Share that can be eliminated	0 – 100%
Effectiveness of mitigation	0 – 99%
Value of life	\$5m – \$15m
Time of perils	5 – 20 years

Mean = 8%. 65% of runs have $s \geq 1\%$

Summary Statistics for Monte Carlo Simulations

	Selfish baseline $(= 0)$ $\delta_0 \sim \texttt{Uniform[0,2\%]}$	Modest altruism $(= 1)$	Higher risk $(= 0)$ $\delta_0 \sim \texttt{Uniform[0,10\%]}$
Optimal share, mean	8.1%	18.4%	20.7%
Fraction with $s_t = 0$	33.1%	15.0%	12.8%
Fraction with $s_t \geq 1\%$	65.1%	84.2%	86.5%

Final Thoughts

Concluding Questions

- Case for investing 1/2% of GDP $\approx$ \$100b seems compelling
- How large is the catastrophic risk from A.I.?
  - How much are we currently spending to mitigate A.I. risk?
  - Effectiveness of mitigation spending?
    - Slow down and invest in safety research?
    - Focus on narrow A.I.? E.g. medical research
- How should we think about A.I. competition and race dynamics?
- How can we get A.I. labs to internalize the x-risk externalities?
  - Should we tax GPUs and use the revenue to fund safety research?