Artificial Intelligence and Economic Growth

P. Aghion, B. Jones, and C. Jones

in Agrawal et al *The Economics of Artificial Intelligence*, 2019
What are the implications of A.I. for economic growth?

• Build some growth models with A.I.
  o A.I. helps to make goods
  o A.I. helps to make ideas

• Implications
  o Long-run growth
  o Share of GDP paid to labor vs capital
  o Firms and organizations

• Singularity?
Two Main Themes

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

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

• Basic model: automating tasks in production

• A.I. and the production of new ideas

• Singularity?

• Some facts
The Zeira 1998 Model
Simple Model of Automation (Zeira 1998)

• Production uses \( n \) tasks/goods:

\[
Y = AX_1^{\alpha_1} X_2^{\alpha_2} \cdot \cdot \cdot X_n^{\alpha_n},
\]

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:**
  - \( \alpha \) reflects the *fraction* of tasks that are automated
  - Embed in neoclassical growth model \( \Rightarrow \)
    \[ g_y = \frac{g_A}{1 - \alpha} \text{ where } y_t \equiv \frac{Y_t}{L_t} \]

• **Automation:** \( \uparrow \alpha \) raises both capital share and LR growth
  - Hard to reconcile with 20th century
  - Substantial automation but stable growth and capital shares
Subsequent Work

  - Old tasks are gradually automated as new (labor) tasks are created
  - Fraction automated can then be steady
  - Rich framework, with endogenous innovation and automation, all cases worked out in great detail

- Peretto and Seater (2013), Hemous and Olson (2016), Agrawal, McHale, and Oettl (2017)
Automation and Baumol’s Cost Disease
Baumol’s Cost Disease and the Kaldor Facts

• Baumol: Agriculture and manufacturing have rapid growth and declining shares of GDP
  ○ ... but also rising automation

• Aggregate capital share could reflect a balance
  ○ Rises within agriculture and manufacturing
  ○ But falls as these sectors decline

• Maybe this is a general feature of the economy!
  ○ First agriculture, then manufacturing, then services
Final good

\[ Y_t = \left( \int_0^1 X_{it}^\sigma \frac{\sigma - 1}{\sigma} di \right)^{\sigma - 1} \text{ where } \sigma < 1 \]

Tasks

\[ X_{it} = \begin{cases} K_{it} & \text{if automated } i \in [0, \beta_t] \\ L_{it} & \text{if not automated } i \in [\beta_t, 1] \end{cases} \]

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)

\[ Y_t = Const_t + I_t \]

Allocation

\[ I_t = \bar{s}_K Y_t \]
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 \( \beta \) interacts with \( K \): two effects
  
  ○ \( \beta \): what fraction of tasks have been automated
  
  ○ \( \beta \): 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 $\beta$ terms into factor-augmenting form:

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

where

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

• Effect of automation: $\uparrow \beta_t \Rightarrow \downarrow B_t \quad \text{and} \quad \uparrow C_t$

*Intuition: dilution effects just get magnified since $\sigma < 1$*
• Suppose a constant fraction of non-automated tasks get automated every period:

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

\[ \Rightarrow \beta_t \to 1 \]

• 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

\[ Y_t = F(B_tK_t, C_tL_t) \text{ where } B_t = \left(\frac{1}{\beta_t}\right)^{\frac{1}{1-\sigma}} \text{ and } C_t = \left(\frac{1}{1 - \beta_t}\right)^{\frac{1}{1-\sigma}} \]

- \( \beta_t \to 1 \Rightarrow B_t \to 1 \)

- But \( C_t \) grows at a constant exponential rate!

\[ \frac{\dot{C}_t}{C_t} = -\frac{1}{1-\sigma} \frac{\dot{m}_t}{m_t} = \frac{\theta}{1-\sigma} \]

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

• Ratio of capital share to labor share:

\[
\frac{\alpha_K}{\alpha_L} = \left( \frac{\beta_t}{1 - \beta_t} \right)^{1/\sigma} \left( \frac{K_t}{L_t} \right)^{\frac{\sigma - 1}{\sigma}}
\]

• Two offsetting effects \((\sigma < 1)\):
  - \(\uparrow \beta_t\) raises the capital share
  - \(\uparrow K_t/L_t\) lowers the capital share

*These balance and deliver constant factor shares in the limit*

\[
\alpha_K \equiv \frac{F_K K}{Y} = \beta_t^{\frac{1}{\sigma}} \left( \frac{K_t}{Y_t} \right)^{\frac{\sigma - 1}{\sigma}} \rightarrow \left( \frac{\bar{s}_K}{g_Y + \delta} \right)^{\frac{\sigma - 1}{\sigma}} < 1
\]
Intuition for AJJ result

- Why does automation lead to balanced growth and satisfy Uzawa?
  - $\beta_t \to 1$ so the KATC piece “ends” eventually (all tasks automated)
  - 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

- Baumol logic
  - Agr/Mfg shrink as a share of the economy...
  - Labor still gets 2/3 of GDP! Vanishing share of tasks, but all else is cheap
    (Baumol)

*Interesting question: What fraction of tasks automated today? $\beta_{2022}$
  (B. Jones and X. Liu 2022 on capital-embodied technical change)*
Simulation: Capital Share and Automation Fraction

Fraction automated, $\beta_t$

Capital share $\alpha_K$

(also automated share of GDP)
Constant Factor Shares?

- Consider $g_A > 0$ — technical change beyond just automation
- Alternatively, factor shares can be constant if automation follows
  $$g_{\beta t} = (1 - \beta_t) \left( \frac{-\rho}{1 - \rho} \right) g_{kt},$$
- Knife-edge condition...
- Surprise: growth rates increase not decrease. Why? Requires
  $$g_{yt} = g_A + \beta_t g_{kt}.$$
- $g_A = 0$ means zero growth. $g_A > 0$ means growth rises
Simulation: Constant Capital Share

Fraction automated, $\beta_t$

Capital share $\alpha_K$
Simulation: Constant Capital Share

GROWTH RATE OF GDP

YEAR

0 50 100 150 200 250 300

2% 3% 4% 5%
Simulation: Switching regimes...
Simulation: Switching regimes...

Fraction automated, $\beta_t$

Capital share $\alpha_K$
A.I. and Ideas
AI in the Ideas Production Function

• Let production of goods and services be $Y_t = A_t L_t$

• Let idea production be:

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

• Assume fraction $\beta_t$ of tasks are automated by date $t$. Then:

$$\dot{A}_t = A_t^\phi F(B_t K_t, C_t S_t)$$

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...
**AI in the Ideas Production Function**

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

\[
F(B_t K_t, C_t S_t) = C_t S_t F \left( \frac{B_t K_t}{C_t S_t}, 1 \right) \to C_t S_t
\]

- 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 $(g_C)$
Can automation replace population growth?

- Maybe! Suppose $S$ is constant, $g_S = 0$
  - Intuition: Fixed $S$ is spread among exponentially-declining measure of tasks
  - So researchers per task is growing exponentially!

- However
  - This setup takes automation as exogenous and at “just the right rate”
  - What if automation is endogenized?
  - Is population growth required to drive automation?
  - Could a smart/growing AI entirely replace humans?
Singularities
Singularities

- Now we become more radical and consider what happens when we go “all the way” and allow AI to take over all tasks.

- **Example 1:** Complete automation of goods and services production.

\[ Y_t = A_t K_t \]

→ Then growth rate can accelerate exponentially

\[ g_Y = g_A + sA_t - \delta \]

we call this a “Type I” growth explosion
Singularities: Example 2

- Complete automation in ideas production function

\[ \dot{A}_t = K_t A_t^\phi \]

- Intuitively, this idea production function acts like

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

- Solution:

\[
A_t = \left( \frac{1}{A_0^{-\phi} - \phi t} \right)^{1/\phi}
\]

- Thus we can have a true singularity for \( \phi > 0 \). \( A_t \) exceeds any finite value before date \( t^* = \frac{1}{\phi A_0^\phi} \).
Singularities: Example 3 – Incomplete Automation

- Cobb-Douglas, $\alpha$ and $\beta$ are fraction automated, $S$ constant

\[ \dot{K}_t = \bar{s}L^{1-\alpha}A_t^\sigma K_t^\alpha - \delta K_t. \]

\[ \dot{A}_t = K_t^\beta S^\lambda A_t^\phi \]

- Standard endogenous growth requires $\gamma = 1$:

\[ \gamma := \frac{\sigma}{1 - \alpha} \cdot \frac{\beta}{1 - \phi}. \]

- If $\gamma > 1$, then growth explodes!
  - Can occur without full automation
  - Example: $\alpha = \beta = \phi = 1/2$ and $\sigma > 1/2$. 


Objections to singularities

1. Automation limits (no $\beta_t \to 1$)

2. Search limits

$$\dot{A}_t = A_t^{1+\phi} \text{ or even } A_t \leq \bar{A}$$

but $\phi < 0$ (e.g., fishing out, burden of knowledge...)

3. Natural Laws

$$Y_t = \left( \int_0^1 (a_{it} Y_{it}) \frac{\sigma - 1}{\sigma} \right)^{\frac{\sigma}{\sigma - 1}} \text{ where } \sigma < 1$$

now can have $a_{it} \to \infty$ for many tasks but no singularity

- *Baumol theme*: growth determined not by what we are good at, but by what is essential yet hard to improve
Final Thoughts
Conclusion: A.I. in the Production of Goods and Services

- Introduced Baumol's “cost disease” insight into Zeira’s model of automation
  - Automation can act like labor augmenting technology (surprise!)
  - Can get balanced growth with a constant capital share well below 100%, even with nearly full automation
Conclusion: A.I. in the Ideas Production Function

- Could A.I. obviate the role of population growth in generating exponential growth?

- Discussed possibility that A.I. could generate a singularity
  - Derived conditions under which the economy can achieve infinite income in finite time

- Discussed obstacles to such events
  - Automation limits, search limits, and/or natural laws (among others)
Extra Slides
Some Facts
Capital Shares in U.S. Industries

- Agriculture
- Oil/Gas Extraction
- Utilities
- Petroleum Mfg.
- Construction
Capital Shares in U.S. Industries

- Motor Vehicles
- Chemicals
- Plastics
- Computers
- Furniture
Capital Shares in U.S. Industries

[Graph showing capital shares for different industries over time, including Wholesale, Retail, Air Trans., Publishing, and Movies.]
Capital Shares in U.S. Industries

Telecommunications

Federal Govt

Health (hospitals)

Health (ambulatory)

Education
Adoption of Robots and Change in Capital Share

Motor Vehicles = 56% of robot investment in 2014
AI, Organizations, and Wage Inequality

- Usual story: robots replace low-skill labor, hence ↑ skill premium (e.g., Krusell et al. 2000)
- But solving future problems, incl. advancing AI, might be increasingly hard, suggesting ↑ complementarities across workers, ↑ teamwork, and changing firm boundaries (Garicano 2000, Jones 2009)
- Aghion et al. (2017) find evidence along these lines
  - outsource higher fraction of low-skill workers
  - pay *increased* premium to low-skill workers kept
AI, Organizations, and Wage Inequality

Log wage relative to labour market mean

In R&D intensity

kernel = epanechnikov, degree = 0, bandwidth = .31, pwidth = .47
AI, Skills, and Wage Inequality

![Graph showing the relationship between log wage relative to labor market mean and In R&D intensity for low, medium, and high skill levels.](image-url)
Examples of Data

- Google, Facebook
- Amazon
- Tesla, Uber, Waymo
- Medical and genetic data
- Location history
- Speech records
- Physical action data
What is Data in this Paper?

- Data as a factor of production

- Data improves the quality of a product
  - We do not model data as helping a consumer or firm make a more informed decision (e.g., consumption, pricing)

- Data can be useful even if anonymous

- Other aspects of the economics of data are interesting (price discrimination, product specialization, etc.), but are purposely left out of the model

Canonical example: data as input into machine learning algorithm. E.g., medical detection algorithms, self-driving cars, voice recognition software.
Policies on Data Are Being Written Now

What policies governing data use maximize welfare?

- European General Data Protection Regulation (GDPR)
  - Privacy vs. social gain from sharing
  - “The protection of natural persons in relation to the processing of personal data is a fundamental right”
  - “The right…must be considered in relation to its function in society…”

- The California Consumer Privacy Act of 2018 (start Jan 1 2020)
  - Allows consumers to opt out of having their data sold

- US Congress: COPRA, ACESS, etc.

- India’s Personal Data Protection bill
Data is Nonrival

- Growth literature: Ideas are nonrival
  - Unlike rival goods, ideas are infinitely usable

- Data is another nonrival good
  - Clearly not a blueprint / recipe \(\Rightarrow\) different from ideas
  - Ideas are production functions, data is a factor of production
  - Multiple engineers/algorithms can use same data at same time (within and across firms)

- Nonrivalry implies increasing returns to scale: \( Y = F(D, X) \)
  - Constant returns to rival inputs: \( F(D, \lambda X) = \lambda F(D, X) \)
  - Increasing returns to data and rival inputs: \( F(\lambda D, \lambda X) > \lambda F(D, X) \)
Data Property Rights Matter

• Key point: allocations with different degrees of data use
  ⇒ different output, welfare, etc.

• How do different property rights affect the use of data?
  o “Firms own data” versus “consumers own data”

• To illustrate, we assume (plausibly?) the Coase theorem fails
  o Consumers can’t commit to selling data to just one firm
  o Firms can’t commit to not using data they acquire
  o Useful for showing the role of data sharing
Data is Nonrival ⇒ Interesting Questions

• Do markets produce the right amount of data?

• Why don’t firms (always) sell their data?

• Who should own data as it’s created?

• Implications of data nonrivalry for antitrust, economic growth, and comparative advantage across countries?

*We develop a framework for thinking through these questions*
Outline

- Economic environment
- Allocations:
  - Optimal allocation
  - Firms own data
  - Consumers own data
  - Extreme privacy protection: outlaw data sharing
- Theory results and a numerical example
Basic Setup
Overview

- Representative consumer with a love for variety
- Innovation $\Rightarrow$ endogenous measure of varieties
- Nonrivalry of data $\Rightarrow$ increasing returns to scale
- How is data produced?
  - Learning by doing: each unit consumed $\rightarrow$ 1 unit of data
  - Alternative: separate PF (Tesla vs Google self-driving car)
- Any data equally useful in all firms $\Rightarrow$ one sector of economy
- Data depreciates fully each period
The Economic Environment

**Utility**
\[ \int_0^\infty e^{-\rho t} L_t \; u(c_t) \; dt \]

**Flow Utility**
\[ u(c_t) = \log c_t \]

**Consumption per person**
\[ c_t = \left( \int_0^{N_t} c_{it}^{\frac{\sigma-1}{\sigma}} \; di \right)^{\frac{\sigma}{\sigma-1}} \text{ with } \sigma > 1 \]

**Data production**
\[ J_{it} = c_{it} L_t \]

**Variety resource constraint**
\[ c_{it} = Y_{it} / L_t \]

**Firm production**
\[ Y_{it} = D_{it}^\eta L_{it}, \; \eta \in (0,1) \]

**Data used by firm \( i \)**
\[ D_{it} \leq \alpha x_{it} J_{it} + (1 - \alpha) B_t \; \text{(nonrivalry)} \]

**Data of firm \( i \) used by others**
\[ D_{sit} \leq \tilde{x}_{it} J_{it} \]

**Data bundle**
\[ B_t = \left( N_t^{-\frac{1}{\epsilon}} \int_0^{N_t} D_{sit}^{\frac{\epsilon-1}{\epsilon}} \; di \right)^{\frac{\epsilon}{\epsilon-1}} \text{ with } \epsilon > 1 \]

**Innovation (new varieties)**
\[ \dot{N}_t = \frac{1}{\chi} \cdot L_{et} \]

**Labor resource constraint**
\[ L_{et} + \int_0^{N_t} L_{it} \; di = L_t \]

**Population growth (exogenous)**
\[ L_t = L_0 e^{\gamma L t} \]

**Creative destruction**
\[ \delta(\tilde{x}_{it}) = \frac{\delta_0}{2} \tilde{x}_{it}^2 \; \text{(equilibrium)} \]
The Economic Environment: Simple Privacy Costs

Utility
\[ \int_{0}^{\infty} e^{-\rho t} L_t \ u(c_t, x_{it}, \tilde{x}_{it}) \, dt \]

Flow Utility
\[ u(c_t, x_{it}, \tilde{x}_{it}) = \log c_t - \frac{\kappa}{2} \frac{1}{N_t^2} \int_{0}^{N_t} x_{it}^2 \, di - \frac{\bar{\kappa}}{2} \frac{1}{N_t} \int_{0}^{N_t} \tilde{x}_{it}^2 \, di \]

Consumption per person
\[ c_t = \left( \int_{0}^{N_t} c_{it} \left( \frac{\sigma-1}{\sigma} \right) \, di \right)^{\frac{1}{\sigma}} \text{ with } \sigma > 1 \]

Data production
\[ J_{it} = c_{it} L_t \]

Variety resource constraint
\[ c_{it} = \frac{Y_{it}}{L_t} \]

Firm production
\[ Y_{it} = D_{it}^{\eta} L_{it}, \quad \eta \in (0, 1) \]

Data used by firm \( i \)
\[ D_{it} \leq \alpha x_{it} J_{it} + (1 - \alpha) B_t \text{ (nonrivalry)} \]

Data of firm \( i \) used by others
\[ D_{sit} \leq \tilde{x}_{it} J_{it} \]

Data bundle
\[ B_t = \left( N_t^{-\frac{1}{\epsilon}} \int_{0}^{N_t} D_{sit}^{\frac{\epsilon-1}{\epsilon}} \, di \right)^{\frac{\epsilon}{\epsilon-1}} \text{ with } \epsilon > 1 \]

Innovation (new varieties)
\[ \dot{N}_t = \frac{1}{\chi} \cdot L_{et} \]

Labor resource constraint
\[ L_{et} + \int_{0}^{N_t} L_{it} \, di = L_t \]

Population growth (exogenous)
\[ L_t = L_0 e^{g_{Lt}} \]

Creative destruction
\[ \delta(\tilde{x}_{it}) = \frac{\delta_0}{2} \tilde{x}_{it}^2 \text{ (equilibrium)} \]
The Planner Problem (using symmetry of firms)

$$\max_{\{L_{pt}, x_t, \tilde{x}_t\}} \int_0^\infty e^{-\tilde{\rho}t} L_0 \left( \log c_t - \frac{\kappa}{2} \frac{1}{N} x_t^2 - \frac{\tilde{\kappa}}{2} \tilde{x}_t^2 \right) dt, \quad \tilde{\rho} := \rho - g_L$$

subject to

\[ c_t = \frac{Y_t}{L_t} \]
\[ Y_t = N_t^{\frac{1}{\sigma - 1}} D_{it}^{\eta} L_{pt} \]
\[ D_{it} = \alpha x_t Y_{it} + (1 - \alpha) N_t \tilde{x}_t Y_{it} \]
\[ Y_{it} = D_{it}^{\eta} \cdot \frac{L_{pt}}{N_t} \]
\[ \dot{N}_t = \frac{1}{\chi} (L_t - L_{pt}) \]
\[ L_t = L_0 e^{g_L t} \]

• More sharing ⇒ negative utility cost but more consumption
• Balance labor across production and entry/innovation
Scale Effect from Sharing Data

\[ D_{it} = \alpha x_t J_{it} + (1 - \alpha) \left( N_t^{-\frac{1}{\epsilon}} \int_0^{N_t} (\tilde{x}_t J_{it}) \frac{e-1}{\epsilon} \, di \right)^{\frac{e}{e-1}} \]

\[ D_{it} = \alpha x_t Y_{it} + (1 - \alpha) N_t \tilde{x}_t Y_{it} \]

\[ = [\alpha x_t + (1 - \alpha) \tilde{x}_t N_t] Y_{it} \]

- No sharing versus sharing:
  - No sharing: Only the \( \alpha x_t \) term = no scale effect
  - Sharing: The \( (1 - \alpha) \tilde{x}_t N_t \) term = extra scale effect

Source of Scale Effect: \( N_t \) scales with \( L_t \)

- Plugging into production function:

\[ Y_{it} = ([\alpha x_t + (1 - \alpha) \tilde{x}_t N_t]^\eta L_{it})^{\frac{1}{1-\eta}} \]
The Optimal Allocation on BGP (asymptotic)

\[
\tilde{x}_{it} = \tilde{x}_{sp} = \left( \frac{1}{\tilde{\kappa}} \cdot \frac{\eta}{1 - \eta} \right)^{1/2}
\]

\[
x_{it} = x_{sp} = \frac{\alpha}{1 - \alpha} \tilde{\kappa} \left( \frac{1}{\tilde{\kappa}} \cdot \frac{\eta}{1 - \eta} \right)^{1/2}
\]

\[
L_{it}^{sp} = \chi \rho \cdot \frac{\sigma - 1}{1 - \eta} := \nu_{sp}
\]

\[
N_{t}^{sp} = \frac{L_{t}}{\chi (g_{L} + \nu_{sp})} := \psi_{sp}L_{t}
\]

\[
L_{pt}^{sp} = \nu_{sp} \psi_{sp}L_{t}
\]

\[
Y_{t}^{sp} = \left( \nu_{sp} (1 - \alpha)^{\eta} \tilde{x}_{sp}^{\eta} \right)^{1_{1-\eta}} \left( \psi_{sp}L_{t} \right)^{\frac{1}{\sigma-1} + \frac{1}{1-\eta}}
\]

\[
c_{t}^{sp} = \frac{Y_{t}}{L_{t}} = \left( \nu_{sp} (1 - \alpha)^{\eta} \tilde{x}_{sp}^{\eta} \right)^{1_{1-\eta}} \left( \psi_{sp}L_{t} \right)^{\frac{1}{\sigma-1} + \frac{\eta}{1-\eta}}
\]

\[
g_{c}^{sp} = \left( \frac{1}{\sigma - 1} + \frac{\eta}{1 - \eta} \right) g_{L}
\]

\[
D_{i}^{sp} = \left( (1 - \alpha) \tilde{x}_{sp} \nu_{sp} \psi_{sp}L_{t} \right)^{\frac{1}{1-\eta}}
\]

\[
D^{sp} = ND_{i} = \left( (1 - \alpha) \tilde{x}_{sp} \nu_{sp} \right)^{\frac{1}{1-\eta}} \left( \psi_{sp}L_{t} \right)^{1 + \frac{1}{1-\eta}}
\]

\[
Y_{it}^{sp} = \left( \nu_{sp} (1 - \alpha)^{\eta} \tilde{x}_{sp}^{\eta} \right)^{1_{1-\eta}} \left( \psi_{sp}L_{t} \right)^{\frac{\eta}{1-\eta}}
\]
The Optimal Allocation: GDP per person

\[ c_t^{sp} = \frac{Y_t}{L_t} = \left( \nu_{sp}(1 - \alpha)\eta \tilde{x}_{sp} \right)^{\frac{1}{1 - \eta}} \left( \psi_{sp}L_t \right)^{\frac{1}{\sigma - 1} + \frac{\eta}{1 - \eta}} \]

\[ g_c^{sp} = \left( \frac{1}{\sigma - 1} + \frac{\eta}{1 - \eta} \right) g_L \]

- Scale effect: \[ \frac{1}{\sigma - 1} + \frac{\eta}{1 - \eta} \]

  Love of Variety \quad Data

- More people make more data and all firms use all shared data
The Optimal Allocation: Data, Firm Size, Variety

\[ \tilde{x}_{sp} = \left( \frac{1}{\tilde{\kappa}} \cdot \frac{\eta}{1 - \eta} \right)^{1/2} \]

\[ L_{it}^{sp} = \chi \rho \cdot \frac{\sigma - 1}{1 - \eta} := \nu_{sp} \]

\[ N_{t}^{sp} = \frac{L_t}{\chi g L + \nu_{sp}} := \psi_{sp} L_t \]

- Data shared increasing in data production elasticity and decreasing in privacy cost
- Firm size constant on BGP. \( N \) has opposite comparative statics
- Higher entry cost, time preference, population growth, and elasticity of substitution raise firm size and reduce varieties
- Higher \( \eta \) raises firm size and reduces varieties: Entry does not create data
Firms Own Data
Firms Own Data: Consumer Problem

- Firms own data and choose one data policy \((x_{it}, \tilde{x}_{it})\) applied to all consumers

- Consumers just choose consumption:

\[
U_0 = \max_{\{c_{it}\}} \int_0^{\infty} e^{-\tilde{\rho}t} L_0 u(c_t, x_{it}, \tilde{x}_{it}) dt
\]

s.t. \quad c_t = \left( \int_0^{N_t} c_{it}^{\sigma-1} \frac{\sigma}{\sigma-1} di \right)^{\frac{\sigma}{\sigma-1}}

\[
\dot{a}_t = (r_t - g_L) a_t + \omega_t - \int_0^{N_t} p_{it} c_{it} di
\]
Firms own Data: Data Decisions

- Firms buy $D_{bit}$ data from intermediary at given price $p_b$

- Firms sell $D_{sit}$ data to intermediary at chosen price $p_{si}$
  - Perfect competition inconsistent with nonrival data!
  - Monopolistically competitive with own data
  - See the intermediary’s downward-sloping demand curve and set price

- How much data to use / sell?
  - $x_{it}$: Use all of own data $\Rightarrow x_{it} = 1$
  - $\tilde{x}_{it}$: Trade off = selling data versus creative destruction
    $\delta(\tilde{x}_{it}) =$ Poisson rate transferring ownership of variety
Firms own the Data: Incumbent Firm Problem

- Monopolistically competitive firm takes demand for variety as given (from FOC of consumer problem): \( p_{it} = \left( \frac{c_t}{c_{it}} \right)^{\frac{1}{\sigma}} = \left( \frac{Y_t}{Y_{it}} \right)^{\frac{1}{\sigma}} \)

\[
 r_t V_{it} = \max_{L_{it}, D_{bit}, x_{it}, \tilde{x}_{it}} \left( \frac{Y_t}{Y_{it}} \right)^{\frac{1}{\sigma}} Y_{it} - w_t L_{it} - p_{bt} D_{bit} + p_{st} \tilde{x}_{it} Y_{it} + \dot{V}_{it} - \delta(\tilde{x}_{it}) V_{it}
\]

s.t. \( Y_{it} = D_{it}^\eta L_{it} \)

\[
 D_{it} = \alpha x_{it} Y_{it} + (1 - \alpha) D_{bit}
\]

\( x_{it} \in [0, 1], \quad \tilde{x}_{it} \in [0, 1] \)

\[
 p_{sit} = \lambda D_{IN}^{-\frac{1}{\epsilon}} \left( \frac{B_t}{\tilde{x}_{it} Y_{it}} \right)^{\frac{1}{\epsilon}}
\]

- Data Intermediary \((p_{bt}, p_{st}, D_{bit})\) and Free Entry complete eqm.
Firms own the Data: Data Intermediary Problem

• A monopolist takes data purchase price as given and sees the downward sloping demand curve for data $p_{bt}(D_{bit})$:

$$\max_{p_{bt}}, D_{sit} \quad p_{bt} \int_0^{N_t} D_{bit} \, di - p_{st} \int_0^{N_t} D_{sit} \, di$$

s.t.

$$D_{bit} \leq B_t = \left( N_t^{-\frac{1}{\epsilon}} \int_0^{N_t} D_{sit}^{\frac{\epsilon-1}{\epsilon}} \, di \right)^{\frac{\epsilon}{\epsilon-1}}$$

$$p_{bt} \leq p_{bt}^*$$

• Free entry at zero cost $\Rightarrow$ zero profits

• Problem incorporates data nonrivalry
  ○ Buys data once from each firm
  ○ But can sell the same bundle multiple times
Entry: Innovation Creates a New Variety

- $\chi$ units of labor needed to create an additional variety

- Free entry condition:

$$ \chi w_t = V_{it} + \frac{\int_0^{N_t} \delta(\tilde{x}_{it}) V_{it} \, di}{\dot{N}_t} $$

- The value of a new variety and the per-entrant share of business stealing from creative destruction
Firms Own Data: A “No Trade” Law

• What if the government, in an attempt to protect consumers privacy, makes data sharing illegal?

• Government chooses
  ○ $x_{it} \in (0, 1]$
  ○ $\tilde{x}_{it} = 0$

• We call this the “Outlaw Sharing” allocation
Consumers Own Data
Consumers own Data: Consumer Problem

- Consumers own data, so now choose how much to sell \((x_{it}, \tilde{x}_{it})\):

\[
U_0 = \max_{\{c_{it}, x_{it}, \tilde{x}_{it}\}} \int_0^\infty e^{-\tilde{\rho}t} L_0 u(c_t, x_{it}, \tilde{x}_{it}) dt
\]

s.t. \[
c_t = \left( \int_0^{N_t} \frac{\sigma - 1}{\sigma} c_{it}^\sigma \, di \right)^{\frac{\sigma}{\sigma - 1}}
\]

\[
\dot{a}_t = (r_t - g_L)a_t + w_t - \int_0^{N_t} p_{it}\epsilon_{it} di + \int_0^{N_t} x_{it}p_{st}^a c_{it} di + \int_0^{N_t} \tilde{x}_{it}p_{st}^b c_{it} di
\]

- Firm problem similar to before, but now takes \(x, \tilde{x}\) as given, can’t sell data, and has to buy “own” data
Consumers own the Data: Incumbent Firm Problem

- Monopolistically competitive firm takes demand for variety as given (from FOC of consumer problem):
  \[ q_{it} = \left( \frac{c_t}{c_{it}} \right)^{\frac{1}{\sigma}} = \left( \frac{Y_t}{Y_{it}} \right)^{\frac{1}{\sigma}} = p_{it} - x_{it}p_{st}^a - \tilde{x}_{it}p_{st}^b \]

- Firm buys data on its own variety \((D_{ait})\) and data on other firms varieties \((D_{bit})\)

\[
\begin{align*}
  r_t V_{it} &= \max_{L_{it}, D_{ait}, D_{bit}} \left[ \left( \frac{Y_t}{Y_{it}} \right)^{\frac{1}{\sigma}} + p_{st}^a x_{it} + p_{st}^b \tilde{x}_{it} \right] Y_{it} - w_t L_{it} \\
  &\quad - p_{at} D_{ait} - p_{bt} D_{bit} + \dot{V}_{it} - \delta(\tilde{x}_t) V_{it} \\
\text{s.t.} \quad &Y_{it} = D_{it}^n L_{it} \\
& D_{it} = \alpha D_{ait} + (1 - \alpha) D_{bit} \\
& D_{ait} \geq 0, \quad D_{bit} \geq 0
\end{align*}
\]
Key Forces: Consumers vs. Firms vs. Outlaw Sharing

- Firms
  - use all data on own variety, ignoring consumer privacy
  - restrict data sharing because of creative destruction
- Consumers
  - respect their own privacy concerns
  - sell data broadly, ignoring creative destruction
- Outlaw sharing
  - maximizes privacy gains
  - missing scale effect reduces consumption
Results: Comparing Allocations

1. Planner Problem
2. Firms Own Data
3. Outlaw Data Sharing
4. Consumers Own Data
Key Allocations: \( alloc \in \{sp, f, c, ns\} \)

- **Firm size:** \( L_{i}^{\text{alloc}} = L_{pt}/N_{t} = \nu_{\text{alloc}} \)

\[
\begin{align*}
\nu_{sp} & := \chi \rho \cdot \frac{\sigma - 1}{1 - \eta} \\
\nu_{os} & := \chi \rho \cdot \frac{\sigma - 1}{1 - \sigma \eta} \\
\nu_{c} & := \chi \rho \cdot \frac{\rho + \delta(\Tilde{x}_{c})}{g_{L} + \delta(\Tilde{x}_{c})} \cdot \frac{\sigma - 1}{1 - \sigma \eta} \\
\nu_{f} & := \chi \rho \cdot \frac{\rho + \delta(\Tilde{x}_{f})}{g_{L} + \delta(\Tilde{x}_{f})} \cdot \frac{\sigma - 1}{1 - \sigma \eta} \frac{\epsilon - 1}{\epsilon}
\end{align*}
\]

- **Number of firms:** \( N_{t}^{\text{alloc}} = \psi_{\text{alloc}} L_{t} \)

\[
\psi_{\text{alloc}} := \frac{1}{\chi g_{L} + \nu_{\text{alloc}}}
\]
• For \( alloc \in \{sp, c, f\} \):

\[
Y_{t}^{alloc} = \left[ \nu_{alloc} (1 - \alpha) \eta \tilde{x}_{alloc} \right]^{\frac{1}{\eta - 1}} \left( \psi_{alloc} L_{t} \right)^{1 + \frac{1}{\sigma - 1} + \frac{\eta}{\eta - 1}}
\]

• For Outlaw Sharing:

\[
Y_{t}^{os} = \left[ \nu_{os} \alpha \eta x_{os} \right]^{\frac{1}{\eta - 1}} \left( \psi_{os} L_{t} \right)^{1 + \frac{1}{\sigma - 1}}
\]

• Two source of increasing returns to scale:
  
  o Standard variety effect: \( \frac{\sigma}{\sigma - 1} \)
  
  o Data sharing: \( \frac{\eta}{1 - \eta} \)

• Recall \( \tilde{x}_{t} > 0 \) from data sharing \( \Rightarrow \) scale effect
Data Sharing

<table>
<thead>
<tr>
<th>Own Firm Data</th>
<th>Sharing with Other Firms</th>
</tr>
</thead>
<tbody>
<tr>
<td>$x_{sp} = \frac{\alpha}{1-\alpha} \frac{\tilde{\kappa}}{\kappa} \left( \frac{1}{\tilde{\kappa}} \cdot \frac{\eta}{1-\eta} \right)^{1/2}$</td>
<td>$\tilde{x}_{sp} = \left( \frac{1}{\tilde{\kappa}} \cdot \frac{\eta}{1-\eta} \right)^{1/2}$</td>
</tr>
<tr>
<td>$x_f = 1$</td>
<td>$\tilde{x}_f = \left( \frac{2\Gamma \rho}{(2-\Gamma)\delta_0} \right)^{1/2}$, $\Gamma := \frac{\eta(\sigma-1)}{e^{-1} - \sigma\eta}$</td>
</tr>
<tr>
<td>$x_{os} \in (0, 1]$</td>
<td>$\tilde{x}_{os} = 0$</td>
</tr>
<tr>
<td>$x_c = \frac{\alpha}{1-\alpha} \frac{\tilde{\kappa}}{\kappa} \left( \frac{1}{\tilde{\kappa}} \cdot \frac{\eta}{1-\eta} \cdot \frac{\sigma-1}{\sigma} \right)^{1/2}$</td>
<td>$\tilde{x}_c = \left( \frac{1}{\tilde{\kappa}} \cdot \frac{\eta}{1-\eta} \cdot \frac{\sigma-1}{\sigma} \right)^{1/2}$</td>
</tr>
</tbody>
</table>

- Firms fear creative destruction and share less than planner ($\delta_0$)
- Consumers share less than planner because of mark up
- No sharing law restricts data even more
- Firms use more own-variety data compared to consumer/planner
**Numerical Example: Parameter Values**

<table>
<thead>
<tr>
<th>Description</th>
<th>Parameter</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Importance of data</td>
<td>$\eta$</td>
<td>0.06</td>
</tr>
<tr>
<td>Elasticity of substitution</td>
<td>$\sigma$</td>
<td>4</td>
</tr>
<tr>
<td>Weight on privacy</td>
<td>$\kappa = \tilde{\kappa}$</td>
<td>0.20</td>
</tr>
<tr>
<td>Population level</td>
<td>$L_0$</td>
<td>100</td>
</tr>
<tr>
<td>Population growth rate</td>
<td>$g_L$</td>
<td>0.02</td>
</tr>
<tr>
<td>Rate of time preference</td>
<td>$\rho$</td>
<td>0.025</td>
</tr>
<tr>
<td>Labor cost of entry</td>
<td>$\chi$</td>
<td>0.01</td>
</tr>
<tr>
<td>Creative destruction</td>
<td>$\delta_0$</td>
<td>0.4</td>
</tr>
<tr>
<td>Weight on own data</td>
<td>$\alpha$</td>
<td>1/2</td>
</tr>
<tr>
<td>Elasticity of Substitution (data)</td>
<td>$\epsilon$</td>
<td>50</td>
</tr>
<tr>
<td>Use of own data in NS</td>
<td>$\bar{x}$</td>
<td>1</td>
</tr>
</tbody>
</table>
Numerical Example: How large is $\eta$? (Approach 1 - Data Share)

- Share of GDP spent on data = $\frac{\eta (\sigma - 1)}{1 - \eta}$
- Similar formula/quantity when consumers or firms own data
- Set $\sigma = 4$
- If data share of GDP is 5% $\Rightarrow \eta = 0.0625$
- If data share of GDP is 10% $\Rightarrow \eta = 0.12$
- Approach will be to explore $\eta \in \{0.03, 0.06, 0.12\}$
Numerical Example: Consumption Equivalent Welfare

\[ U_{ss}^{alloc} = \frac{1}{\tilde{\rho}} \left( \log c_0^{alloc} - \frac{\tilde{\kappa}}{2} \tilde{x}_{alloc}^2 + \frac{g_c^{alloc}}{\tilde{\rho}} \right). \]

Let \( U_{ss}^{alloc}(\lambda) \) denote steady-state welfare when we perturb the allocation of consumption by some proportion \( \lambda \):

\[ U_{ss}^{alloc}(\lambda) = \frac{1}{\tilde{\rho}} \left( \log(\lambda c_0^{alloc}) - \frac{\tilde{\kappa}}{2} \tilde{x}_{alloc}^2 + \frac{g_c^{alloc}}{\tilde{\rho}} \right). \]

Define consumption equivalent welfare as \( \lambda^{alloc} \):

\[ U_{ss}^{sp}(\lambda^{alloc}) = U_{ss}^{alloc}(1) \text{ with } \]

\[ \log \lambda^{alloc} = \log c_0^{alloc} - \log c_0^{sp} - \frac{\tilde{\kappa}}{2} \left( \tilde{x}_{alloc}^2 - \tilde{x}_{sp}^2 \right) + \frac{g_c^{alloc} - g_c^{sp}}{\tilde{\rho}} \]

Level term Privacy term Growth term

Note: The \( x_{it} \) terms drop out because scaled by \( 1/N \)
Welfare Sensitivity Analysis ($\eta$, $\delta$, $\kappa$): $\lambda^c / \lambda^f$
### Allocations: Baseline

<table>
<thead>
<tr>
<th>Allocation</th>
<th>Data Sharing “own” $x$</th>
<th>Data Sharing “others” $\bar{x}$</th>
<th>Firm size $\nu$</th>
<th>Variety $N/L = \psi$</th>
<th>Consumption $c$</th>
<th>Growth $g$</th>
<th>Creative Destruct. $\delta$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Social Planner</td>
<td>0.66</td>
<td>0.66</td>
<td>1304</td>
<td>665</td>
<td>18.6</td>
<td>0.67%</td>
<td>0.0870</td>
</tr>
<tr>
<td>Consumers Own Data</td>
<td>0.59</td>
<td>0.59</td>
<td>1482</td>
<td>594</td>
<td>18.3</td>
<td>0.67%</td>
<td>0.0696</td>
</tr>
<tr>
<td>Firms Own Data</td>
<td>1</td>
<td>0.16</td>
<td>1838</td>
<td>491</td>
<td>16.0</td>
<td>0.67%</td>
<td>0.0052</td>
</tr>
<tr>
<td>Outlaw Sharing</td>
<td>1</td>
<td>0</td>
<td>2000</td>
<td>455</td>
<td>7.3</td>
<td>0.50%</td>
<td>0</td>
</tr>
</tbody>
</table>

- Firms overuse their own data and undershare with others
- Consumers share less data than planner, but not by much
- Growth rate scale effect is modest, level differences are large
## Consumption Equivalent Welfare

<table>
<thead>
<tr>
<th>Allocation</th>
<th>Welfare</th>
<th>Level term</th>
<th>Privacy term</th>
<th>Growth term</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$\lambda$</td>
<td>$\log \lambda$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Optimal Allocation</td>
<td>1</td>
<td>0</td>
<td>..</td>
<td>..</td>
</tr>
<tr>
<td>Consumers Own Data</td>
<td>0.9886</td>
<td>-0.0115</td>
<td>-0.0202</td>
<td>0.0087</td>
</tr>
<tr>
<td>Firms Own Data</td>
<td>0.8917</td>
<td>-0.1146</td>
<td>-0.1555</td>
<td>0.0409</td>
</tr>
<tr>
<td>Outlaw Sharing</td>
<td>0.3429</td>
<td>-1.0703</td>
<td>-0.9399</td>
<td>0.0435</td>
</tr>
</tbody>
</table>

- Outlaw sharing: particularly harmful law (66 percent worse!)
- Firms own data: substantially lower welfare (11 percent worse)
- Consumers own data: nearly optimal (1 or 2 percent worse)
Implications for IO

• Firms that use data might grow fast compared to those that don’t

• Firms would like to merge into one single economy-wide firm
  o Implications for antitrust
  o Price/quantity behavior

• What are the costs of forced sharing?
  o Disincentive to collect/create data
  o Data as a barrier to entry
    (extension to quality ladder model)
  o Markets unraveling

• Targeted mandatory sharing?
  o E.g., airplane safety (after a crash)
Data versus Ideas: Excludability

• Maybe technologically easier to transmit data than ideas (usb key vs. education) . . .

• But data can be encrypted and monitored

• Data seems highly excludable
  – Idea: use machine learning to train self-driving car algorithm
  – ML needs lots of data. Each firm gathering own data
The Boundaries of Data Diffusion: Firms and Countries

• How does data diffuse across firms and countries?
  o Ideas eventually diffuse across firms or countries, so no country scale effect (e.g. HK vs China)
  o What about data?

• Scale effects and country size
  o Larger countries may have an important advantage as data grows in importance

• Scale effects and institutions
  o What if China mandates data sharing across Chinese firms and U.S. has no such policy
  o What if consumers in China have different privacy concerns than in the U.S. or Europe?
Conclusion

• Nonrival data $\Rightarrow$ large social gain from sharing data

• If firms own data, they may:
  ◦ privately use more data than consumers/planner would
  ◦ sell less data across firms than consumers/planner would

• Nonrivalry $\Rightarrow$ Laws that outlaw sharing could be very harmful

• Consumers owning data good at balancing privacy and sharing