Nonrivalry and the Economics of Data

Chad Jones and Chris Tonetti

U.C. Berkeley
5 November 2018
Examples of Data

- Google, Facebook
- Amazon
- Tesla, Uber, Waymo
- Medical and genetic data
- Location history
- Speech records
- Physical action data

Canonical example: data as input into machine learning algorithm. E.g. self-driving car.
Data is Nonrival

- Data is infinitely usable
  - Contrast with rival goods: coffee, computer, doctor
  - Multiple engineers/algorithms can use same data at same time (within and across firms)

- Key ways that data enters the economy:
  - Nonrivalry $\Rightarrow$ social gain from sharing data
  - Privacy
    - Firm: competitive advantage ("moat")

- Social planner and consumers only care about the first two. But firms care a lot about the last one $\Rightarrow$ inefficiency
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 and be balanced against other fundamental rights. . . ”

- The California Consumer Privacy Act of 2018
  - Allows consumers to opt out of having their data sold
Nonrivalry of Data ⇒ Increasing Returns

- 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: $F(\lambda D, \lambda X) > \lambda F(D, X)$

- When firms hoard data, a firm learns only from its own consumers

- But when firms share data, all firms learn from all consumers
  - Firms, fearing creative destruction, will not do this
  - But if consumers own the data, they appropriately balance data sharing and privacy
Outline

• Economic environment

• Allocations:
  o Optimal allocation
  o Firms own data
  o Consumers own data
  o 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

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

Flow Utility

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

Consumption per person

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

Data production

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

Variety resource constraint

\[ c_{it} = \frac{Y_{it}}{L_t} \]

Firm production

\[ Y_{it} = D^\eta_{it} L_{it}, \quad \eta \in (0, 1) \]

Data used by firm \( i \)

\[ D_{it} \leq \alpha x_{it} J_{it} + (1 - \alpha) B_t \quad \text{(nonrivalry)} \]

Data of firm \( i \) used by others

\[ D_{sit} \leq \tilde{x}_{it} J_{it} \]

Data bundle

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

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

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

\[
\max_{\{L_{pt}, x_{it}, \tilde{x}_{it}\}} \int_0^\infty e^{-\tilde{\rho}t} L_0 u(c_t, x_{it}, \tilde{x}_{it}) \, dt, \quad \tilde{\rho} := \rho - g_L
\]

subject to

\[
c_t = \frac{Y_t}{L_t}
\]

\[
Y_t = N_t^{\sigma-1} D_{it}^\eta L_{pt}
\]

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

\[
Y_{it} = D_{it}^\eta \cdot \frac{L_{pt}}{N_t}
\]

\[
\dot{N}_t = \frac{1}{\chi} (L_t - L_{pt})
\]

\[
L_t = L_0 e^{g_L t}
\]

- More sharing \(\Rightarrow\) negative utility cost but more consumption
- Balance labor across production and entry/innovation
Scale Effect from Sharing Data

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

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

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

- 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:

\[ Y_{it} = \left( [\alpha x_t + (1 - \alpha) \tilde{x}_t N_t]^\eta L_{it} \right)^{\frac{1}{1-n}} \]
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:

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

subject to

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

\[
\dot{a}_t = (r_t - g_L) a_t + w_t - \int_0^{N_t} p_{it} c_{it} di
\]
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
    $\delta(\tilde{x}_{it}) = \text{Poisson rate transferring ownership of variety}$
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. \[ Y_{it} = D_{it}^\eta L_{it} \]

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

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

  \[ p_{sit} = \lambda_{DI} N_{t}^{-\frac{1}{\epsilon}} \left( \frac{B_t}{\tilde{x}_{it} Y_{it}} \right)^{\frac{1}{\epsilon}} \]

- 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 $p_{bt}(D_{bit})$:

$$\max_{p_{bt}, D_{sit}} p_{bt} \int_{0}^{N_t} D_{bit} \, di - p_{st} \int_{0}^{N_t} D_{sit} \, di$$

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:

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

- The value of a new variety and the per-entrant share of business stealing from creative destruction
Firms Own the Data: Definition of Equilibrium

The equilibrium in which firms own the data consists of quantities \( \{c_t, Y_t, c_{it}, x_{it}, \tilde{x}_{it}, a_t, Y_{it}, L_{it}, D_{it}, D_{bit}, B_t, D_{sit}, N_t, L_{pt}, L_{et}\} \) and prices \( \{p_{it}, p_{bt}, p_{sit}, w_t, r_t, V_{it}\} \) such that

1. \( \{c_t, c_{it}, a_t\} \) solve the Household Problem.
2. \( \{L_{it}, Y_{it}, p_{it}, p_{sit}, D_{bit}, D_{it}, x_{it}, \tilde{x}_{it}, V_{it}\} \) solve the Firm Problem.
3. \((D_{sit}, B_t)\) Data markets clear: \( D_{bit} = B_t \) and \( D_{sit} = \tilde{x}_{it} Y_{it} \)
4. \((p_{bt})\) Free entry into data intermediation gives zero profits there (constrains \( p_b \) as a function of \( p_s \))
5. \((L_{et})\) Free entry into producing a new variety leads to zero profits:
   \[ \chi w_t = V_{it} + \frac{\int_0^{N_t} \delta(\tilde{x}_{it}) V_{it} \, di}{N_t} \]
6. Definition of \( L_{pt} \): \( L_{pt} = \int_0^{N_t} L_{it} \, di \)
7. \( w_t \) clears the labor market: \( L_{pt} + L_{et} = L_t \)
8. \( r_t \) clears the asset market: \( a_t = \frac{\int_0^{N_t} V_{it} \, di}{L_t} \)
9. \( N_t \) follows its law of motion: \( \dot{N}_t = \frac{1}{\chi} (L_t - L_{pt}) \)
10. \( Y_t := c_t L_t \) denotes aggregate output.
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 share $(x_{it}, \tilde{x}_{it})$:

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

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

\[
\dot{a}_t = (r_t - g_L) a_t + \omega_t - \int_0^{N_t} p_{it} c_{it} dt + \int_0^{N_t} x_{it} p_{st} a_{it} dt + \int_0^{N_t} \tilde{x}_{it} p_{st} \tilde{c}_{it} dt
\]

- 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 \(D_{bit}\)

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

  s.t. \[
  Y_{it} = D_{it}^\eta L_{it} \\
  D_{it} = \alpha D_{ait} + (1 - \alpha) D_{bit} \\
  D_{ait} \geq 0, \quad D_{bit} \geq 0
  \]
Consumers own the Data: Data Intermediary Problem

- The DI chooses the price at which it sells a firm its own data and the price of other firms data, given its purchase price

\[
\max_{p_{ait}, p_{bit}, D^a_{cit}, D^b_{cit}} \int_0^{N_t} (p_{ait}D_{ait} + p_{bit}D_{bit})di - \int_0^{N_t} (p^{a}_{st}D^{a}_{cit} + p^{b}_{st}D^{b}_{cit})di
\]

s.t.

\[
D_{ait} \leq D^a_{cit} \quad \forall i
\]

\[
D_{bit} \leq B_t = \left( N_t^{-\frac{1}{\epsilon}} \int_0^{N_t} (D^b_{cit})^{\frac{\epsilon-1}{\epsilon}} di \right)^{\frac{\epsilon}{\epsilon-1}} \quad \forall i
\]

\[
p_{ait} \leq p^*_{ait} \quad \text{and} \quad p_{bit} \leq p^*_{bit}
\]

- Can not sell more data on firm \( i \) than it buys from consumers

- Can sell all data purchased as “type-b” data to each firm (nonrivalry)
Consumers own the Data: Equilibrium

An equilibrium in which consumers own data consists of quantities \( \{c_t, Y_t, c_{it}, x_{it}, \tilde{x}_{it}, a_t, Y_{it}, L_{it}, D_{ait}, D_{bit}, D_{cit}, D_{cit}, B_{it}, N_{it}, L_{pt}, L_{et}\} \) and prices \( \{q_{it}, p_{it}, p_{ait}, p_{bit}, p_{st}^a, p_{st}^b, w_t, r_t, V_{it}\} \) such that

1. \( \{c_t, c_{it}, x_{it}, \tilde{x}_{it}, a_t\} \) solve the Household Problem.
2. \( \{L_{it}, Y_{it}, p_{it}, D_{ait}, D_{bit}, D_{it}, V_{it}\} \) solve the Firm Problem.
3. \( (q_{it}) \) The effective consumer price is \( q_{it} = p_{it} - x_{it}p_{st}^a - \tilde{x}_{it}p_{st}^b \)
4. \( D_{cit}^a, D_{cit}^b, B_{it}, p_{ait}, \) and \( p_{bit} \) solve the Data Intermediary Problem (with zero profits).
5. \( p_{st}^a \) clears the data market: \( D_{cit}^a = x_{it}c_{it}L_t \).
6. \( p_{st}^b \) clears the data market: \( D_{cit}^b = \tilde{x}_{it}c_{it}L_t \).
7. \( (L_{et}) \) Free entry into new varieties leads to zero profits:
   \[ \chi w_t = V_{it} + \int_0^{N_t} \delta(\tilde{x}_{it})V_{it} \, di/N_t \]
8. Definition of \( L_{pt} \): \( L_{pt} = \int_0^{N_t} L_{it} \, di \)
9. \( w_t \) clears the labor market: \( L_{pt} + L_{et} = L_t \)
10. \( r_t \) clears the asset market: \( a_t = \int_0^{N_t} V_{it} \, di/L_t \)
11. \( N_t \) follows its law of motion: \( \dot{N}_t = \frac{1}{\chi}(L_t - L_{pt}) \)
12. \( Y_t := c_tL_t \) denotes aggregate GDP.
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: $\text{alloc} \in \{sp, f, c, ns\}$

- Firm size: $L_i^{\text{alloc}} = \frac{L_{pt}}{N_t} = \nu_{\text{alloc}}$

  $$
  \nu_{sp} := \chi \rho \cdot \frac{\sigma - 1}{1 - \eta}
  $$

  $$
  \nu_{os} := \chi \rho \cdot \frac{\sigma - 1}{1 - \sigma \eta}
  $$

  $$
  \nu_{c} := \chi g_L \cdot \frac{\rho + \delta(\tilde{x}_c)}{g_L + \delta(\tilde{x}_c)} \cdot \frac{\sigma - 1}{1 - \sigma \eta}
  $$

  $$
  \nu_{f} := \chi g_L \cdot \frac{\rho + \delta(\tilde{x}_f)}{g_L + \delta(\tilde{x}_f)} \cdot \frac{\sigma - 1}{1 - \sigma \eta^{\frac{\epsilon - 1}{\epsilon}}}
  $$

- Number of firms: $N_t^{\text{alloc}} = \psi_{\text{alloc}} L_t$

  $$
  \psi_{\text{alloc}} := \frac{1}{\chi g_L + \nu_{\text{alloc}}}
  $$
## Data Sharing

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

$$Y^\text{alloc}_t = [\nu_{\text{alloc}} (1 - \alpha) \eta \tilde{x}^\eta_{\text{alloc}}]^{\frac{1}{1-\eta}} (\psi_{\text{alloc}} L_t)^{1 + \frac{1}{\sigma - 1} + \frac{\eta}{1-\eta}}$$

For Outlaw Sharing:

$$Y^{os}_t = [\nu^{os} \alpha \eta \chi^{\alpha}_{os}]^{\frac{1}{1-\eta}} (\psi^{os} L_t)^{1 + \frac{1}{\sigma - 1}}$$

Two source of increasing returns to scale:

- Standard variety effect: $\frac{\sigma}{\sigma - 1}$
- Data sharing: $\frac{\eta}{1-\eta}$

Recall $\tilde{x}_t > 0$ from data sharing $\Rightarrow$ scale effect
Consumption per person and Growth

- Consumption per person:

  For $\text{alloc} \in \{sp, c, f\}$:  
  \[
  c_t^{\text{alloc}} = \text{Const}_{\text{alloc}} \cdot L_t^{\frac{1}{\sigma-1} + \frac{\eta}{1-\eta}}
  \]

  For outlaw sharing:  
  \[
  c_t^{\text{os}} = \text{Const}_{\text{os}} \cdot L_t^{\frac{1}{\sigma-1}}
  \]

- Per capita growth:

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

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

**Intuition:** No sharing means you learn from 10 workers (constant firm size), sharing means you learn from the entire population
Numerical Example: How large is $\eta$?

- Error rate is proportional to $M^{-\eta}$. Productivity = $1/(\text{error rate})$

- Average $\eta = 0.08$. Double data $\Rightarrow$ 6% reduction in error rate
## Numerical Example: Other Parameters

<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>0.08</td>
</tr>
<tr>
<td>Elasticity of substitution</td>
<td>( \sigma )</td>
<td>5</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>0.03</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>Use of own data in NS</td>
<td>( \bar{x} )</td>
<td>1</td>
</tr>
</tbody>
</table>
Allocations

<table>
<thead>
<tr>
<th>Allocation</th>
<th>Data Sharing</th>
<th>Firm size</th>
<th>Variety</th>
<th>Consumption</th>
<th>Growth</th>
<th>Creative Destruct.</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>“own” (x)</td>
<td>“others” (\tilde{x})</td>
<td>(\nu)</td>
<td>(N/L = \psi)</td>
<td>(c)</td>
<td>(g)</td>
</tr>
<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>
</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>
</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>
</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>
</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
  - Implications for antitrust
- Targeted mandatory sharing?
  - E.g., airplane safety (after a crash)
- What are the costs of forced sharing?
  - Disincentive to create data
    (in MS, data is aggregate, good approx?)
  - Data as a barrier to entry
    (extension to quality ladder model)
Data versus Ideas: Excludability

- Maybe technologically easier to transmit data than idea (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?
  - Ideas eventually diffuse across firms or countries, so no country scale effect (e.g. HK vs China)
  - What about data?
    - Intellectual property rights

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

- Scale effects and institutions
  - What if China mandates data sharing across Chinese firms and U.S. has no such policy
Conclusion

- Nonrival data ⇒ large social gain from sharing data

- If firms own data, they may:
  - privately use more data than consumers/planner would
  - share less data across firms than consumers/planner would

- Nonrivalry ⇒ Laws that outlaw sharing could be very harmful

- Consumers owning data good at balancing privacy and sharing