Nonrivalry and the Economics of Data

Presenter: Christopher Tonetti
based on research with Chad Jones

Federal Reserve Bank of Atlanta
Financial Markets Conference
20 April 2019
Examples of Data

• Google search history
• Amazon purchase history
• Tesla, Waymo car sensors
• Medical and genetic data
• Location history
• Speech records
• Physical action data
How is data used in the economy?

- People make choices in uncertain environments. Data helps to reduce uncertainty. Data informs models, formally or informally.

- Many modern goods and services have at their core algorithms that make choices

- Can think of data as a factor of production

- Data improves the quality or lowers the cost of a product
  - e.g., voice recognition software, self-driving cars, medical detection algorithms

- There are many factors of production (machines, buildings, labor, land, etc.). Why is data special?
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 ⇒ 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 ⇒ inefficiency
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
  - Allows consumers to opt out of having their data sold
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?
  - “Firms own data” versus “consumers own data”

- Our research builds a mathematical model with a market for buying and selling data

- We model data as being created as the byproduct of consumption

- We study the outcomes of the economy under different ownership regimes
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, may not do this enough
  
  ○ But if consumers own the data, they appropriately balance data sharing and privacy
Data is Nonrival ⇒ Interesting Questions

- Adam Smith’s invisible hand breaks down in environments with nonrival goods

- 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
The Economic Environment: Preferences and Technology

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

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

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

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

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

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

Creative destruction
\[ \delta(\tilde{x}_{it}) = \frac{\delta_0}{2} \tilde{x}_{it}^2 \quad \text{(equilibrium)} \]
The Benevolent Social Planner

- Imagine an all-powerful benevolent social planner who makes choices about the use of data

- Imagine the planner chooses which hospitals get to see which medical scans and biopsy results

- Why might the planner want each hospital to use data collected from patients at other hospitals?

- Why might the planner not make all medical data available to all hospitals?

- In a model, we can formalize the trade-off between privacy and improved quality of medical services
Firms Own the Data

- Imagine a world in which firms own data as it is created

- Let’s think about companies trying to develop self-driving car algorithms, e.g., Tesla and Waymo

- Why might Tesla want to buy data produced by Waymo cars?

- Why might Tesla sell data produced by people driving Teslas?
  - Note, Tesla would still have their data even after selling it because data is nonrival

- Why might Tesla not sell all their data to Waymo?

- What is the social cost of limited data-sharing across firms?
  - Imagine if every car manufacturer could produce with every factory (workers, robots, machines, etc.) simultaneously
Consumers Own the Data

- Imagine a world in which consumers own data as it is created
- Why might a Tesla owner want to sell data to Tesla?
- Why might a Tesla owner want to sell data to Waymo?
- Why wouldn’t a Tesla owner sell all their data to all firms?
- Asymmetry between how consumer thinks about selling data broadly and how firm thinks about it
  - I don’t care what is the name of the company that sells me a car, I care about the quality/price of the car
  - Firm owners do care if their company makes profits or if they go out of business
Summarizing Key Forces in Model

• Firms
  o use all data on own variety, ignoring consumer privacy
  o restrict data sharing because of creative destruction

• Consumers
  o respect their own privacy concerns
  o sell data broadly, ignoring creative destruction

• Outlaw sharing
  o maximizes privacy gains
  o missing scale effect reduces consumption
Quantitative Results: Many Open Questions

- We have a simple model designed to illustrate basic forces
- There are many difficulties in trying to quantify the welfare gains and losses associated with selling data across firms
  - How large are privacy costs? Utility costs per se, or concerns about firm behavior (prices and quantities)?
  - What are the returns to more data? Are we close to being saturated in data? How substitutable are different types/sources of data?
  - How concerned are firms about creative destruction due to leakage of data about their products?
  - How does the incentive to collect and create data change under different property-right regimes?
Implementation of Consumers own Data

- There are difficulties in understanding how to implement consumers owning data
  - Technologies, Legal frameworks, Market design

- Main takeaway is that there may be benefits to broadly using data across firms

- Broad use is technologically possible because data is nonrival

- Markets might not deliver optimal use of data without the right laws and institutions (especially an issue with nonrival goods)

- Counterpoint to the position that protecting privacy should be the single mandate for policy makers thinking about regulating data
Implications for Industrial Organization

- Firms that use data might grow fast compared to those that don’t
- Data-sharing within the firm is a force towards mergers
  - Implications for antitrust
  - Price/quantity behavior?
- Targeted mandatory sharing?
  - E.g., airplane safety (after a crash)
- What are the costs of prioritizing sharing?
  - Data as a barrier to entry
  - Markets unraveling?
  - Incentives to collect/create 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?

- 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 state-owned firms and the U.S. has no such policy or even outlaws selling data across firms
  - What if consumers in China have different privacy concerns than in the U.S. or Europe?
Conclusion

- Nonrival data $\Rightarrow$ large social gain from broad use of 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