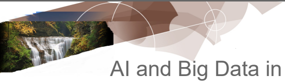


Antitrust, Regulation, and User Union in the Era of Digital Platforms and Big Data

Lin William Cong

Cornell SC Johnson College of Business | NBER

Simon Mayer

Chicago Booth | HEC Paris

June 2022

Rise of Digital Platforms and Big Data

- Digital economy, platforms, data explosion, etc.
  - Small number of large companies/platforms dominate and hold most of the world's data.
  - Privacy concerns, regulation of digital assets, anti-competitive concerns of platforms.
- Regulations: GDPR, open banking, etc.
- FinTech (r)evolution, decentralization, blockchains and Web3.
- Users acquire a new role in firm's production and service quality:
  - Network scale matters and userbase adds to network effects.
  - User data help platforms to develop products and improve (reducing search friction, facilitating decision-making, etc.).
- Antitrust and regulation in the presence of network effects and data feedback effects.

Data Economy and Data Sharing

- Network effects (e.g., Rochet & Tirole, 2003; switching cost in dynamic settings).
- Data reinforcement/feedback effect:
  - Platform price discrimination (e.g., Ichihashi, 2020).
  - Increasing versus decreasing return to scale on firm mark-up (e.g., Eeckout & Veldkamp, 2021).
- Over-supply and over-sharing:
  - Incentives, cognitive/attention frictions, informational asymmetry (e.g., Carrascal et al, 2013; Athey, Catalini, & Tucker, 2017; Arrieta-Ibarra et al., 2018; Agarwal et al., 2020).
  - Negative externalities of data contribution on others (e.g., Kirpalani & Philippon, 2020; MacCarthy, 2010;, Fairfield & Engel, 2015; Choi, Jeon, & Kim, 2019; Acemoglu et al., 2019).
  - Privacy concerns (Tang, 2020; Liu, Sockin, & Xiong, 2021).
- Under-supply and under-sharing:
  - Data feedback (Bergemann & Bonatti, 2015; Goldfarb & Tucker, 2022; Calvano & Polo, 2021; Veldkamp, 2020; Biglaiser, Calvano, and Cremer, 2019; Farboodi et al., 2020; Haigu & Wright, 2020; Prufer & Shottmuller, 2021).
  - Non-rivalry (e.g., Jones & Tonetti, 2020; Cong, Xie, & Zhang, 2021; Cong et al., 2021) and divisible.

Antitrust and User Union in the Digital Era

- The core features of data production and generation.
  1. Non-rivalry (horizontal, vertical, dynamic).
  2. By-product of economic activities (and non-separable from data subjects).
  3. Platforms also crucial in data production.
  4. Decentralized nature of data (Hayek, 1945): limited value if siloed; need for aggregation and exchange.

- This paper's insights and contribution:
  - Users (and platforms/intermediaries) do not internalize the impact of data generation and sharing on (i) other users, (ii) market power, and (iii) platforms' incentives.
  - Both over- and under- supply of data.
  - A unified framework to evaluate various policies.
  - User union and data trust as a radical solution to the antitrust challenge.

Outline

- Introduction
- **Model Setup**
- Equilibrium Characterization and Key Inefficiencies
- Policy Evaluations
- User Union
- Discussion and Conclusion

Model Setup

- Two periods, $t = 1, 2$; no discounting.
- Two platforms $x \in \{A, B\}$ compete for a unit measure of users uniformly distributed in location $z \in [0, 1]$.
- Production/service cost normalized to 0.
- Price p_t^x in the numeraire.
- Userbase N_t^x (level of adoption or usage).
- User utility: $Y_t^x - \kappa_t^x(z)$.
- Transport cost

$$\kappa_t^x(z) = \begin{cases} \hat{\kappa} z & \text{for} \quad x = A \\ \hat{\kappa}(1-z) & \text{for} \quad x = B, \end{cases} \tag{1}$$

- Service quality: $Y_t^x = K^x + \phi^x D^x \ \mathbb{I}_{\{t=2\}} + \gamma^x N_t^x$.
- Heterogeneous data processing power; WLOG, $\phi^A \geq \phi^B$.

Data Generation and Collection

- Data as by-product of economic activities; $e^x\theta^x(z)$ units of data generated and collected.
- Platform's investment/effort e_t^x with quadratic (and private) cost $\frac{1}{2}\lambda(e_t^x)^2$.
- Fraction of data endogenously shared by user z: $\theta_t^x(z) \in [\underline{\theta}, 1]$.
- Data compensation and privacy cost: $(q^x - c^x)\theta^x e^x$ and $c^x = c(1+\eta^x)$, where η^x is fraction shared with $-x$.
- User's net utility:

$$u_t^x(z) = Y_t^x - p_t^x - \kappa_t^x(z) + (q^x - c^x)e^x\theta^x \;\; \mathbf{1}_{\{t=1\}}.$$

- $\hat{D}^x$: data generated in $t=1$; $D \equiv \hat{D}^A + \hat{D}^B$ total data for $t=2$.
  - $D^x \leq D$ (it needs not hold that $D_t^A + D_t^B = D_t$). .
  - $D = N_1^A e_1^A \theta_1^A + N_1^B e_1^B \theta_1^B$.

Timeline, Objective, & Equilibrium

- Timeline of the game:
  - $t = 1$: Platforms simultaneously set (q^x, e^x, p_1^x), followed by users' choices of platforms (no multi-homing).
  - $t = 2$: Platforms simultaneously set p_2^x, followed by users' choices of platforms (no sticky assumption).
- Equilibrium concept: Subgame perfect in pure strategies.
- Objectives:

$$\pi_2^x := N_2^x p_2^x.$$

$$\pi_1^x := N_1^x p_1^x - q^x N_1^x e^x \theta_1^x + N_2^x p_2^x - \frac{\lambda (e^x)^2}{2}.$$

Outline

- Introduction
- Model Setup
- **Equilibrium Characterization and Key Inefficiencies**
- Policy Evaluations
- User Union
- Discussion and Conclusion

The Marginal User

- Indifference condition:

$$Y_t^A - p_t^A - \kappa_t^A(\hat{z}_t) + e^A\theta^A(q^A - c^A)\mathbb{I}_{\{t=1\}} \quad (2)$$
$$= \quad Y_t^B - p_t^B - \kappa_t^B(\hat{z}_t) + e^B\theta^B(q^B - c^B)\mathbb{I}_{\{t=1\}}.$$

$$\hat{z}_t = \quad \frac{1}{2} + \frac{1}{2\kappa}\left[\Delta_K - (p_t^A - p_t^B) + \left[\phi^A D^A - \phi^B D^B\right]\mathbb{I}_{\{t=2\}} + \left[e^A\theta^A(q^A - c^A) - e^B\theta^B(q^B - c^B)\right]\mathbb{I}_{\{t=1\}}\right.$$

(3)

where $\kappa := \hat{\kappa} - \frac{\gamma^A+\gamma^B}{2}$ (transport costs mitigated by average network externality) and $\Delta_K := K^A - K^B + \frac{\gamma^A-\gamma^B}{2}$ (non-data-related platform heterogeneity).

Unique Nash equilibrium in period $t = 2$ subgame

- FOC: $\frac{\partial \pi_2^x}{\partial p_2^x} = N_2^x + \frac{\partial N_2^x}{\partial p_2^x} = 0.$
- $t = 2$ long-term price:

$$p_2^A = \bar{p}^A + \frac{D^A\phi^A - D^B\phi^B}{3} \quad \text{and} \quad p_2^B = \bar{p}^B - \frac{D^A\phi^A - D^B\phi^B}{3},$$

$$\bar{p}^A = \kappa + \frac{\Delta_K}{3} \quad \text{and} \quad \bar{p}^B = \kappa - \frac{\Delta_K}{3} \tag{4}$$

- The marginal user $\hat{z}_2$ is characterized by

$$\hat{z}_2 = \frac{1}{2} + \frac{\Delta_K + D^A\phi^A - D^B\phi^B}{6\kappa}.$$

- Market share $\hat{z}_2$ captures A's market and price-setting power.

Platform's Optimal Pricing

First-stage optimization:

$$\pi_1^x := \max_{q^x, p_1^x, e_1^x} \left(N_1^x p_1^x + \max_{p_2^x} N_2^x (p_2^{xx} - N_1^x \theta^x e^x q^x - \frac{\lambda (e^x)^2}{2} \right). \quad (5)$$

Platform's initial pricing of service:

$$\frac{\partial \pi_1^x}{\partial p_1^x} = \underbrace{N_1^x + \left(\frac{\partial N_1^x}{\partial p_1^x}\right) p_1^x}_{\text{Static revenue maximization}} + \underbrace{\sum_{x'=A,B} \left(\frac{\partial \pi_2^x}{\partial D^{x'}} \frac{\partial D^{x'}}{\partial N_1^{x'}} \frac{\partial N_1^{x'}}{\partial p_1^x} \right)}_{\text{Data benefits}} \quad (6)$$

$$- \underbrace{\left(\frac{\partial N_1^x}{\partial p_1^x}\right) e^x \theta^x q^x}_{\text{Data cost}} = 0. \quad (7)$$

Platform's Optimal Effort and Data Compensation

Platform's effort/investment:

$$\frac{\partial \pi_1^x}{\partial e^x} = \underbrace{\left(\frac{\partial N_1^x}{\partial e^x}\right) p_1^x}_{\text{Static revenue maximization}} + \underbrace{\sum_{x'=A,B} \left(\frac{\partial \pi_2^x}{\partial D^{x'}} \left[\frac{\partial D^{x'}}{\partial e^x} + \frac{\partial D^{x'}}{\partial N_1^{x'}} \frac{\partial N_1^{x'}}{\partial e^x}\right]\right)}_{\text{Data benefits}}$$

$$- \underbrace{\left(\left(\frac{\partial N_1^x}{\partial e^x}\right) e^x \theta^x q^x + \theta^x N_1^x q^x + \lambda e^x\right)}_{\text{Data cost}} = 0. \quad (8)$$

Compensation for data:

$$q^x = \begin{cases} 0 & \text{if} \quad \underline{\theta} = 1 \\ c^x & \text{if} \quad \underline{\theta} \in [0, 1)]. \end{cases} \quad (9)$$

Cornell University

Baseline: Database Directives

- Platforms own data, $\underline{\theta} = \eta^x = 1$ and there is no data sharing by platforms $\eta^x = 0$.
- $p_1^x = p_1^{x,Base} + q^x e^x$: platforms pass the costs of collecting data one-to-one onto users; one can set $q^x = 0$.
- $p_1^A < p_1^B$; platforms compete fiercely and price below benchmark $\bar{p}^x$; users benefit.
- $p_2^A > p_2^B$ and $p_1^x < p_2^x$; users hurt in $t = 2$ even without assumptions on stickiness.
- $N_t^A = \hat{z}_t$ increases in:
  1. A's data processing capacity: $\hat{\phi}^A - \hat{\phi}^B$.
  2. Importance of data: $\hat{\phi}^A + \hat{\phi}^B$.
  3. Network effects: $\gamma^A + \gamma^B$.

Two Inefficiencies/Challenges in the Era of Platforms and Big Data

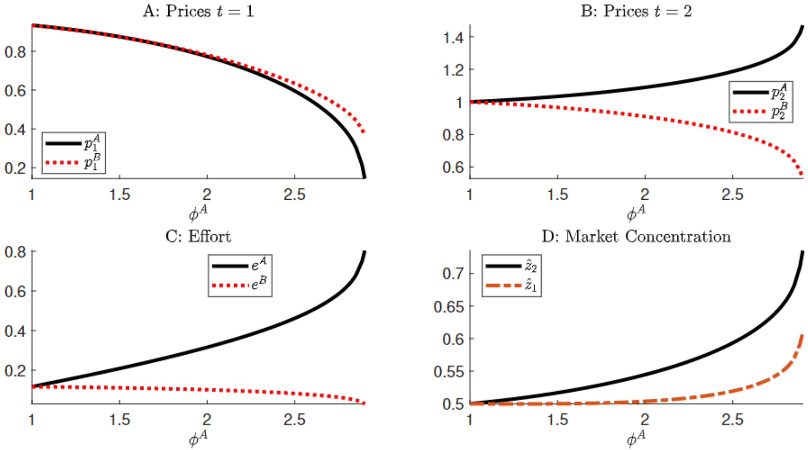
- Market power and “winner takes all” due to “data feedback effect,” not (directly) network effects.

Two Inefficiencies/Challenges in the Era of Platforms and Big Data

- Market power and "winner takes all" due to "data feedback effect," not (directly) network effects.
- Symmetric equilibrium for symmetric platforms:
  $p_1^x = \kappa + e^x \left(q^x - \frac{2\phi^x}{3}\right)$ and $e^x = \min\left\{\left[\frac{\phi^x - 3c^x}{6\lambda}\right]^+, 1\right\}$.
- Effort decreasing in privacy cost as users require direct compensation or lower price.
- Individuals fail to internalize the externality of data contribution:
  1. on service quality improvement for other users.
  2. on platform's market power.
  3. on the incentives to invest in infrastructure and scenarios for generating data.

Illustration of the Two Inefficiencies

Prices, Effort, and Market Shares under $\Delta_K = 0$ and $\phi^B = 1$.



Outline

- Introduction
- Model Setup
- Equilibrium Characterization and Key Inefficiencies
- **Policy Evaluations**
- User Union
- Discussion and Conclusion

Data Privacy Protections

- GDPR (2018), CCPA (2020), China PIPL (2021), CPA (Colorado), UCPA (Utah), etc.
- Data access, control, and portability.
- Opt in options for users: i.e., $\underline{\theta} = 0$.
- But data compensation is passed onto consumers one-for-one.
- Barely giving data ownership to users has no material effects within our framework.

Open Data Initiatives

- Open Banking: European vs. U.S. approaches.
- Two potential benefits: (i) it increases any platform's stock of data and thus overall service quality and (ii) it may reduce market power.
- An illustration: $D^x = e^A N_1^A + e^B N_1^B$; $\eta = 1$ without monetary transfer.
- Holding effort fixed at $e^A = e^B = 1$, data sharing increases platform A's market share $\hat{z}_2$ relative to the baseline iff Δ_K is small and $\phi^A + \phi^B < 3\sqrt{6}\kappa$.
- Open data initiatives may undermine (ii).
- Open data hold up platform's effort and increases privacy costs.

Cornell University

Data Market with User Ownership

- Unbundling platform service consumption and data contribution.
- Platform x' posts a per-unit price for data and the atomistic users take this price as given and decide on whether they sell their data to x'. Platform x' optimally offers user z per unit price of data of c dollars (via, e.g., Nash Bargaining).
- Data market may hurt consumer welfare and increase platform A's market share if two platforms are close.
- Undermines (i) platforms' incentives to collect data and to attract users via low prices, (ii) platform's effort, and (iii) may increase or decrease market power by large platforms.

Data Market with Platform Ownership

- Platforms determine the optimal allocation of data and the transfer via Nash Bargaining or competitive data intermediaries.
- Platform collusion via data trading (when effort is fixed), i.e., “data collusion” or “data merger.” Weakly strong platform possesses all data.
- When $\hat{D}^A \phi^A + \Delta_K \geq \hat{D}^B \phi^B$, joint payoff $\pi_2^A + \pi_2^B$ maximized with $D^A = \hat{D}^A + \hat{D}^B$ and $D^B = \hat{D}^B$.
- Symmetric platforms:
    1. One platform gets all data and market is concentrated: $\hat{z}_2 = \frac{1}{2} + \frac{\theta^X e^X \phi^X}{12\kappa}$.
    2. high-powered incentives to generate data (lower prices and higher efforts).
- Platforms are not fully competitive.
- Data intermediary does not ensure ex-ante platform incentives.

Outline

- Introduction
- Model Setup
- Equilibrium Characterization and Key Inefficiencies
- Policy Evaluations
- **User Union**
- Discussion and Conclusion

Introducing User Union

- Investor protections through SECs, labor unions, etc.
- Users are contributors but are atomistic and decentralized.
- User union as a coordination device.
- They may under-supply data, even though data in aggregate is beneficial.
- They may over-supply to aggravate market power or disincentivize platforms to generate more data.
- User union addressing both over- and under-sharing.
- Data governance structures in practice (Houser and Bagby, 2022): Data Trust, etc.

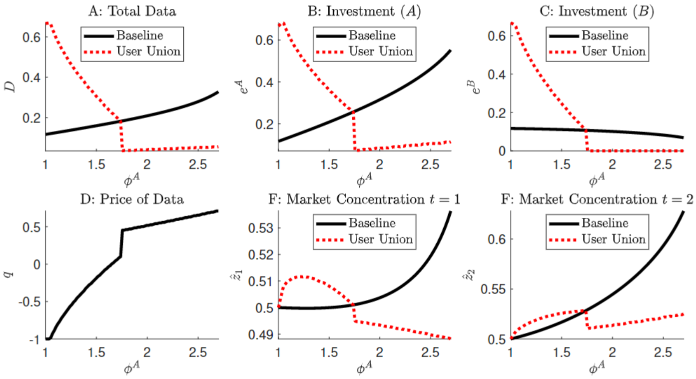
Implementation of User Union

- UU pre-commits to reward f per unit of data; negative f means users pay a union fee.
- Implementation 1: UU collects all data; platforms can purchase their own data pro-rata payouts to users.
- Platform solves: $\max_{D^x \in [0, N_1^x e^x]} \pi_2^x - q^x D^x$.
- UU sets $\theta^x = 1$ and solves: $\max_{q^A, q^B} \left[u_1 + N_1^A e^A (q^A - c^A) + N_2^B e^B (q^B - c^B)\right] \quad \text{s.t.} \quad \pi_1^x \geq 0.$
- Alternative implementation: UU pre-commits to f dollar per unit of data contribution; negative f means users pay a union fee.
- UU solves: $\max_f (u_1 - f\theta^x e^x)$ s.t. $\pi_1^x \geq 0$.

Cornell University

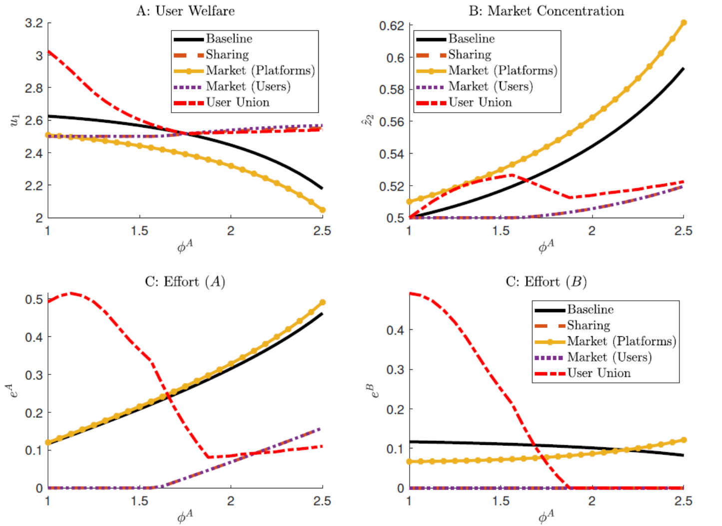
User Union as an Intuitive Solution

Platform asymmetry and equilibrium outcomes.



User Union as an Intuitive Solution

Comparison of various antitrust policies and designs.



Cornell University

Discussion and Takeaways

- Network effects, interoperability, and data heterogeneity.
- Blockchains, smart contracts, and tokenization as commitments to future pricing.
  - e.g., Cong, et al. (2020); Goldstein, Gupta, and Sverchkov, (2021); Lyandres, (2022).
  - Categories and functions of crypto-tokens.
  - Monetary policy for tokenized platforms.
  - User union implemented through DataDAO.
- Antitrust and regulation in the digital era:
  - Users contribute via network effects and data reinforcement effects.
  - Data and data subjects are atomistic and dispersed (Hayek's decentralization).
  - Giving ownership and control to platforms means market power.
  - Individual users do not internalize the effect of data contribution on (i) other users, (ii) market power of platforms, and (iii) platforms' incentives.

Discussion and Takeaways

- Network effects, interoperability, and data heterogeneity.
- Blockchains, smart contracts, and tokenization as commitments to future pricing.
  - e.g., Cong, et al. (2020); Goldstein, Gupta, and Sverchkov (2021); Lyandres (2022).
  - Categories and functions of crypto-tokens.
  - Monetary policy for tokenized platforms.
  - User union implemented through DataDAO.
- Antitrust and regulation in the digital era:
  - Users contribute via network effects and data reinforcement effects.
  - Data and data subjects are atomistic and dispersed (Hayek's decentralization).
  - Giving ownership and control to platforms means market power.
  - Individual users do not internalize the effect of data contribution on (i) other users, (ii) market power of platforms, and (iii) platforms' incentives.
- Data privacy protection, open data initiatives, splits and interoperability, data markets, etc., may all have their own issues.
- User union as an intuitive yet effective solution:
  - Data trust or government-run data union.
  - Relevance for network effects too.