

The Value of Data

Theory and Evidence

Siavash Mohades^{1,2} & Maria Savona^{3,4}

¹Maastricht University, ²Ellison Institute of Technology, ³Luiss Guido Carli, ⁴ University of Sussex

February 5, 2026

PRIN Final Workshop
Luiss Guido Carli

① Paper 1: Uncertainty, Data and R&D Investments

② Paper 2: Data Rent and Surplus Value in the Digital Economy

- Adoption of digital technologies and increased use of AI due to Industry 4.0 -
Both are data intensive

- Adoption of digital technologies and increased use of AI due to Industry 4.0 -
Both are data intensive
- Data collection at scale → information → more knowledge stock (Mansell and Steinmueller, 2020), (Brynjolfsson, Jin and Wang, 2023; Ennen and Richter, 2010) → competitive advantage and growth (Eeckhout and Veldkamp, 2022)

- Adoption of digital technologies and increased use of AI due to Industry 4.0 -
Both are data intensive
- Data collection at scale → information → more knowledge stock (Mansell and Steinmueller, 2020), (Brynjolfsson, Jin and Wang, 2023; Ennen and Richter, 2010) → competitive advantage and growth (Eeckhout and Veldkamp, 2022)
- Used for improving sales and demand forecasting (Hallikainen, Savimäki and, Laukkanen, 2020; Hofmann and Rutschmann, 2018), productivity (Müller, Fay and vom Brocke, 2018)

- Adoption of digital technologies and increased use of AI due to Industry 4.0 -
Both are data intensive
- Data collection at scale → information → more knowledge stock (Mansell and Steinmueller, 2020), (Brynjolfsson, Jin and Wang, 2023; Ennen and Richter, 2010) → competitive advantage and growth (Eeckhout and Veldkamp, 2022)
- Used for improving sales and demand forecasting (Hallikainen, Savimäki and, Laukkanen, 2020; Hofmann and Rutschmann, 2018), productivity (Müller, Fay and vom Brocke, 2018)
- Data → information asymmetries ↓ → operational and market uncertainty and risks ↓ → potential to afford higher 'intrinsic' uncertainty from R&D? → complementarity of investments in data & R&D ↑ → productivity gains ↑

Research Questions

- 1 Does investing in data affect R&D investments?

Research Questions

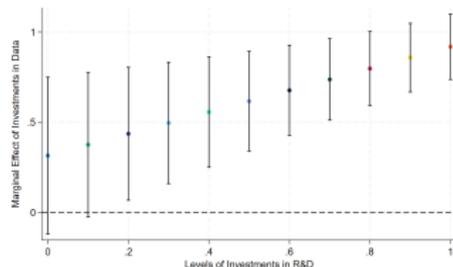
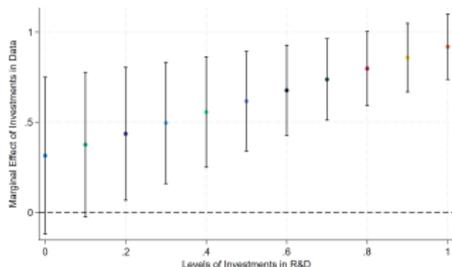
- 1 Does investing in data affect R&D investments?
- 2 Is there a multiplicative effect on labour productivity of the joint investments in data and R&D, with respect to sole investments in data, sole investments in R&D?

- Investing in data reduces (demand) uncertainty,

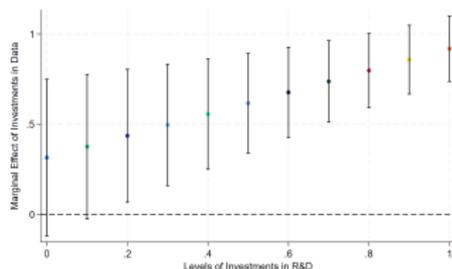
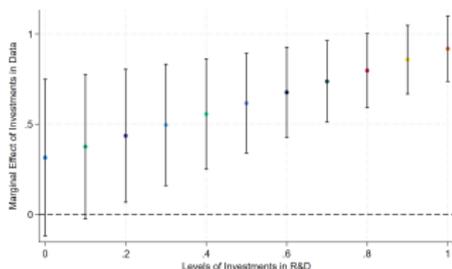
- Investing in data reduces (demand) uncertainty, freeing up room for the firm to become more willing to take risk on innovation.

Results

- Investing in data reduces (demand) uncertainty, freeing up room for the firm to become more willing to take risk on innovation. Data also makes firms more productive!
- The effect of data on both innovation and productivity is not uniform across firms



- Investing in data reduces (demand) uncertainty, freeing up room for the firm to become more willing to take risk on innovation. Data also makes firms more productive!
- The effect of data on both innovation and productivity is not uniform across firms



- Status: Awaiting decision at the Journal of Industrial Economics

① Paper 1: Uncertainty, Data and R&D Investments

② **Paper 2: Data Rent and Surplus Value in the Digital Economy**

The Data Paradox

People always asked us “What data are you talking about?”

The Data Paradox

People always asked us “What data are you talking about?” “Who owns it?”

The Data Paradox

People always asked us “What data are you talking about?” “Who owns it?”

Conventional View: Data is a symmetric production factor receiving its marginal product

The Data Paradox

People always asked us “What data are you talking about?” “Who owns it?”

Conventional View: Data is a symmetric production factor receiving its marginal product

Some Important Empirical Realities:

- Higher markups in the intangibilisation era: (De Loecker et al. 2020), Rising concentration in data-intensive sectors, Persistent monopoly profits (Google, Meta, Amazon)
- Falling profit rates in the last decades
- GDPR *increased* concentration by 17% (Johnson et al. 2023)

The Data Paradox

People always asked us “What data are you talking about?” “Who owns it?”

Conventional View: Data is a symmetric production factor receiving its marginal product

Some Important Empirical Realities:

- Higher markups in the intangibilisation era: (De Loecker et al. 2020), Rising concentration in data-intensive sectors, Persistent monopoly profits (Google, Meta, Amazon)
- Falling profit rates in the last decades
- GDPR *increased* concentration by 17% (Johnson et al. 2023)

Key Question: Who is the winner in the age of datafication? How is the value of data distributed?

Two Channels of Value Surplus Creation: Productivity vs. Rent

Data operates through two analytically distinct channels:

1. Productivity Channel (Value Creation)

Better information → efficiency, matching, prediction
Would persist under any ownership arrangement

2. Rent Channel (Value Appropriation)

Exclusive control → monopoly extraction
Depends critically on private ownership structures

Marxian Value Theory

Total value created in production:

$$W = c + v + s$$

where:

- c = constant capital (means of production)
- v = variable capital (wages)
- s = surplus value (appropriated by capitalists)

Rate of exploitation: $\sigma = s/v$

Rate of profit: $r = \frac{s}{c+v} = \frac{\sigma}{q+1}$ where $q = c/v$

Surplus distribution:

$$s = \pi_I + \pi_F + \pi_C + R_L + \boxed{R_D}$$

Data as Exploitation-Intensifier

Data increases the rate of surplus value extraction:

$$\sigma(D) = \sigma_0 + \beta D^\gamma \quad (0 < \gamma < 1)$$

Mechanisms:

- 1 Surveillance → real-time monitoring
- 2 Algorithmic management → intensified labour
- 3 Speedup of circulation (logistics, inventory)
- 4 Reduction of necessary labour (automation)
- 5 Monopolistic appropriation (price discrimination)

Data enables extracting more surplus per labour-hour

The Rate of Profit with Data

Let $\theta_R \in [0, 1]$ = fraction of data-enabled surplus appropriated as rent

Profit rate:

$$r = \frac{\sigma_0 + \beta D^\gamma (1 - \theta_R)}{q + 1}$$

Dual effect of data:

- **Positive:** Data increases $\sigma(D) \rightarrow$ raises surplus
- **Negative:** Rent appropriates fraction $\theta_R \rightarrow$ reduces profit for accumulation

When $\theta_R = 0$: data unambiguously raises r

When $\theta_R = 1$: all data-enabled surplus extracted as rent

Data Sharing as Counteracting Tendency

Introduce sharing parameter $\psi \in [0, 1]$:

- $\psi = 0$: Full monopolization
- $\psi = 1$: Complete data commons

As $\psi \rightarrow 1$, differential rent vanishes:

$$R_D(\psi) \rightarrow 0 \quad \text{while} \quad \sigma(D^{\text{total}}) \text{ preserved}$$

If sharing follows some conditions (not unconventional assumptions discussed in the paper...), We show that:

$$r(\psi = 1) > r(\psi = 0)$$

Privacy Protection and Data Sharing

Extend model with privacy parameter $\rho \in [0, 1]$:

$$\sigma(D, \rho, \psi) = \sigma_0 + \beta D^\gamma (1 - \rho)^\theta (1 + \eta\psi)$$

$$R_D(\rho, \psi) = \beta D^\gamma V (1 - \rho)^\theta (1 + \eta\psi) (1 - \psi)^\lambda$$

Comparative statics:

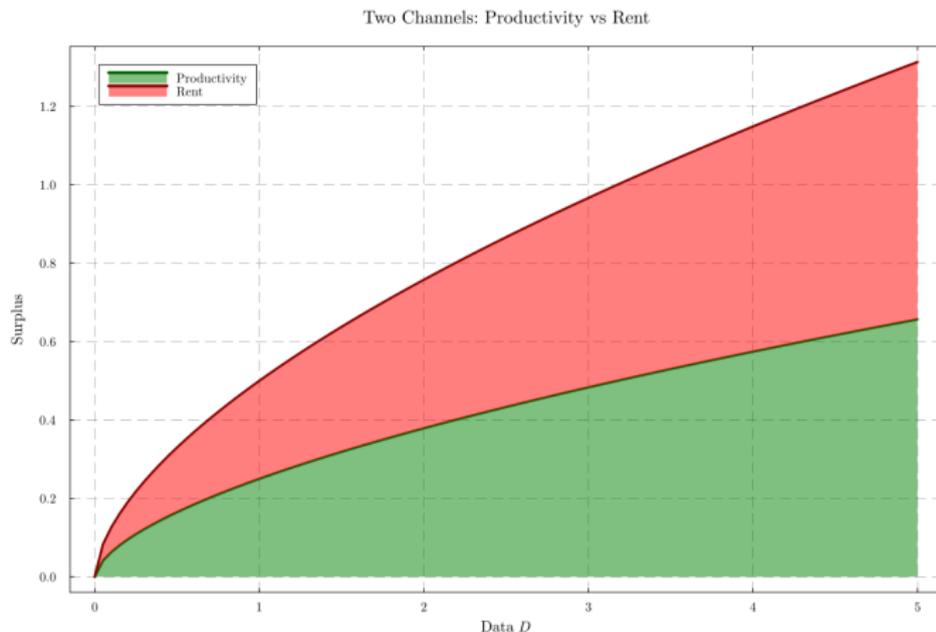
$$\frac{\partial r}{\partial \rho} < 0 \quad (\text{privacy reduces profitability})$$

$$\frac{\partial r}{\partial \psi} > 0 \quad (\text{sharing increases profitability})$$

$$\frac{\partial^2 r}{\partial \rho \partial \psi} < 0 \quad (\text{substitutes!})$$

Implication: Privacy and sharing are policy substitutes, not complements

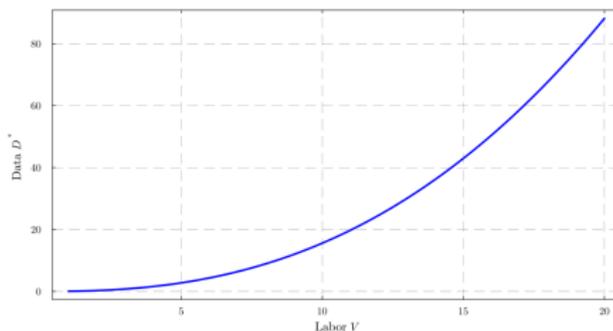
Data's Dual Nature: Productivity vs. Rent



- Productivity (green): Value creation that persists under any ownership
- Rent (red): Monopolistic appropriation dependent on exclusion

Winner-Take-All: Data Intensity Grows with Scale

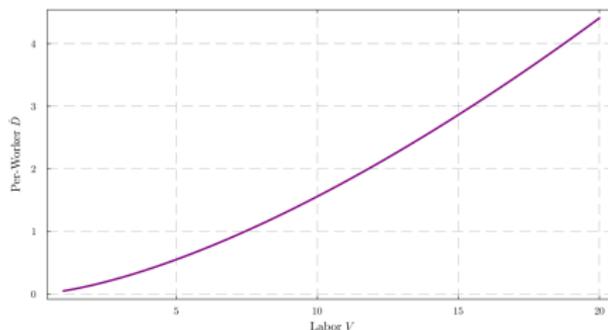
Total Data Grows Super-Linearly



- Per-worker data intensity increases with platform size

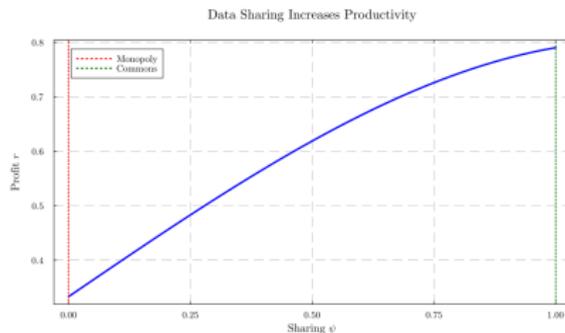
$$\tilde{D} \propto V^{\gamma/(1-\gamma)}$$

Data Intensity Increases (Flywheel)



- “Data flywheel” generates persistent advantages for dominant platforms

Data Sharing Eliminates Rent, Preserves Productivity



- Profit increases by 140% when moving from monopoly to commons



- Rent falls to zero: data's productive value does not require private ownership

- 1 **Mandate data sharing:** Interoperability, portability ($\uparrow \psi$)
- 2 **Privacy-preserving tech:** Differential privacy, federated learning
- 3 **Antitrust:** Break up data monopolies

Optimal trajectory: (Strong privacy + data commons)

Current trajectory: (Moderate privacy + monopolization)

- 1 Calibrating the model using country, sector, year data (EUKLEMS - IntanProd)
- 2 Estimate a reduced-form model using data on labour, investments in data (requires collaboration with Bdl/ISTAT)
- 3 Think of alternative ways of modelling data to check the robustness of our main takeaways