A Model of the Data Economy

Maryam Farboodi Laura Veldkamp

Is the Data Economy New?

- the economy is changing and we need new tools!
  - the largest firms are valued primarily for their data
  - do the economics change? or is data just new capital?
- challenges
  - economic activity generates informative data

    production/transaction is a form of *active experimentation*
  - data is a non-rival good whose value declines when it is sold

    $\rightarrow$ semi-rival
  - value of data: a piece of data is used for multiple periods, how much is it valued?

    $\Rightarrow$ dynamic programming with information as a state variable
  - data depreciation rate depends on economic conditions

A Macro Model of Data

- A continuum of competitive firms i uses capital $k_{i,t}$ to produce $k_{i,t}^{\alpha}$ units of goods

- Quality of goods depends on chosen production technique $a_{i,t}$ and distance to optimum $(\theta_t + \varepsilon_{a,i,t})$:

$$A_{i,t} = g\left((a_{i,t} - \theta_t - \varepsilon_{a,i,t})^2\right)$$

  $g'(.) < 0$ (accuracy is good)

- The optimal technique $\theta_t + \varepsilon_{a,i,t}$ changes over time.
  - θ_t: AR(1), innovation $\eta_t \sim N(\mu, \sigma_\theta^2)$

$$\theta_t = \bar{\theta} + \rho(\theta_{t-1} - \bar{\theta}) + \eta_t$$

  - $\varepsilon_{a,i,t} \sim N(0, \sigma_a^2)$ is unlearnable and i.i.d.

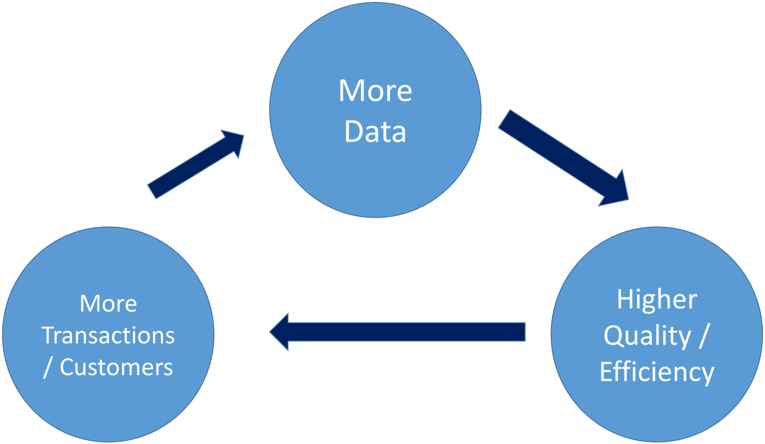
- Demand: Price is decreasing in aggregate output.

Model: Data is Information for Forecasting

- at time t, firm obtains $n_{i,t}$ data points about θ_{t+1}
  - data is a byproduct of production with **data-mining ability** z_i
  - $n_{i,t} = z_i k_{i,t}^{\alpha}$
- each data point $m \in [1 : n_{i,t}]$ reveals

$$s_{i,t,m} = \theta_{t+1} + \xi_{i,t,m} \qquad \text{where} \qquad \xi_{i,t,m} \sim N(0, \sigma_{\varepsilon}^2)$$

Data Feedback Loop



Model: Market for Data

- $\delta_{i,t}$: amount of data traded by firm *i* at time *t*
  - $\delta_{i,t} > 0$: data purchases (< 0: data sales)
  - firm can buy or sell, not both
- data price π_t clears the data market
- multi-use data: firm can sell it and still use it
  - ι**: fraction of sold data that is lost** ($\iota > 0$)
  - many data contracts include prohibitions on seller use, or this captures imperfect competition
- data adjustment cost: $\Psi(\cdot)$: avoid 1-period convergence

Results Overview

- Data is an asset: depreciate and value it
- Long run
  - no long-run growth without innovation
- Short run
  - increasing returns, negative initial losses
  - data barter
- Welfare and business stealing

The contribution is a tool.
Realistic predictions support the idea that the framework is useful.

How does data depreciate?

- Ex: Data to forecast an AR(1): $\theta_{t+1} = \rho\theta_t + \eta_{t+1}$, $\varepsilon_{t+1} \sim N(0, \sigma_\theta^2)$.
- Precision: $V[\theta_t|\mathscr{I}_t]^{-1} := \Omega_t$. Call this a "stock of knowledge."
- Prior variance of tomorrow's state: $V[\theta_{t+1}|\mathscr{I}_t] = \rho^2\Omega_t^{-1} + \sigma_\theta^2$.
- If data forecasts θ_{t+1}, then a data point is: $s_t = \theta_{t+1} + e_{st}$.
- Bayes law for normals says: $t+1$ precision Ω_{t+1} is prior precision plus precision of n_s data points $n_s\sigma_s^{-2}$.

$$\Omega_{t+1} = \underbrace{(\rho^2\Omega_t^{-1} + \sigma_\theta^2)^{-1}}_{depreciated\ t\ data} + \underbrace{n_s\sigma_s^{-2}}_{new\ data\ inflows}$$

  Similar to $k_{t+1} = (1-\delta)k_t + i_t$, where $\delta = 1 - (\rho^2 + \sigma_\theta^2\Omega_t)^{-1}$.

- Data depreciates faster when it's abundant Ω_t and the environment has volatile innovations σ_θ^2.

VALUING DATA: A RECURSIVE SOLUTION

- $a^*_{i,t} = \mathbb{E}[\theta_t + \varepsilon_{i,t} | \mathscr{I}_{i,t}] \rightarrow$ Quality $A_{i,t} \approx$ a fn of squared forecast error
- **state variable: stock of knowledge**

$$\Omega_{i,t} \equiv \mathbb{E}\left[\left(\mathbb{E}[\theta_t | \mathscr{I}_{i,t}] - \theta_t\right)^2 | \mathscr{I}_{i,t}\right]^{-1} \qquad \text{(posterior precision)}$$

LEMMA

optimal sequence of capital / data choices $\{k_{i,t}, \delta_{i,t}\}$ solves:

$$V(\Omega_{i,t}) = \max_{k_{i,t}, \delta_{i,t}} P_t \mathbb{E}_i\left[A_{i,t}(\Omega_{i,t})\right] k_{i,t}^{\alpha} - \Psi(\Delta\Omega_{i,t+1}) - \pi_t \delta_{i,t} - r k_{i,t} + \frac{V(\Omega_{i,t+1})}{1+r}$$

where (depreciated data + data inflows)

$$\Omega_{i,t+1} = \left[\rho^2(\Omega_{i,t} + \tilde{\sigma}_a^{-2})^{-1} + \sigma_\theta^2\right]^{-1} + \left(z_i k_{i,t}^{\alpha} + \delta_{i,t}(\mathbf{1}_{\delta_{i,t}>0} + \iota \mathbf{1}_{\delta_{i,t}<0})\right) \sigma_\varepsilon^{-2}$$

Semi-Rivalry and Data Market

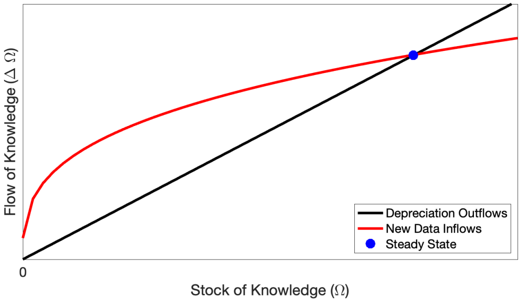
- benefit to buying one unit of data: $V'(\Omega_t) - \pi_t$
- cost of selling one unit of data: $-\iota V'(\Omega_t) + \pi_t$
- *negative* bid-ask spread
- *effective* price of data

$$\pi_{i,t} = \frac{\pi_t}{\mathbf{1}_{\delta_{i,t}>0} + \iota \mathbf{1}_{\delta_{i,t}<0}}$$

- data market active even in steady state with identical firms

Understanding Growth. Data Inflows and Outflows

- **inflow**: $z_i k_{it}^{\alpha} \sigma_{\varepsilon}^{-2}$ (# of data points $\times$ precision)
- **outflow**: data depreciation



- steady state: inflows $=$ outflows $\rightarrow$ **growth stops**

How General is Diminishing Returns?

Proposition

For sustained growth $g_t > \underline{g} > 0$, both most hold:

1. **Infinite output from one-period-ahead forecasts**
   There exists a level of forecast error $\underline{v}$ where the quality function approaches infinity, $\lim_{v \to \underline{v}} g(v) = \infty$

2. **No fundamental randomness**
   The state $\theta_{t+1} + \varepsilon_{t+1}$ has no time-t fundamental randomness - the future must be a deterministic function of time-t observables.

Endogenous Growth

- alternative quality formulation: data for idea creation

$$A_{i,t} = A_{i,t-1} + \max\{0, \hat{\Delta} A_{i,t}\}$$
$$\hat{\Delta} A_{i,t} = \bar{A} - (a_{i,t} - \theta_t - \varepsilon_{a,i,t})^2$$

- data increases step size in a quality ladder $\rightarrow$ growth
- data reduces the variance: R&D that focuses on risk-reduction

Data used for R&D needs to be measured separately

Long run: data looks similar to capital.

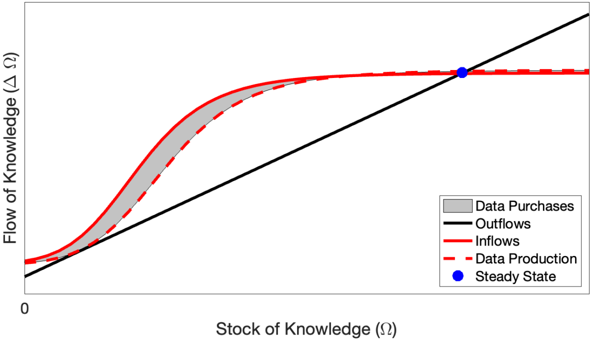
data market

Short Run: Increasing Returns

single firm enters a steady state

Proposition (Convex Data Flow)

there exist parameters such that when knowledge is scarce $\Omega_{it} < \hat{\Omega}$, net data flow $d\Omega_{it}$ increases over time.



Data Barter. Why Produce At a Loss?

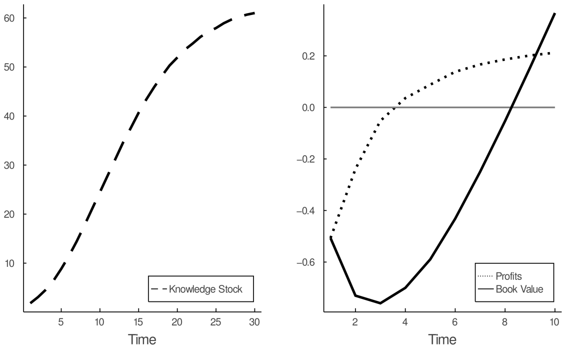
- *barter*: data is "exchanged" for the good
  - at good price $P_t = 0$
- **result:** data barter arises early in a firm's life
  - firms produce goods at a loss to generate data

$$\partial V_t / \partial \Omega_{i,t} > 0$$

- reality: lots of data is bartered for services (phone apps)
  Lots of partial data barter from firms that want more data.
- *GDP is missing lots of digital economic activity because price does not reflect the value of the transaction.*

Initial Losses and Low Book-to-Market

Our data firms look like tech firms:



- Early profit losses are an investment in data (Amazon, Uber, ...).
- Book value: only includes purchased data
- But market value is high because data stock is valuable!

Welfare: Decentralized Problem

- **Household problem**

$$\max_{c_t, m_t} \sum_{t=0}^{+\infty} \frac{u(c_t) + m_t}{(1+r)^t}$$

$$\text{s.t.} \quad P_t c_t + m_t = \Phi_t = \text{aggregate profits of all firms} \qquad \forall t$$

- **Two types of firms: efficient and inefficient data-miners**

$$\max_{\{k_{i,t}, \delta_{i,t}\}_{t=0}^{\infty}} V(0) = \sum_{t=0}^{+\infty} \frac{1}{(1+r)^t} \left(P_t \mathbb{E}[A_{i,t} | \mathscr{I}_{i,t}] k_{i,t}^{\alpha} - \Psi(\Delta\Omega_{i,t+1}) - \pi_t \delta_{i,t} - r k_{i,t} \right)$$

$$\Omega_{i,t+1} = \left[\rho^2 (\Omega_{i,t} + \sigma_a^{-2})^{-1} + \sigma_\theta^2 \right]^{-1} + \left(z_i k_{it}^{\alpha} + \delta_{it} (\mathbf{1}_{\delta_{it}>0} + \iota \mathbf{1}_{\delta_{it}<0}) \right) \sigma_\varepsilon^{-2}$$

- **Market clearing**

$$c_t = \lambda A_{L,t} k_{L,t}^{\alpha} + (1-\lambda) A_{H,t} k_{H,t}^{\alpha} \qquad \text{(retail good)}$$

$$m_t + r \left(\lambda k_{L,t} + (1-\lambda) k_{H,t} \right) + \Sigma_i \lambda_i \Psi(\Delta\Omega_{i,t+1}) = 0 \qquad \text{(numeraire good)}$$

$$\lambda \delta_{L,t} + (1-\lambda) \delta_{H,t} = 0 \qquad \text{(data)}$$

Solution conincides with social planner solution. Efficiency!

DATA AS A BUSINESS STEALING TECHNOLOGY

- Of course, there are inefficiencies – lots of data used for advertising.
- Suppose data processing helps the firm that uses it, but has no social value (keeing up with Joneses externality). Morris-Shin (2002)

$$A_{i,t} = \bar{A} - \left(a_{i,t} - \theta_t - \varepsilon_{a,i,t}\right)^2 + \int_{j=0}^{1} \left(a_{j,t} - \theta_{j,t} - \varepsilon_{a,j,t}\right)^2 dj$$

- equilibrium unchanged, welfare changed
  - **inefficient capital choice:** over-investment in capital
  - **inefficient data choice:** over-supply of data $\rightarrow$ **too much trade**

CONCLUSIONS

- macroeconomics of big data
- knowledge economies have quirky features:

  economic transactions generate data, semi-rivalry, data accumulation and depreciation, increasing and decreasing returns

- flexible tool that captures many features of the data economy:

  endog growth, data platforms, data barter, business stealing, welfare

- lots of new directions to explore:

  measurement, data pricing and valuation theory, firms dynamics with entry/exit, imperfect competition, optimal policy...

Steady State Data Market: Single Type of Firm

- **non-exclusivity of data:** no data trade is not an equilibrium

  $V'(\Omega^{ss}) =$ marginal value of one unit of data in a no trade/symmetric eq

  marginal cost of selling one unit $= \iota \times$ marginal benefit of buying one unit

$\Rightarrow$ **no symmetric equilibrium:** λ^* fraction of firms buy, $1 - \lambda^*$ sell

λ^* determined endogenously

$$\Omega_s^{ss} < \Omega_b^{ss}, \ \frac{dV'(\Omega)}{d\Omega} < 0$$

$$\pi = \frac{\iota}{1+r} V'(\Omega_s^{ss}) = \frac{1}{1+r} V'(\Omega_b^{ss})$$

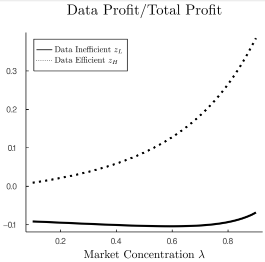
SS Data Market: Two Types of Firms
Specialization & Concentration

- data mining ability: $z_L < z_H$, $\lambda =$ measure of z_L firms

Proposition (Data Efficient Firms Accumulate Less Knowledge & Specialize in Data Sales)

For sufficiently low ι, $\Omega_H < \Omega_L$.

- few efficient data producers
  $\equiv$ high concentration
  $\Rightarrow$ more specialization



distributional consequences of data economy is different from capital economy

back