Data Macroeconomics

Maryam Farboodi   Laura Veldkamp

Tools to Model the Data Economy

- The economy is changing. 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 is a form of active experimentation.
  - How do we value information used for multiple periods? Requires dynamic programming with information as a state variable.
  - Data depreciation depends on economic fluctuations.
  - Data is a non-rival good, whose value declines when it is sold.

Overview

- Model: recursive framework, as tractable as std DSGE.
  - A tool to value data and data-intensive firms.
- Long run growth
  - Diminishing returns
- Short run tools
  - Increasing returns
- Efficiency and externalities
- Imperfect Competition and Market Power

A Macro Model of Data

- continuum of competitive firms i
- each uses capital $k_{i,t}$ to produce $k_{i,t}^{\alpha}$ units of goods
- these goods have quality $A_{i,t}$
- Output / demand

$$Y_t = \int_i A_{i,t} k_{i,t}^{\alpha} di$$
$$P_t = \overline{P} Y_t^{-\gamma}$$

Model: Quality Depends on Forecasts

- firm has one optimal technique: $\theta_t + \varepsilon_{a,i,t}$
  - 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.
  - quality depends on chosen production technique $a_{i,t}$ and distance to optimum $(\theta_t + \varepsilon_{a,i,t})$:

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

- $g(.)$: monotonically decreasing (accuracy is good)

Model: Data is Information for Forecasting

- at time t, firm obtains $n_{i,t}$ data points about θ_{t+1}
  - $n_{i,t} = z_i k_{i,t}^{\alpha}$
  - data is a byproduct of production with **data-mining ability** z_i
- 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**
  - We need $\iota > 0$ for a competitive equilibrium to exist.
  - Many data contracts include prohibitions on seller use. Or this captures imperfect competition.
- Data adjustment cost: $\Psi(\cdot)$: avoid 1-period convergence

Perfect Competition Results

- Valuing and depreciating data as an asset.
- What happens in the long run?
  - Diminishing returns: no long-run growth without innovation.
  - Endogenous growth: If data is used for R&D
- What happens in the short run?
  - Increasing returns, data poverty traps
  - data barter and book-to-market dynamics
- Welfare and business stealing

Solution: Depreciating Data with Bayes Law

- Goal is to forecast $\theta_{t+1} = \rho\theta_t + \varepsilon_{t+1}, \quad \varepsilon_{t+1} \sim N(0, \sigma_\theta^2)$.
- Priors: $E[\theta_t|\mathscr{I}_t]$ and $V[\theta_t|\mathscr{I}_t] := \Omega_t^{-1}$. Ω_t is "stock of knowledge."
- What are today's beliefs about tomorrow's state?

$$E[\theta_{t+1}|\mathscr{I}_t] = \rho E[\theta_t|\mathscr{I}_t]$$

$$V[\theta_{t+1}|\mathscr{I}_t] = \rho^2\Omega_t^{-1} + \sigma_\varepsilon^2$$

- Bayes law for normal variables says:
  posterior precision = prior precision + signal precision σ_s^{-2}.
- A law of motion for stock of knowledge (Kalman filter, Ricatti eqn):

$$\Omega_{t+1} = (\rho^2\Omega_t^{-1} + \sigma_\theta^2)^{-1} + \sigma_s^{-2}$$

  Discount more for low persistence ρ and volatile innovations σ_θ^2.

Semi-rival data as a negative bid-ask spread

- For data purchases ($\delta_{i,t} > 0$) data added to stock is $n_{i,t} + \delta_{i,t}$.
- For data sales ($\delta_{i,t} < 0$), net data additions are $n_{i,t} + \iota \delta_{i,t}$.
- An adjusted price of data: $\tilde{\pi}_{i,t} \equiv \pi / (\mathbf{1}_{\omega_{i,t} > n_{i,t}} + \iota \mathbf{1}_{\omega_{i,t} < n_{i,t}})$.
- This price is higher for data sales than it is for data purchases. We can use tools from finance for modeling economies with bid-ask spreads.

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**

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

Lemma

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

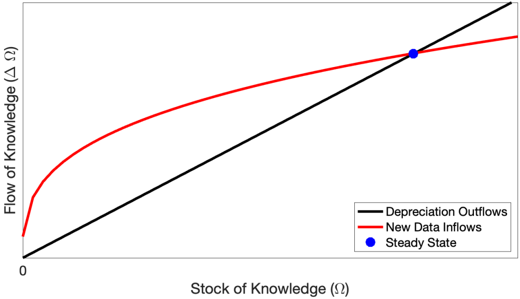
$$V(\Omega_{i,t}) = \max_{k_{i,t}, \delta_{i,t}} P_t \mathbb{E}_i\big[A_{i,t}(\Omega_{i,t})\big] k^{\alpha}_{i,t} - \Psi(\Delta\Omega_{i,t+1}) - \pi\delta_{i,t} - rk_{i,t} + \frac{V(\Omega_{i,t+1})}{1+r}$$

where $n_{i,t} = z_i k^{\alpha}_{i,t}$ and (Kalman filter)

$$\Omega_{i,t+1} = \Big[\rho^2(\Omega_{i,t} + \tilde{\sigma}_a^{-2})^{-1} + \sigma_\theta^2\Big]^{-1} + \big(n_{it} + \delta_{it}(\mathbf{1}_{\delta_{it}>0} + \iota \mathbf{1}_{\delta_{it}<0})\big)\, \sigma_\varepsilon^{-2}$$

Understanding Growth. Data Inflows and Outflows

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



- Growth stops for two reasons: Can't reduce errors below zero and unlearnable risk.

Endogenous Growth

- So far, data added to current productivity.
  Did not augment a stock of ideas.

- Alternative model: Data $\rightarrow$ idea creation

$$A_{i,t} = A_{i,t-1} + \max\{0, \Delta A_{i,t}\}, \tag{1}$$

$$\Delta A_{i,t} = \bar{A} - (a_{i,t} - \theta_t - \varepsilon_{a,i,t})^2. \tag{2}$$

- Data increases step size in a quality ladder $\rightarrow$ growth

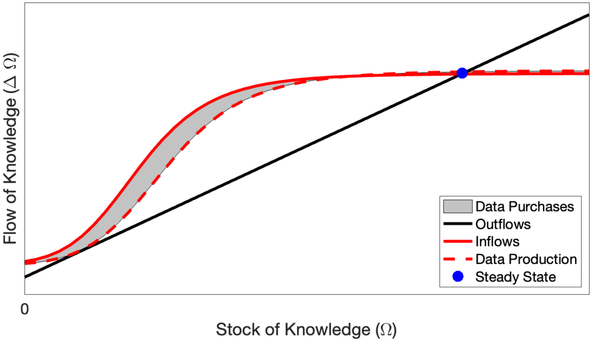
- One should distinguish data used for research from data used for process optimization. (Just like capital vs. R& D investment.)

Long run, data looks just like capital

Short Run: Single firm enters a steady state

Proposition (**Convex Data Flow**)

There exist parameters α and $\bar{P}$, and a threshold $\hat{\Omega}$ such that when knowledge is scarce $\Omega_{it} < \hat{\Omega}$, net data flow $d\Omega_{it}$ increases over time.



Could explain firm size divergence, . . . and eventually convergence?

Data Barter.
Why Produce At a Loss?

- *Barter means*: data is “exchanged” for the good
  - at good price $P_t = 0$
- Reality: lots of data is bartered for services (phone apps)
- Result: Data barter arises early in a firm’s life (depends on pref.s)
  - firms produce goods at a loss to generate data

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

- GDP is missing lots of digital economic activity because price does not reflect value.
  How to correct it?

Welfare, Decentralization and Efficiency

- Household problem: c_{it} is quality units consumed

$$\max \sum_{t=0}^{+\infty} \beta^t \left(\bar{u} \frac{c_t^{1-\gamma}}{(1-\gamma)} + m_t - \frac{h_t^{1+\nu}}{(1+\nu)} \right)$$

$$\text{subject to} \qquad p_t c_t + m_t = \Pi_t + w_t h_t, \ \ \forall t$$

- Numeraire good m production linear in labor h_t, with wage w_t

$$\max_{n_t} \bar{m} n_t - w_t h_t$$

- Retail good c produced with data and capital, with $\mathbb{E}[A_{i,t}]$ and data law of motion, as before. Firm solves

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

- Consumption and data markets clear: $\sum_i c_{it} = A_{it} k_{jt}^{\alpha}$
  Numeraire used to pay capital rental at open economy r.

$\rightarrow$ Efficiency. No externalities.

Modeling Data Exetrnalities

- Lots of data used for advertising. Arguable business stealing.
- We can adjust quality so that data processing helps the firm that uses it, but has no social value:

  Morris-Shin (2002)

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

- Unchanged: firm choices, firm dynamics, aggregate quality
- Changed: welfare

Imperfect Competition (Eeckhout-Veldkamp, 2022)

- Key features of the model:
  - Data is information. Information resolves uncertainty (risk).
    We model this as uncertain consumer demand for a variety of products.
  - Firms price risk.
  - Firms choose an up-front investment and then choose how much to produce (Cournot).

- Data's Competing Effects on Markups

- In the paper: Markup Measurement
  - Data creates composition effects: How one measures markups matters.
  - Procyclical product and counter-cyclical firm / industry markups

Model (1/2): Firms

- n_F firms, indexed by i.
- The product space has N attributes, indexed by j.
  Goods, indexed by k, are combinations of attributes with weights a_{jk}.
- Firm chooses: investment in lowering marginal cost c_i at price $g(\chi_c, \boldsymbol{c}_i)$, and a quantity to produce $\mathbf{q}_i$ (vector) to maximize $E[U_i]$ and U_i:

$$U_i = \mathbf{E}\left[\pi_i | \mathscr{I}_i\right] - \frac{\rho_i}{2}\mathbf{Var}\left[\pi_i | \mathscr{I}_i\right] - g(\chi_c, \boldsymbol{c}_i) \tag{3}$$

$$\pi_i = q_i'(\mathbf{p} - \boldsymbol{A}\boldsymbol{c}_i) \tag{4}$$

- ρ_i is firm i's price of risk. From CAPM? From preferences?

Model (2/2): Demand and Data

- Demand: Customers' willingness to pay decreases in the quantity that all firms produce of attribute j + firm-specific shock b_i:

$$\tilde{v}_i = \underline{p} - \frac{1}{\phi} \sum_{i'=1}^{N} \tilde{q}_{i'j} + b_i \tag{5}$$

attribute demand shock $\mathbf{b}_i \sim N(0, I)$ is i.i.d. across firms (relax). ϕ is price elasticity of demand.

- Customers' value of a good is linear in attributes $v_k = \sum_{j=1}^{N} a_{jk} \tilde{v}_j$.

- Data is information about demand shocks: (exogenous) n_{di} data points for firm i, independent across firms. Each data point is : $\tilde{\boldsymbol{s}}_{i,z} = \boldsymbol{b}_i + \tilde{\boldsymbol{\varepsilon}}_{i,z}$, where $\tilde{\boldsymbol{\varepsilon}}_{i,z} \sim N(\mathbf{0}, \Sigma)$. Information set: $\mathscr{I}_i := \{\{\tilde{\boldsymbol{s}}_{i,z}\}_{z=1}^{n_{di}}\}_{i=1}^{n_F}$ (All info public. Private info in appdx)

Equilibrium

- FOC: Production depends on risk and price impact (denominator) and expected profit (numerator) Kyle '89 or Back and Zender '93.

$$\boldsymbol{q}_i = \left(\rho_i \mathbf{Var}\left[\boldsymbol{p}_i|\mathscr{I}_i\right] + \frac{\partial \mathbf{E}\left[\boldsymbol{p}_i|\mathscr{I}_i\right]}{\partial \boldsymbol{q}_i}\right)^{-1} \left(\mathbf{E}\left[\boldsymbol{p}_i|\mathscr{I}_i\right] - \boldsymbol{c}_i\right)$$

- Define sensitivity of production to a change in the expected price:
  $\boldsymbol{H}_i := \left(\frac{2}{\phi} I_N + \rho_i \mathbf{Var}\left[\boldsymbol{b}_i|\mathscr{I}_i\right]\right)^{-1}$ and $\bar{H} = \sum_{i=1}^{n_F} \mathbf{H}_i$
  Data lowers *Var*, raises $\mathbf{H}_i$.
  $\mathbf{H}_i$ governs the $cov(q_i, p_i)$.
  Data allows a firm to choose quantities that covary with prices.

- Optimal choice of cost (firm size):

$$\frac{\partial \mathbf{E}[U_i]}{\partial \tilde{\boldsymbol{c}}_{ij}} = \frac{1}{2} \underbrace{\frac{\partial \mathbf{E}[\tilde{\boldsymbol{p}} - \tilde{\boldsymbol{c}}_i]' \boldsymbol{H}_i \mathbf{E}[\tilde{\boldsymbol{p}} - \tilde{\boldsymbol{c}}_i]}{\partial \tilde{\boldsymbol{c}}_{ij}}}_{marginal\ benefit} - \underbrace{\frac{\partial g(\chi_c, \tilde{\boldsymbol{c}}_i)}{\partial \tilde{\boldsymbol{c}}_{ij}}}_{marginal\ cost} = 0 \quad \forall j$$

Product Markups

Product-level markup for k produced by firm i:

$$M_{ik}^{p} := \mathbf{E}[\boldsymbol{p}_i(k)]/\boldsymbol{c}_i(k)$$

$$= \frac{1}{a_k' \widetilde{\boldsymbol{c}}_i} \left(a_k' \underline{\boldsymbol{p}} - a_k' \boldsymbol{A}^{-1} \left(\phi I + \bar{\boldsymbol{H}} \right)^{-1} \left(\sum_{i'} \hat{\boldsymbol{H}}_{i'} \boldsymbol{A} \left(\overline{\boldsymbol{p}} - \widetilde{\boldsymbol{c}}_{i'} \right) \right) \right)$$

What makes product markups large?

- Having lots of valuable attributes: big a_k's.
- Low price elasticity of demand: Low ϕ.

  New:
- Scarce data or high price of risk. High $\rho\, Var[p|\eta_j]$ makes $\bar{H}$ low. Markup has a risk premium in there.

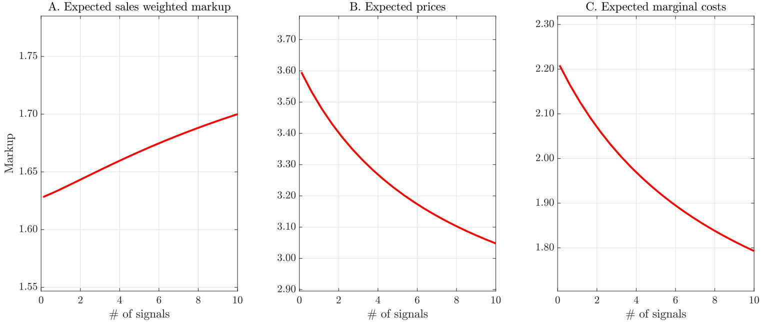
Product Markups: Firm Size Effect

Proposition

Data-investment complementarity. *Data lowers choice of c_i.*

Proposition

Higher investment raises product markups. *Lower c_i increases M^p_{ik}.*

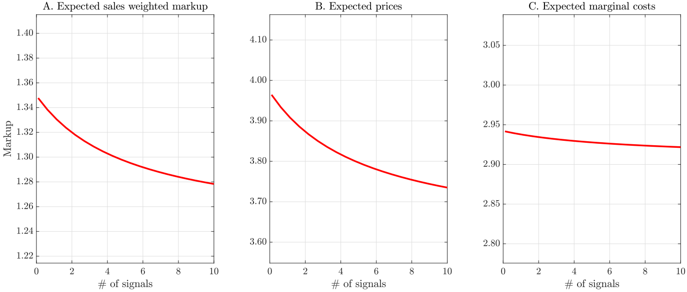


Single-good duopoly. $g(\chi_c, c_i) = \chi_c(\bar{c} - c_i)^2/2$ with $\chi_c = 1$ and $\bar{c} = 3$; $\underline{p} = 5$, $\phi = 1$, $\sigma_b = 1$,

Markups: Risk Premium Channel

Proposition

For investment cost χ_c or ρ sufficiently high, data reduces M_{ik}^{p}.



Proposition

Net effect: Data in(de)creases product markups when risk price or marginal cost of investment is sufficiently low (high).

Dynamic Imperfect Competition

How to combine the recursive economy with the static imperfect competition model?

Challenges: Every firm's information matters for price.
All information sets are state variables.

Potential payoffs:
Understand new firm dynamics in a digital economy.

How to measure competition when many prices are zero?

How to value firms that have never made a profit?

What if we add firm entry?

Welfare? Should we tax data? How?

Conclusions

- Economists cannot continue to study industrial economies.
  We live in a knowledge economy.
  We need modern tools that reflect this reality.
- Knowledge economies are not easy to work with.
  Production generates data, depreciation not std, semi-rivalry, increasing returns and decreasing returns, information state variables.
- Lots of new directions to explore
  - measurement
  - firms dynamics
  - data pricing and valuation theory