Data and the Aggregate Economy

Maryam Farboodi Laura Veldkamp

2021


amazon
UBER
Microsoft
Google
TESLA
How does this economy work?
David Parkins

Economics of Data as Economics of Prediction

- top-5 most valuable U.S. companies are data-intensive
  - will AI & ML spawn a new data economy?
- what is big data technology about?
  - **prediction technology**
  - data-driven innovation is also a way to go! but this is not what big data revolution is about
- can data accumulation *alone* sustain aggregate growth?
  - hold innovation fixed (as in Solow 1956)
- incorporating data-driven innovation is next

Overview

- model: recursive framework with data accumulation
  - transactions generate new data
  - an evolving world makes old data depreciate
- *decreasing returns*
  - data can**not** sustain growth without technological innovation
- *increasing returns*
  - when firm is young and data-poor
- applications
  1. data poverty traps: entry barriers
  2. data barter: initial losses
  3. specialization

A Macro Model of Data

- continuum of competitive firms i
- each firm use $k_{i,t}$ units of capital to produce $k_{i,t}^{\alpha}$ units of goods
- these goods have quality $A_{i,t}$
- firms take equilibrium price P_t as given and their quality-adjusted outputs are perfect substitutes

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

Model. Data Improves Good's Quality

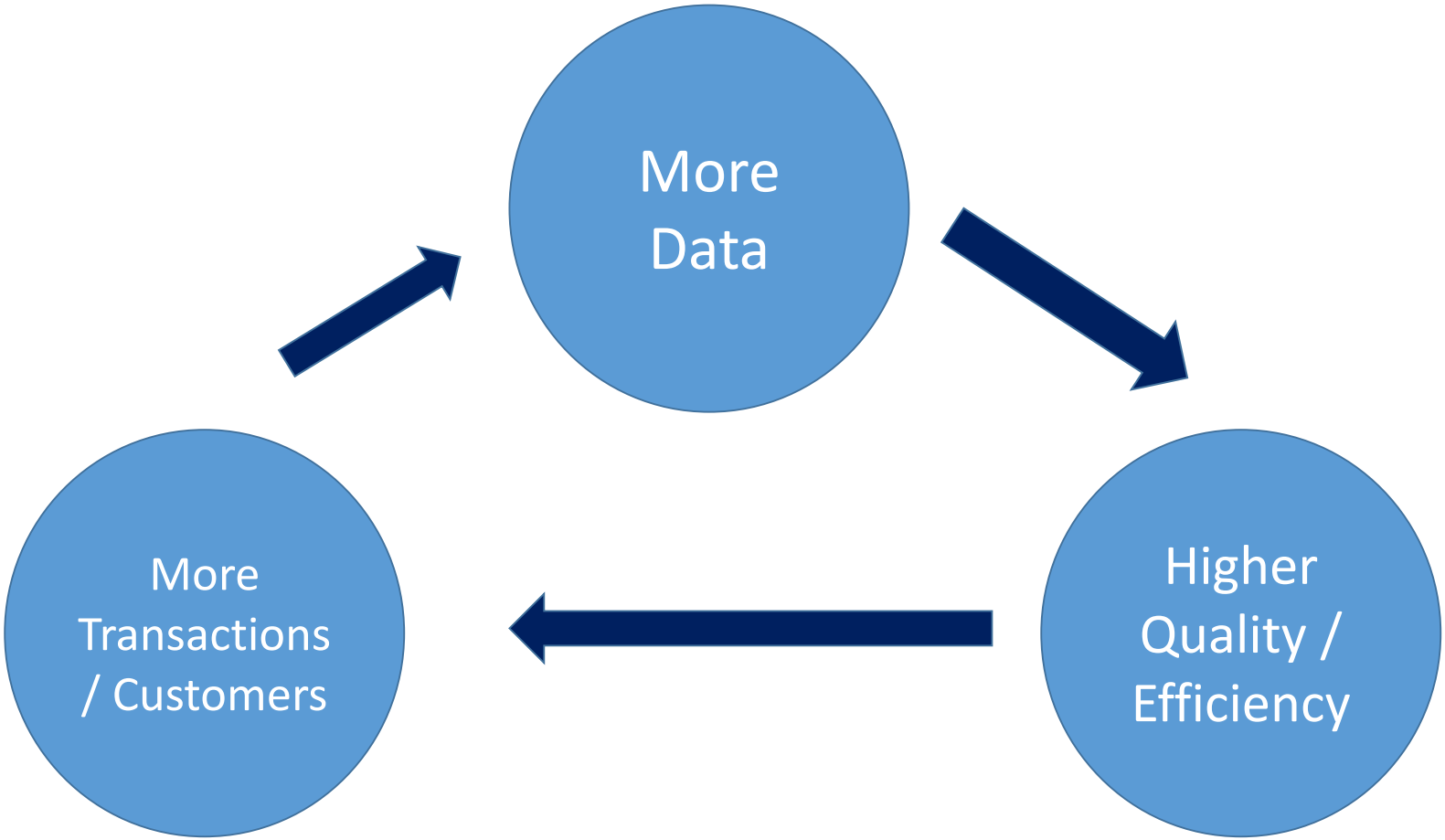
- quality depends on chosen production technique $a_{i,t}$
- firm has one optimal technique: $\theta_t + \varepsilon_{a,i,t}$
  - θ_t is AR(1) with param ρ, innovation $\eta_t \sim N(\mu, \sigma_\theta^2)$
  - $\varepsilon_{a,i,t} \sim N(0, \sigma_a^2)$ is unlearnable and i.i.d.

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

- at time t, the firm obtains $n_{i,t}$ data points about θ_{t+1}
  - $n_{i,t} = z_i k_{i,t}^{\alpha}$
  - data is a bi-product 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
  - $\delta_{i,t} > 0$ is data purchases
  - $\delta_{i,t} < 0$ is data sales
  - firm can buy or sell, not both
- data price π_t clears the data market
- multi-use data (non-rival): You can sell it and still use it
  - why isn't all data sold? There must be some cost/regulation
  - ι**: fraction of sold data that is lost**
  - could be privacy restrictions on multiple users, or market power
- data adjustment cost: $\Psi(\cdot)$ (avoid 1-period convergence)

A Simple Recursive Solution

- **state variable: stock of knowledge**

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

$$\mathbb{E}\left[A_{i,t}\right] = \bar{A} - \Omega_{i,t}^{-1} - \sigma_a^2 \quad \text{(expected quality)}$$

Lemma

The optimal sequence of capital investment choices $\{k_{i,t}\}$ *solves the recursive problem:*

$$V(\Omega_{i,t}) = \max_{k_{i,t},\delta_{i,t}} P_t\left(\bar{A} - \Omega_{i,t}^{-1} - \sigma_a^2\right) k_{i,t}^{\alpha} - \Psi(\Delta\Omega_{i,t+1}) - \pi\delta_{i,t} - rk_{i,t} + \left(\frac{1}{1+r}\right) V(\Omega_{i,t+1})$$

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

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

Data Accounting

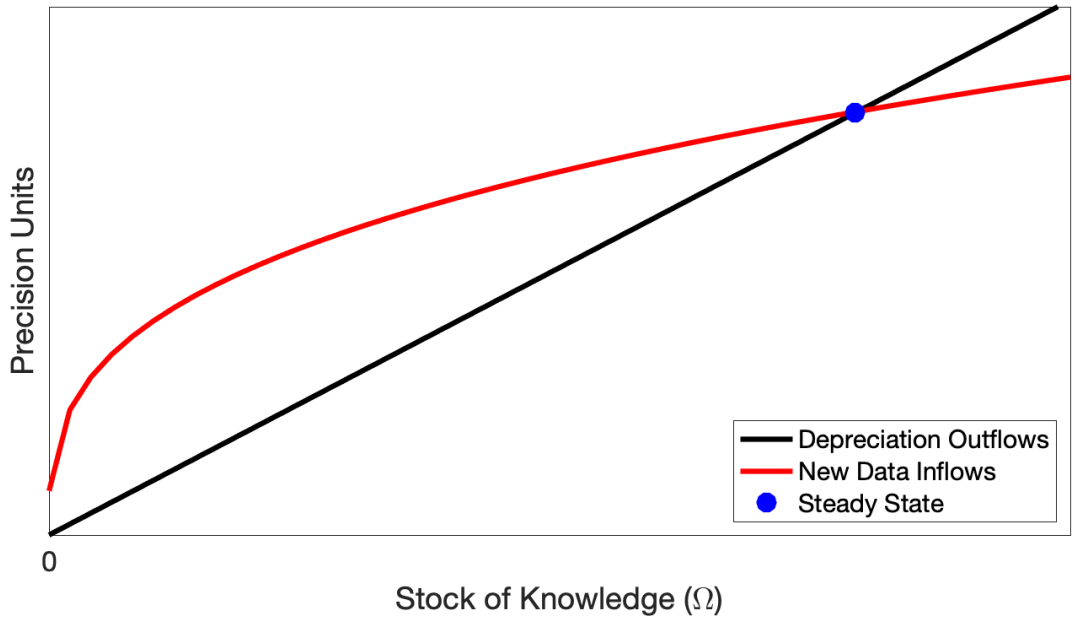
- Ω_t is the discounted sum of investments in data
- $V(\Omega_t)$ is the value of data
- $V'(\Omega_t)$ is the marginal value of a unit of data

Results Overview

- diminishing returns: no long-run growth
- increasing returns
- applications: market for data
  - data poverty trap
  - data barter
  - **specialization**

Understanding Growth. Data Inflow and Outflow

- **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 specific is this to our modeling assumptions?

What if Data-Driven Growth is Sustainable without Data-Driven Innovation?

Consider an arbitrary model where data is used to forecast and this affects output. Growth $g_t > \underline{g} > 0$ can be sustained only if

Proposition (Data and Infinite Output)

perfect forecasts generates infinite output.

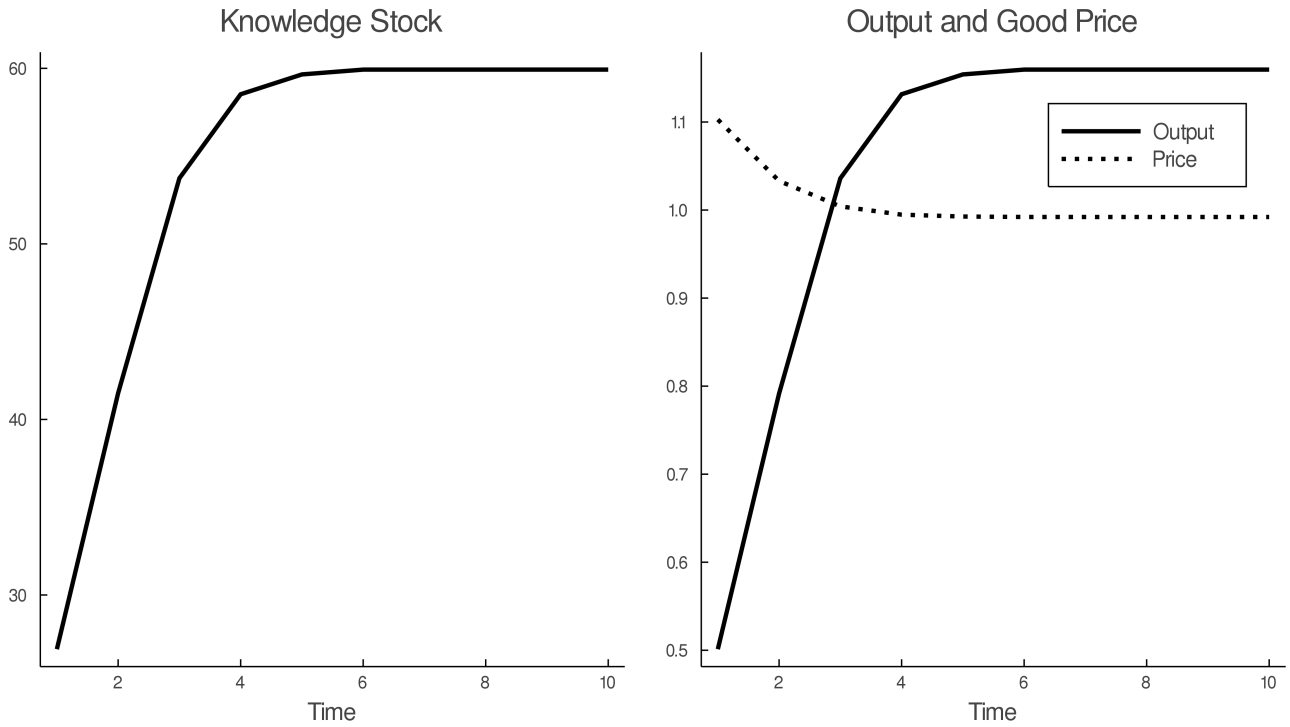
and

Proposition (Data and Infinite Precision)

future is perfectly deterministic, for some output relevant parameter.

- $y_{i,t} \to \infty \Rightarrow n_{i,t} \to \infty \Rightarrow \Omega_{i,t} \to \infty$
  $\Rightarrow$ zero variance forecasts $\equiv$ deterministic future

Decreasing Return.
Homogeneous Contemporaneous Firms



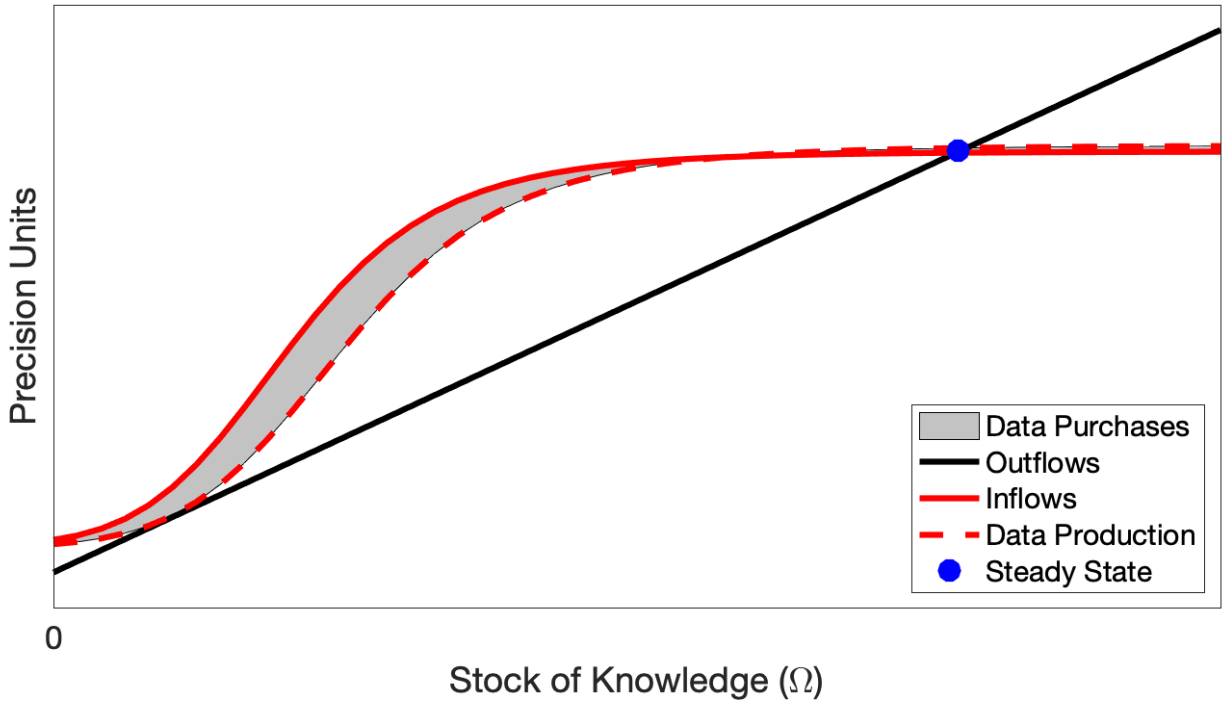
- homogeneous firms, grow together:
  knowledge and output grow on a concave path till convergence
- keep homogeneity, but one firms enters steady state of others:
  different dynamics

Increasing Return.
Entry and Trade

Single firm enters a steady state

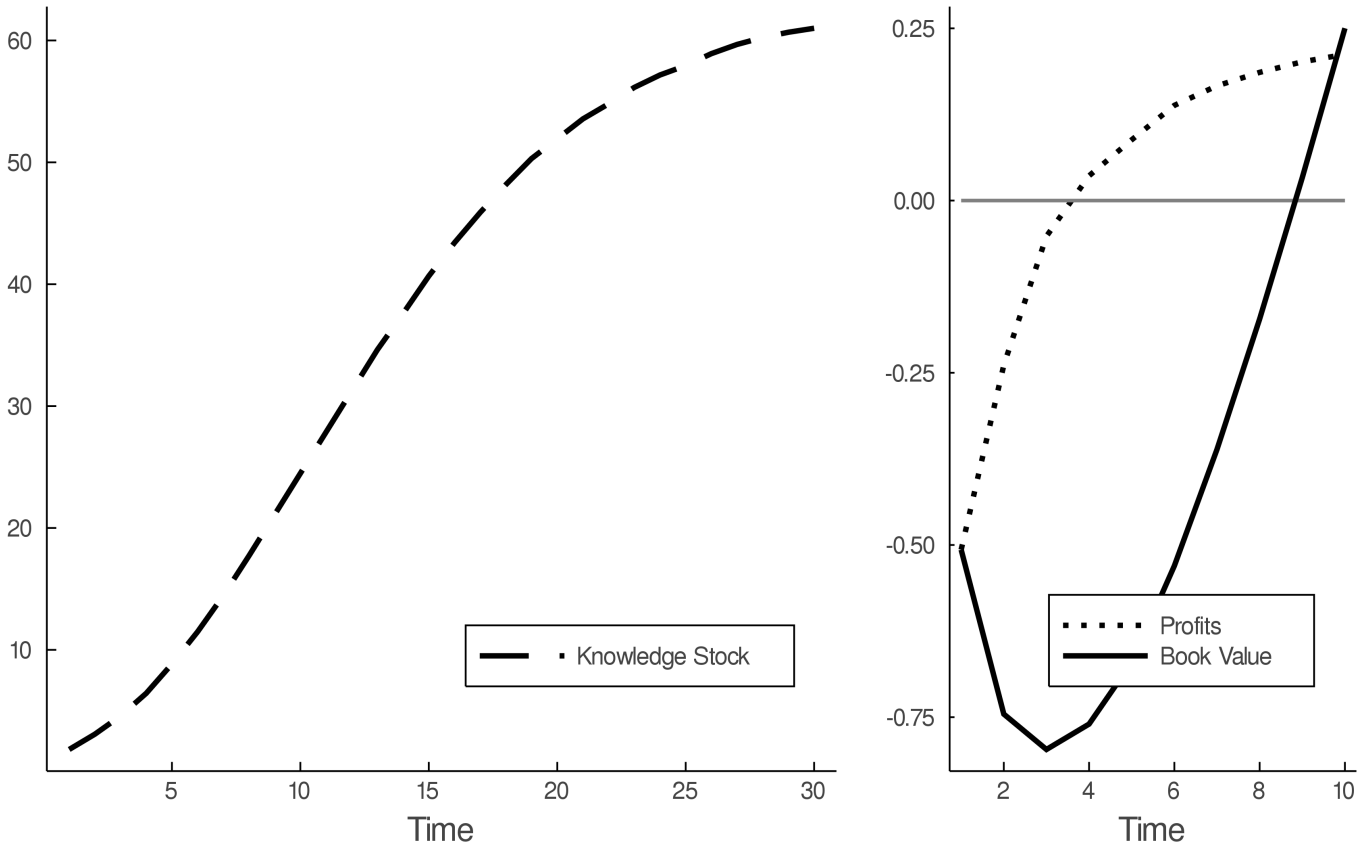
Proposition (Convex Data Flow)

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.



Data Poverty Trap.
Does Data Impede Firm Entry?

- one new firm entrant: negative profits, slow catch-up



- S-shaped accumulation of aggregate stock of knowledge

Data Barter.
Why Produce At A Loss?

- *barter* means data is "exchanged" for the good
  - at good price $P_t = 0$
- lots of data is bartered for services
- data barter arises early in a firm's life (Proposition 5)
  - firms produce goods at a loss to generate data: $\partial V_t / \partial \Omega_{i,t} > 0$

Return to Specialization

- who specializes in using data and producing high quality goods?
- **efficient data producers** do not use data
  instead the specialize in **selling data**
  - data is their comparative advantage

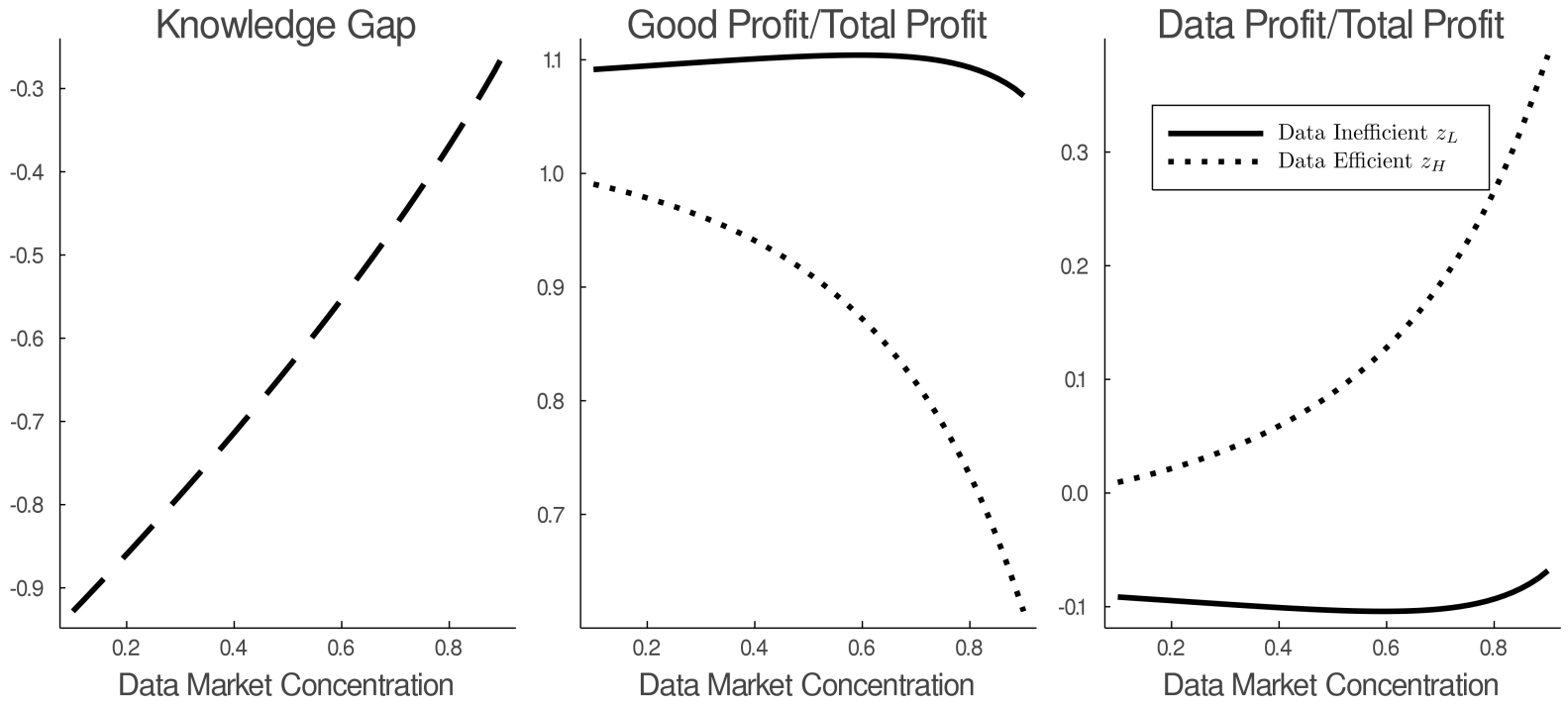
Proposition

for γ & α low, $\lambda < 1$, and $\iota < \bar{\iota}$, then **knowledge gap** *is negative:*

$$\Omega_H - \Omega_L < 0$$

Concentration & Data Specialization

- λ measures **concentration**
- $\lambda \uparrow \equiv$ fewer efficient data producers $\equiv$ high concentration



- efficient data producers specialize more in selling data: **emergence of data platforms?**

Conclusions

- assumptions
  - data raises productivity and is not always perfectly rival
  - "big data" is information used for forecasting
- results
  - diminishing returns to forecast precision in the long run
  - increasing returns when knowledge low: 0 prices and $(-)$ profits
  - returns to specialization: data platforms?
- lots of new directions
  - data-driven innovation and endogenous growth: $\bar{A}(n_{i,t})$
  - data pricing theory
  - "different" data has "different" degrees of relevance
  - firms dynamics with entry/exit and imperfect competition