Valuing Data as a New Asset Type

Laura Veldkamp
Columbia

based on work with Simona Abis, Juliana Begenau, Cindy Chung, Jan Eeckhout, Maryam Farboodi, Adrien Matray, Roxana Mihet, Thomas Philippon, Dhruv Singal and Venky Venkateswaran

2024

Big Data is Raising Big Questions

The most valuable firms in the world today are valued largely for their data.

Raises a whole agenda of questions, theoretical and empirical, that touch on many aspects of economics:

- Data is a new asset class. How do we price it? (finance)
- How do we measure monopoly power? What are reasonable harms to assess from data malfeasance? (law, IO)
- Is GDP measuring the value of an economy correctly? (macro)
- Does the value of data vary? Is it interest-rate sensitive? (monetary)

Our industrial-era economic tools need updating for the modern data economy.

Outline

This talk: How do we start on this agenda?
It begins with measuring the amount and value of firms' data.

- What do we mean by data?
- Data economy mechanics
  - A by-product of transactions
  - Buying and selling data
  - Depreciating data
- Measuring and valuing data: 6 approaches
- Conclude: Where next?

What do we mean by data?

- Data is digitized information
- Data of interest is big data:
  Often generated by economic activity: search history, traffic patterns, purchases...
- AI / ML are prediction technologies. Data used for prediction.
- Data is distinct from tech, patents, learning-by-doing and algorithms.
  - Ideas/ technologies: procedures or concepts. Data may be an input.
  - Learning by doing: Human capital, owned by workers. Not tradeable.
  - Data is the fuel for algorithms / AI

How data is generated and accumulated?

"It's free, but they sell your information."

Data price is zero. Data value is positive.

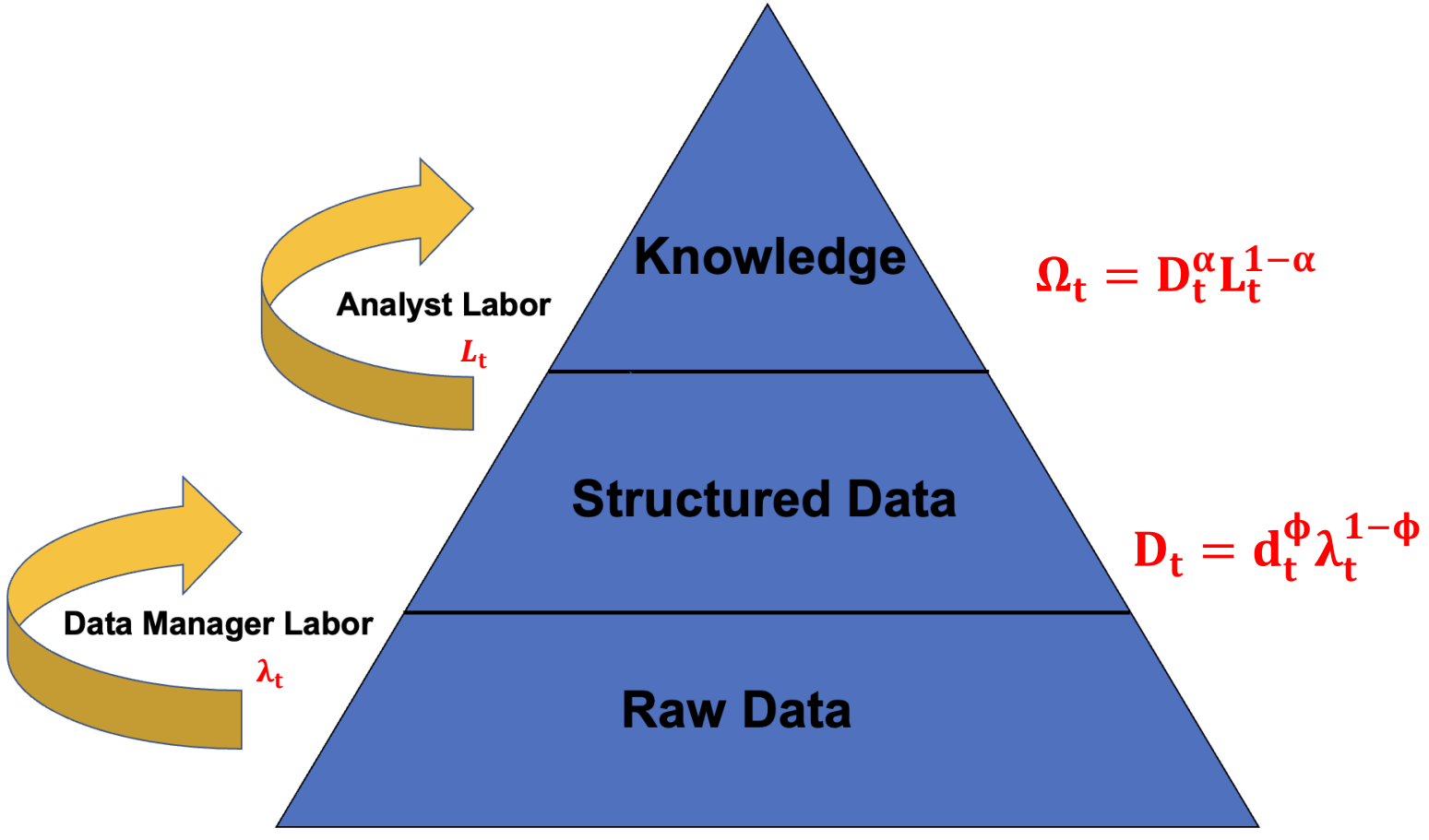
A Data Feedback Loop

What is Higher Quality?
How Does Data Create Value for Firms?

- Raises current profits: Choose better products, inventory, transportation, advertise to better customers. (Tucker '22)
- Creates market power
  - Firms with more data can grow bigger, exert monopoly power.
- Reduces risk
  - Data is information. Information resolves uncertainty (risk).
  - Finance tools here are crucial.
  - This could be big: Risk compensation is 2x expected return.

Accumulating Data: Raw Data, Structured Data and Knowledge

Maybe labor is an input into useable data?



Accumulating Data: Buying/Selling It

- Data is non-rival: You can sell it and keep it.
- Data leaks: through prices, through neighbors' transactions (Acemoglu et al '22)
- Data known by others loses value (strategic substitutability).
- Firms selling data face a commitment problem: (Liu, Ma, Veldkamp '23)
  To extract monopoly rents, sell data to few clients.
  But tomorrow, data seller can sell more copies at a lower price and dilute the value of data.

How does data depreciate?

- A key question for valuation.
- Ex: Data to predict an moving target: $\theta_{t+1} = \rho\theta_t + \epsilon_{t+1}$
- Data about today's state θ_t is less valuable tomorrow because the state θ_{t+1} has changed.
- Similar to capital depreciation: $k_{t+1} = (1-\delta)k_t + i_t$, where $\delta = 1 - (\rho^2 + \sigma_\epsilon^2 \Omega_t)^{-1}$.
- Data depreciates faster when it's abundant Ω_t and the environment has volatile innovations σ_ϵ^2.

Measuring and Valuing Data

Six Approaches to Measuring Data

1. Complementary inputs
2. Value functions
3. Revenue
4. Choice covariance
5. Market prices
6. Cost accounting

Measurement Approach 1: Complementary Inputs

- Knowledge is produced using structured data and "analyst" labor:

$$K_{it} = {D_{it}}^{\alpha} {L_{it}}^{1-\alpha}, \tag{1}$$

- New structured data is added to the existing stock of structured data with "data management" labor. Depreciates at rate δ:

$$D_{i,t+1} = (1-\delta) D_{it} + \lambda_{it}^{1-\phi} \tag{2}$$

- Estimate data stock from hiring and wages of each.
  What amount of data would make employing L_{it}, λ_{it} workers at wages w_{Lt}, $w_{\lambda t}$ optimal? (Abis-Veldkamp '23)

- Another observable complementary input: IT capital
  (Bresnahan, Brynjolfsson '02)

Measurement Approach 2: Value Function Approach

- The same tools macro uses to value capital work for data, with an adjusted law of accumulation.

$$V(\text{data}_t) = max_{K,L} A(\text{data}_t) K_t^{\alpha} L_t^{1-\alpha} - wL_t - rK_t + \beta V(\text{data}_{t+1})$$

- The state law of motion is the depreciation equation:

$$data_{t+1} = \underbrace{(\rho^2 data_t^{-1} + \sigma_\epsilon^2)^{-1}}_{depreciated\ \ data} + \underbrace{n_s \sigma_s^{-2}}_{new\ \ data\ \ inflows}$$

- Pair with theories of data inflows:
  - By-product of transactions
  - Data purchases / sales
  - Using labor to process raw data

VALUE FUNCTION ESTIMATION RESULTS ABIS, VELDKAMP '22

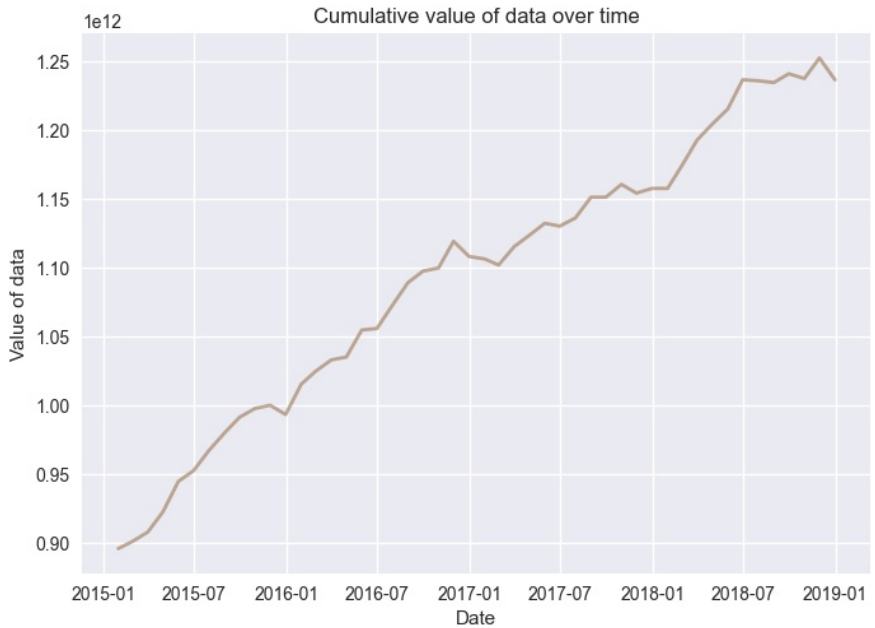


FIGURE: Estimated Value of the Aggregate Stock of Data, used for financial analysis, in hundreds of billions of current U.S. dollars, 2015-2018.

Data value is growing for 3 reasons:

1. Firms manage more data.
2. More analysis workers make each data point more valuable.
3. Firms are becoming more productive at using AI.

Measurement Approach 3: Revenue Approach

- The value of data is the pdv of the revenue it generates
- How to isolate data revenue from other revenue?
- This is doable. But you need a clear idea of how data generates revenue. A model is essential to compute counter-factuals with more/less data.
  (Manela, Kadan '21; Davila, Parlatore '21; Cong, Xie, Zhang '21)
- Problem: Data has different values to different agents (a private value asset)

Next: an example of valuing data as a private value asset, using a revenue approach.

Revenue Valution and Data's Private Value

- Suppose data is used to purchase a portfolio of risky assets.
- Value of data in $\mathcal{I}_{it}$ from an equilibrium with heterogeneous investors, correlated information and learning from noisy prices depends on 3 sufficient statistics:
  Expected return, variance of return and the variance of return, conditional on data.
- One can estimate these moments from return forecasting regressions.
- Finding: The same data is worth $\$10 - \$1.2m$, depending on the investor's wealth, investment style, price impact or trading frequency.
  (Farboodi, Singal, Veldkamp, Venkateswaran, 2022)

Measurement Approach 4: Choice Covariance

- Data allows agents to make better choices or matches.
- Better choices means actions q_t that covary with payoffs r_t.

$$E[q_t r_t] = E[q_t]E[r_t] + cov(q_t, r_t)$$

- Agents cannot achieve high covariance without information.
  They could have good data and no skill at analyzing it. This is a lower bound on data.
- Measure the covariance.
  Ex: Portfolio alpha or a customer click-conversion rate.
  Covariances may show up as aggregation effects: In Eeckhout-Veldkamp ('23): (firm markup - product markup) measures data

Measurement Approach 5: Market Prices

- Market prices for data
  - Data marketplaces: Snowflake, Datarade, etc.
  - This is the intersection of supply and demand.
    Lower bound on buyer firm's private value.
  - The value to a buyer often depends on who else knows the data.
- Market prices for firms
  - Market-to-book values can measure intangibles
    Crouzet-Eberly '20, Peters-Taylor '17
  - Already used for: Branding, patents, organizational capital, ...
    e.g., Belo, Gala, Salomao, Vittorino '21, Eisfeldt-Papanikolaou '13, '14
  - How to tease apart the value of data?
- Presumes that market participants know how to value data.

Measurement Approach 6: Cost Accounting

- A book value approach to valuing assets is to cumulate the sum of costly investments.
  Why not add up data costs?

- Most data is a by-product of some other economic transaction.
  - There was no explicit cost for it.
  - GAAP accounting rules only count data if it was purchased.

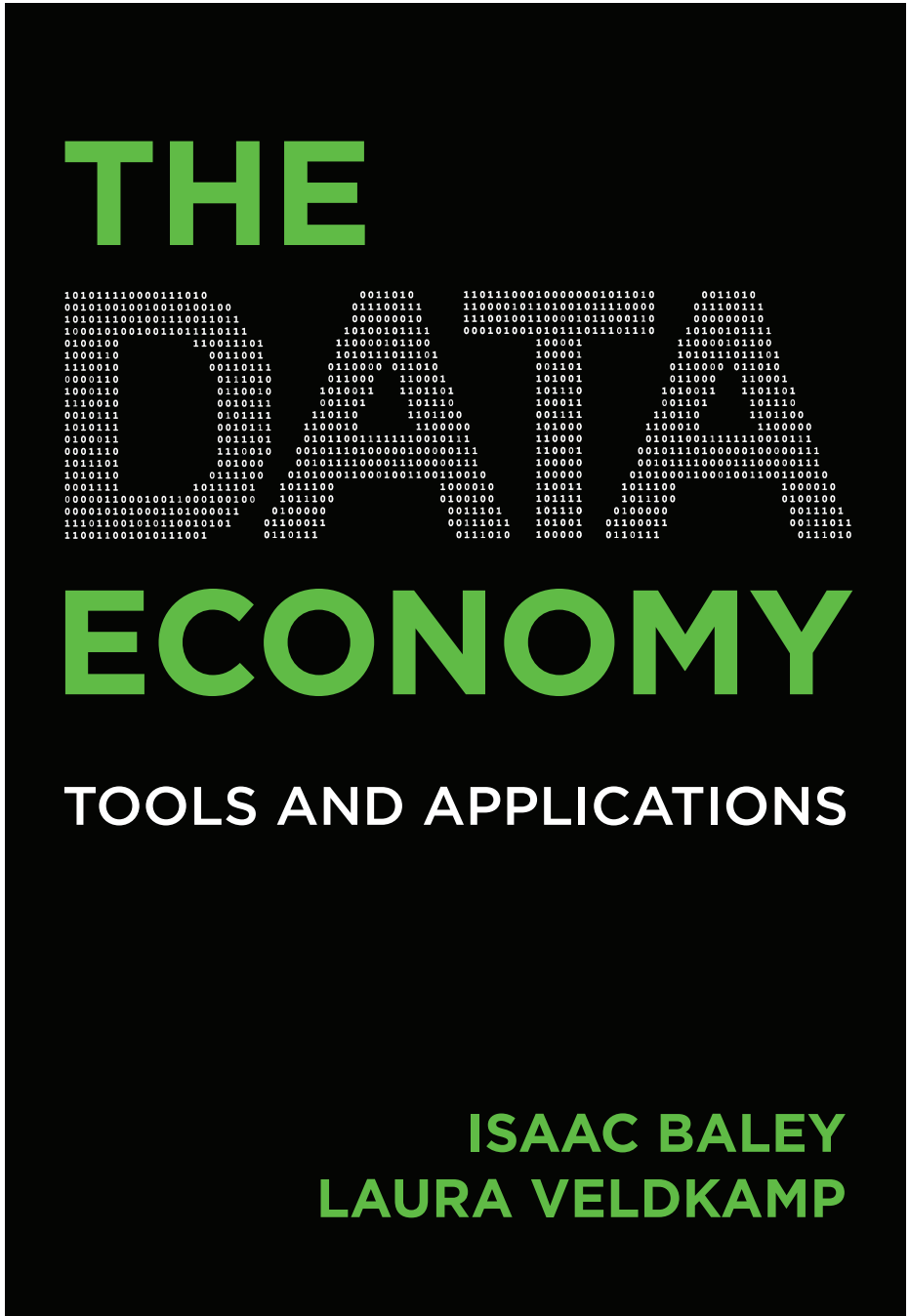
- An implicit cost: Data is bartered.
  - That shows up as a discount in the price of the good.

- This could work, if we can impute the data barter discount.

Conclusions

- Data is one of the most important and highly-valued assets in the modern economy.
  Also one of the hardest to observe, measure and put a price on.
- Different approaches needed for different situations.
- Theory and measurement need to work together here.
- Next steps:
  - Quantify the data barter value.
  - Private vs. social value: Do privacy laws prevent or worsen price discrimination, by undermining competition?
  - Data accumulation and firm life-cycles. Size, scope and competition.
- Textbook coming Soon

Coming Soon


THE
ECONOMY
TOOLS AND APPLICATIONS
ISAAC BALEY
LAURA VELDKAMP