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

2022

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 every aspect of finance:

- Data is a new asset class. Is it over-valued? (asset pricing)
- Do data-intensive firms have valuations that are realistic? (corporate fin)
- Are large troves of data an entry barrier for new firms? (entrepreneurship) What does market power look like if digital services are "free"?

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
  - Cost
  - Value function estimation
  - Complementary inputs
  - Choice covariance
  - Revenue
  - Intangibles approach
- Conclude: Where next?

What do we mean by data?

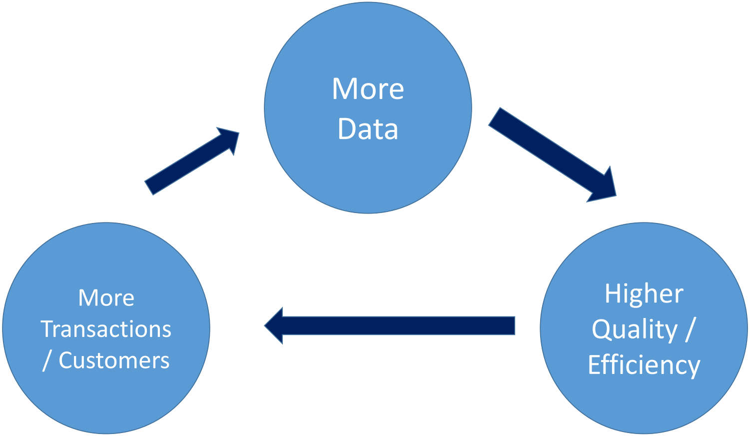
- Data is digitized information
- Data of interest is big data:
  Often generated by economic activity: search history, traffic patterns, purchases...
  Used for forecasting.
- Data is distinct from tech, patents, and learning-by-doing.
  - Ideas/ technologies: procedures or concepts. Data may be an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

How data is generated and accumulated?

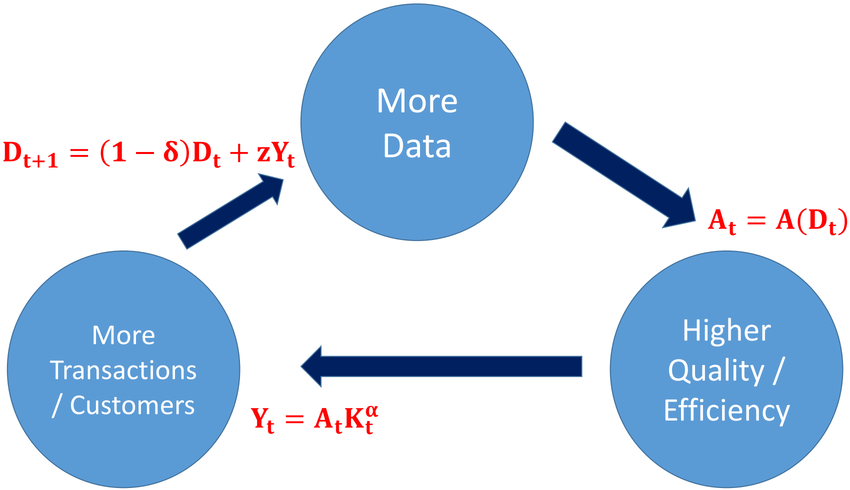
Most big data firms use is transactions, browsing/search history, GPS location, etc.

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

A Data Feedback Loop



A Data Feedback Loop



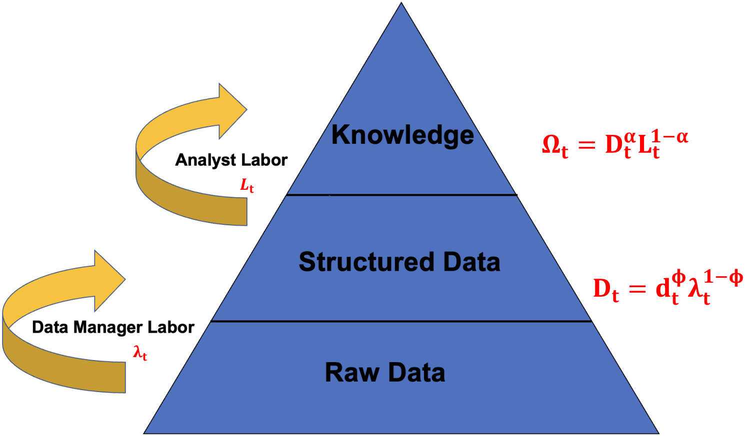
Data-fueled monopolies? (Farboodi, Veldkamp '22, Begenau-F-V '18)

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

- Raises current profits: Choose better products, inventory, transportation, advertise to better customers.
- Creates market power
  - Firms with more data can grow bigger, exert monopoly power.
  - Is data an entry barrier?
- 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 it

- Indirect and direct data sales
  - Just like financial information can be monetized through analyst reports (direct) or the services of a managed fund (indirect).
  - Google could sell you names and zip codes of people who bought iPads (direct, structured).
    Or, they can place ads for you, using their information (indirect, knowledge).

- Data is (imperfectly?) non-rival: You can sell it and keep it.
  - But does sold data lose value? How much?
  - Losing some of the data you sell is like a negative bid-ask spread.
    Seller loses less than buyer gains. (as in Farboodi, Veldkamp '22)

$$\text{data}_{t+1} = (1-\delta)\text{data}_t + \underbrace{\gamma_t}_{\textit{data purchases}} - \underbrace{\iota|\gamma_t|\mathbb{1}_{\gamma<0}}_{\textit{loss from data sales}}$$

How does data depreciate?

- A key question for valuation.
- Ex: Data to forecast an AR(1): $\theta_{t+1} = \rho\theta_t + \epsilon_{t+1}, \quad \epsilon_{t+1} \sim N(0, \sigma_\epsilon^2)$.
- Precision: $V[\theta_t|\mathcal{I}_t]^{-1} := \Omega_t$. Call this a "stock of knowledge."
- Prior variance of tomorrow's state: $V[\theta_{t+1}|\mathcal{I}_t] = \rho^2\Omega_t^{-1} + \sigma_\epsilon^2$.
- 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 $n_s\sigma_s^{-2}$.

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

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

- 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. Cost accounting
2. Complementary inputs
3. Value functions
4. Revenue
5. Choice covariance
6. Intangibles approach

Measurement Approach 1: 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.

- Maybe customers are paid for data.
  Data is bartered for goods/ services.
  But that shows up as a discount in the price of the good.

- This may work well if we can impute the discount or if most of a firm's data sets are purchased.

Measurement Approach 2: Complementary Inputs

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

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

- 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 '21)

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

Measurement Approach 3: Value Function Approach

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

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

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

- The state law of motion is the depreciation equation:

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

VALUE FUNCTION ESTIMATION RESULTS ABIS, VELDKAMP '22

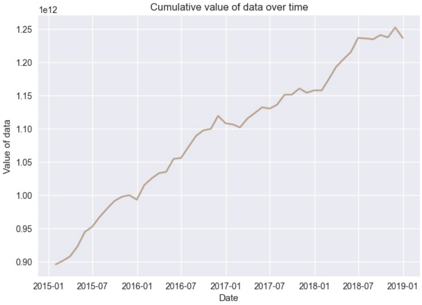


FIGURE: Estimated Value of the Aggregate Stock of Data, used for financial analysis, in hundreds of billions of current U.S. dollars, 2015-2018.

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 4: Revenue Approach

- The value of data is the pdv of the revenue it generates
- How to isolate data revenue from other revenue?
- Young, data-intensive firms may operate at a loss.
- 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 M-V investors, correlated information and learning from noisy prices:

$$\begin{aligned}\text{Value of data}_i \approx \; & \frac{\rho_i}{2}\mathbb{E}R_{t+1}{}'(\mathbb{V}\left[R_{t+1} \mid \mathcal{I}_{it}\right]^{-1} - \mathbb{V}R_{t+1}^{-1})\mathbb{E}R_{t+1} \\ & + \frac{\rho}{2}\mathrm{Tr}\left[\mathbb{V}R_{t+1} \cdot \mathbb{V}\left[R_{t+1} \mid \mathcal{I}_{it}\right]^{-1} - I\right]\end{aligned}$$

- One can estimate the conditional variances with return data, using forecasting regressions (sufficient statistics).
- 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 5: Choice Covariance

- Data allows agents to make better choices (matching and signals)
- 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.
- Measure the covariance.
  Ex: Portfolio alpha, firm vs product markups (Eeckhout-Veldkamp '22) , or a customer click-conversion rate.

Measurement Approach 6: Intangibles Approach

- Typical intangible valuation uses book-to-market values.
  Crouzet-Eberly '20, Peters-Taylor '17
  Why not do this for data?

- Intangibles include: Branding, patents, organizational capital, ...
  e.g., Belo, Gala, Salomao, Vittorino '21, Eisfeldt-Papanikolaou '13, '14

  Data may contribute to each, but is not the same.

- How to tease apart the value of data from these other intangibles in market value?

- Presumes that equity market participants know how to value data.

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:
  - Explore data supply: data markets, platforms, data ownership
  - Demand estimates and supply $\rightarrow$ equilibrium prices.
    Valuation, fluctuations, data risk premia.
  - I.O. of data markets: pricing, competition, entry
- Many important open questions to tackle.
  Join us in exploring the data economy!