The Big Data Revolution: Is Data the New Capital?

Maryam Farboodi

IIES/SNS International Policy Talks
May 2022

Why Should We Care?

I. data is a new asset class

II. what is data used for in this "big data revolution"?
   - **prediction**

III. in finance, collection, processing and use of data has exploded
   - which assets and prices are most affected?
   - how to quantify data and its value?

IV. non-financial firms increasingly use data too
   - where does the data come from? customer transactions
   - what is it used for?

V. **data market:** design and regulate functional data markets
   - understand and quantify financial and customer transaction data
   - value it!

Growth of Data Processing

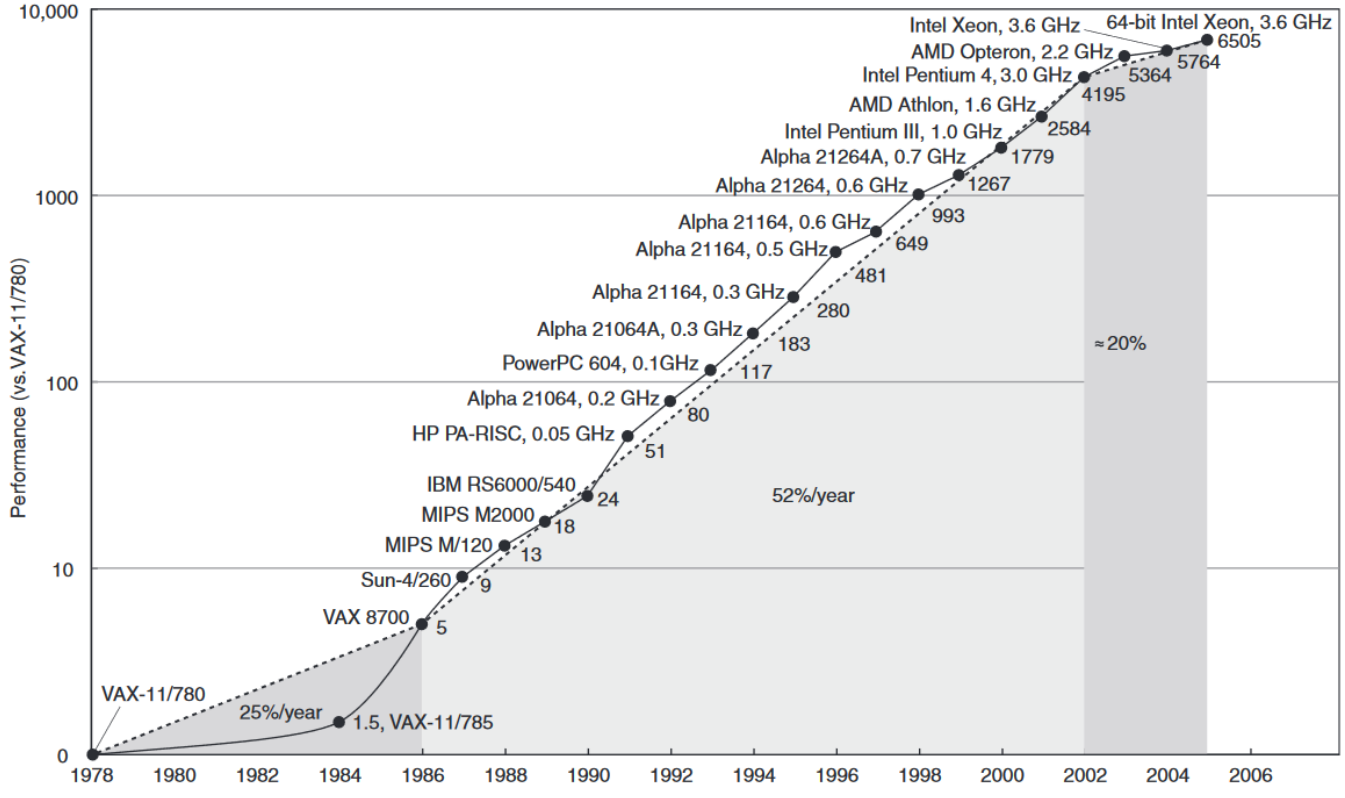


FIGURE 1.16 Growth in processor performance since the mid-1980s. This chart plots performance relative to the VAX 11/780 as measured by the SPECint benchmarks (see Section 1.8). Prior to the mid-1980s, processor performance growth was largely technology-driven and averaged about 25% per year. The increase in growth to about 52% since then is attributable to more advanced architectural and organizational ideas. By 2002, this growth led to a difference in performance of about a factor of seven. Performance for floating-point-oriented calculations has increased even faster. Since 2002, the limits of power, available instruction-level parallelism, and long memory latency have slowed uniprocessor performance recently, to about 20% per year.

Evolution of Processing Performance, 1978–2007
Hennessy and Patterson ('08)

Data is the New Oil

Figure: Economist Cover (left), Economist Inside (right)

models of manufacturing and producing tangible goods are less and less representative of the modern economy

Why does Our Understanding of the Economics of Big Data Still Lag Behind?

I. data is difficult to incorporate into our models

II. data is unobservable

III. data refers to many things!
we don't have great models to systematically disentangle data functions

IV. data and data valuation are difficult to measure and quantify

Acknowledgment

based on the body of work with amazing coauthors:

Julianne Begenau, Adrian Matray, Roxana Mihet, Thomas Philippon, Dhruv Singal, Laura Veldkamp, Venky Venkateswaran

- Long Run Growth of Financial Data Technology
- Big Data in Finance and Growth of Large Firms
- Big Data and Firm Dynamics
- A Model of the Data Economy
- Where Has All the Data Gone?
- Valuing Financial Data

Framework
Data Used for Prediction

- continuum of agents who want to make a decision under uncertainty
- uncertainty creates risk
  - risk of profit loss: producing undesirable products, taking wrong financial position
  - risk of death, injury
- data reduces the uncertainty and enables agents to make better decisions
  - (can) increase profits, (can) improve welfare
- what should we expect as data availability and technology of data collection and processing improves?
  - financial market
  - real economy
  - emergence of data markets

Data

I. **what is the data used for?**
   - prediction

II. **who are the data users?**
   - investors
   - firms

III. **what is the data used to predict?**
   - investors
     - ★ firm fundamentals
     - ★ market sentiments, demand
   - firms
     - ★ optimal production quality, customer taste

IV. **where does the data come from?**
   - information acquisition
   - firm performance history
   - customer transactions

Bayes Law
Bread and Butter in Economics of Prediction

- agents want to predict random variable z
- **data:** signals $s^1, s^2, ..., s^n$ about z

  where do signals come from? prior knowledge, information acquisition, production, God sent them, $\cdots$
- more data improves the precision of agents' pesterior belief about the random variable
- **Bayes Law: posterior precision is additive**

$$s^j = z + \epsilon^j \qquad j = 1, \cdots, n \quad \epsilon^j \sim N(0, \Sigma^j)$$
$$\Omega^j = (\Sigma^j)^{-1}$$
$$\Omega^{\text{posterior}} = \sum_{j=1}^{n} \Omega^j$$

Some Notation!

- random variables to learn about
  - y : firm fundamental, x : market demand, θ : production quality
- data
  - $\mathcal{I}_{it}$: information set of agent i at time t
  - Ω_{it} : stock of knowledge of agent i at time t (posterior precision)
- financial variables
  - R_t : asset return, d_t : firm earnings, g_t : earning growth
- real variables
  - investors: $u(.)$: utility, c_{it} : consumption, w_{it} : wealth, q_{it} : portfolio allocation
  - firms: k_{jt} : capital, A_{jt} : productivity/quality
- prices
  - p_t : asset/good price, π_t : data price
- expections $\mathbb{E}[.]$ and variances $\mathbb{V}[.]$, conditional and unconditional

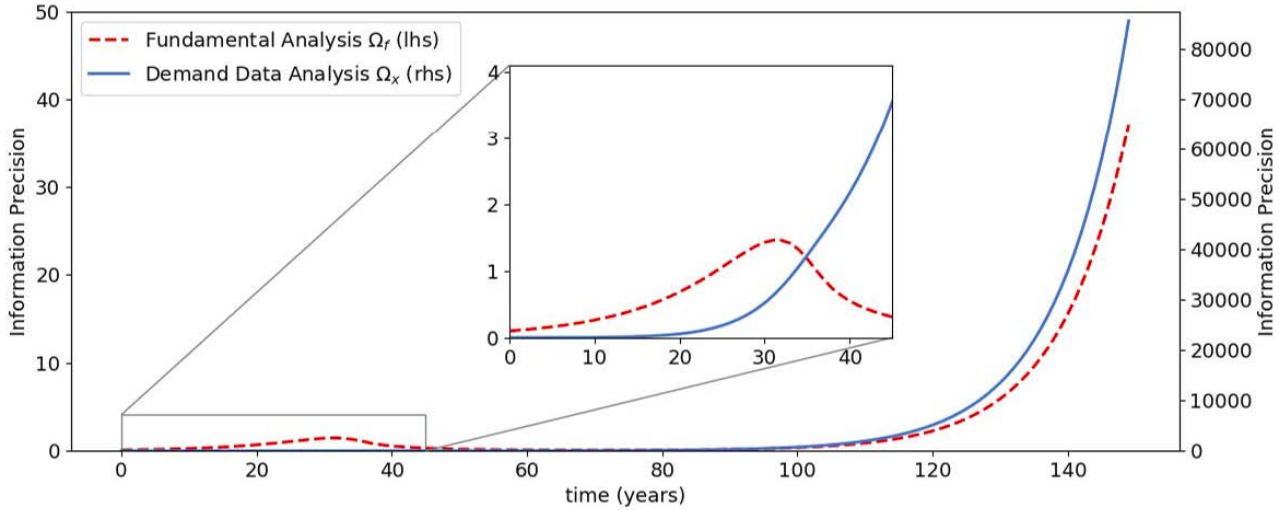
Outline

Financial Markets. Aggregate Trends

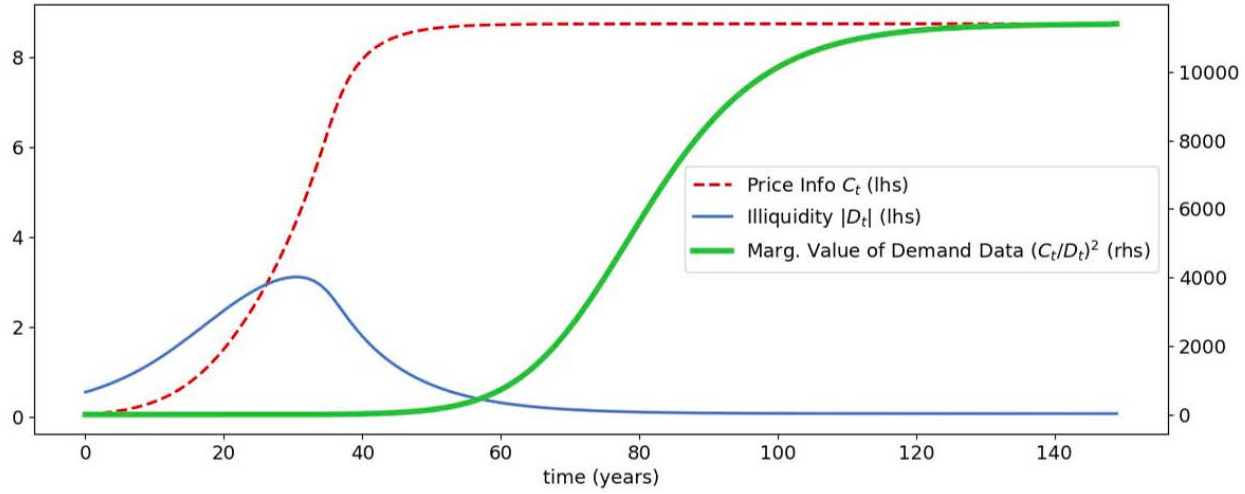
- aggregate consequences of technological progress in data analysis in financial markets
- growth theory describes how technology boosts efficiency.
  but in finance, technology (IT) is blamed for volatility, illiquidity and inefficiency SEC ('15), Ben-David et al ('12), Zhang ('06)
- concern: big data is changing not only how much data we see, but also what kinds of data we choose to use
  - big data can predict asset payoffs, or market demand/sentiments
- **findings**
  - different phases of data analysis: first fundamental analysis, followed by demand/sentiment analysis, finally balanced growth
  - aggregate price informativeness grows
  - market becomes illiquid before reverting and becoming more liquid

Growth of Financial Data Processing

- phases of data analysis



- price informativeness and liquidity

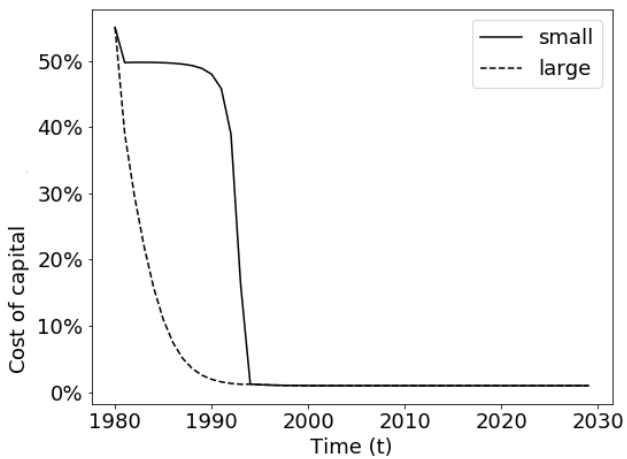


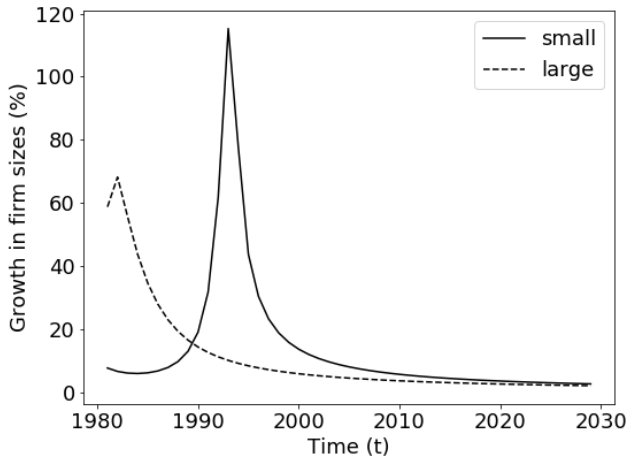
Financial Markets. Cross Sectional Trends

- large degree of heterogeneity in cross-section of firm
- S&P500 price informativeness has improved over time
  but average price informativeness over all public firms has deteriorated!
- have all the firms benefited the same from technological progress in data processing?
- **finding:** divergence in data and informational efficiency of prices
  - most data processing by investors is about *large growth* firms
  - why? investors process that that is most valuable to them
  - size and growth interact to make data more valuable
  - measuring investor data

Spillover from Financial Markets to Firm Distribution

- small firms are being displaced by larger ones
- **big data** benefits growth of large firms disproportionately
  - data comes from economic transactions
  - big firms, with many transactions, produce a lot of data
  - big data in financial markets systematically changes how large and small firm capital is priced: **cost of capital**





key mechanism: data resolves risk $\Rightarrow$ lower risk reduces risk premium

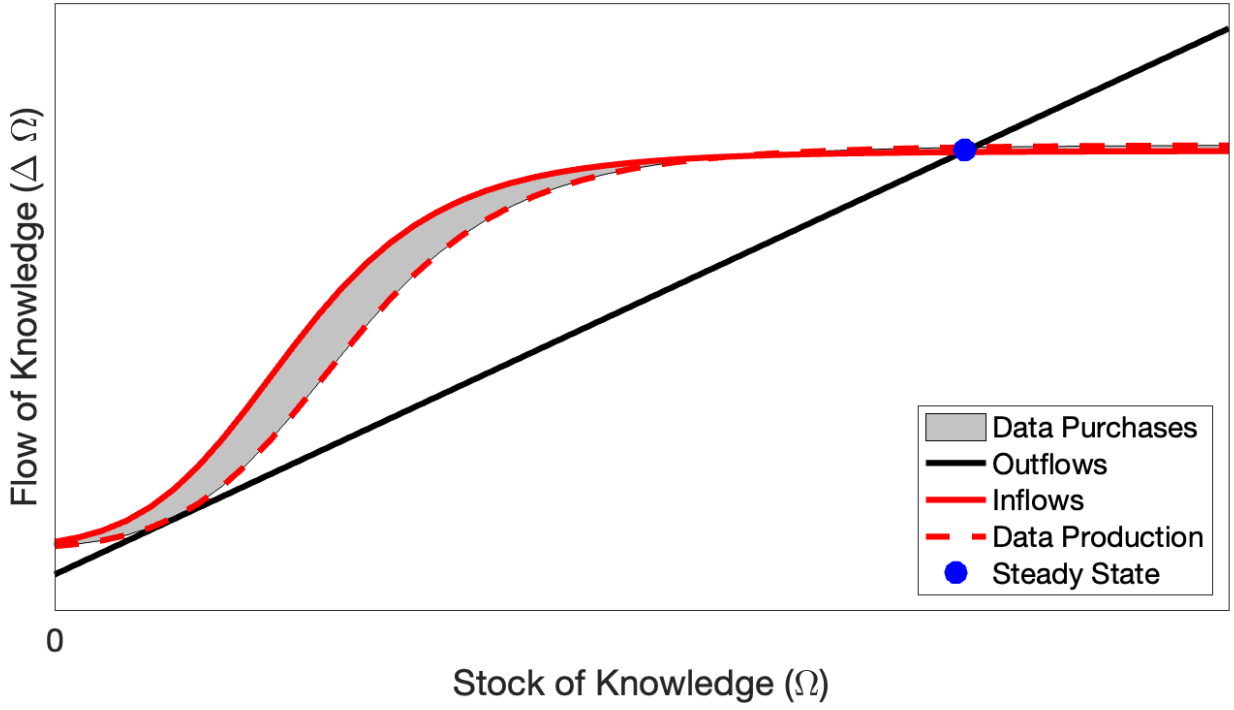
$\Rightarrow$ cost of capital falls $\Rightarrow$ firm grows more

Data Production by Firms

- data is a byproduct of economic transactions
- firms use data to **predict** optimal production technique and improve quality of produced goods
- **data in non-exclusive** (non-rival)
  - firms can sell data and still use some of it
  - key difference with other factors of production
- data is tradable among firms
  - key difference with learning-by-doing
- firms can use data to produce high quality goods and services, and/or sell it and make profits from data sales

Long-Run and short-Run Trends

- whether data sustains growth or not depends on data usage
  - process optimization
  - innovation
- data used for process optimization: S-shaped growth



- short run
  - data barter

Data Markets

- **data non-exclusive**
  - without frictions, data markets are active even in steady state
  - firms voluntarily participate in data market
  - policy encourages data sharing too!
- **data intermediaries** emerge
  - firms with better data processing technology
  - larger
  - specialize in data market, as opposed to high-quality goods production
- new entrants rely on the data market to facilitate growth

Outline

Measuring Financial Data

- group stocks into four groups j:
  {Small-Growth, Large-Growth, Small-Value, Large-Value}
- informativeness of stock prices

$$\text{price informativeness}_t^j = \underbrace{\frac{\Sigma_d^j}{StdDev(p^j)}}_{\text{volatility}} \underbrace{\frac{g^j}{r - g^j}}_{\text{growth}} \underbrace{\left[1 - \frac{\Sigma_d^{j\,-1}}{\overline{\Omega}^j}\right]}_{data}{}^t$$

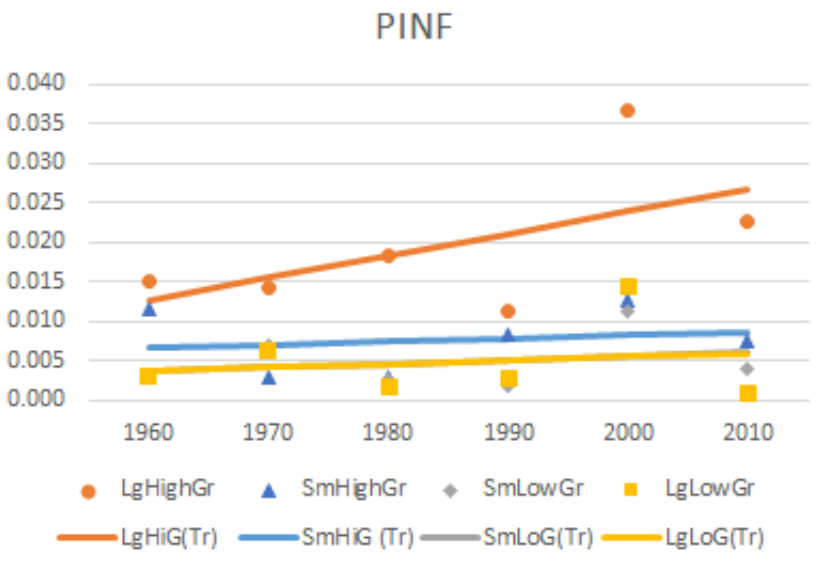
- estimate for each decade (stock f, group j, time t):

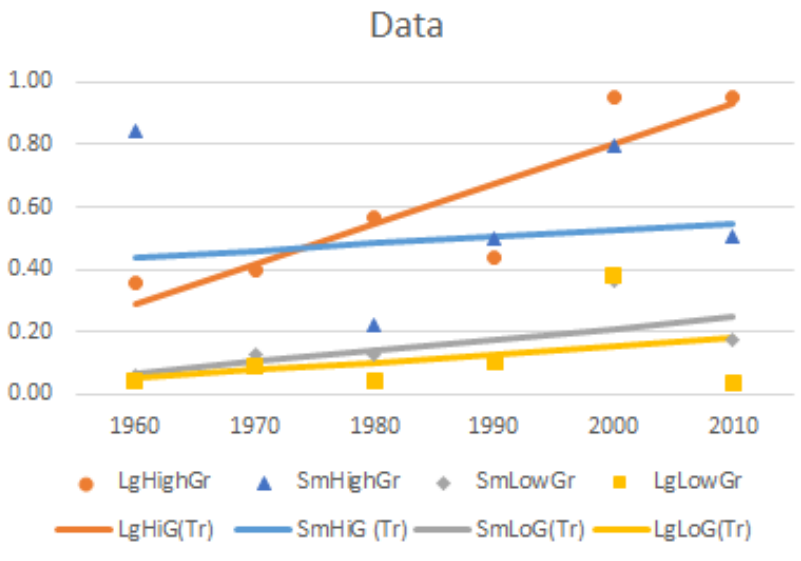
$$\frac{EBIT_{fjt+s}}{ASSET_{fjt}} = \alpha_{js} + \beta_{js} log\left(\frac{MKVAL_{fjt}}{ASSET_{fjt}}\right) + \gamma_j X_{fjt} + \epsilon_{fjts}$$

- price informativeness

$$PINF_{js} = \beta_{js} \cdot StdDev\left(\frac{MKVAL_{fjt}}{ASSET_{fjt}}\right)$$

Financial Data





cross-sectional divergence in financial data

Measuring Value of Data to Investors

- very heterogeneous: investor characteristics, stock characteristics, market conditions
- **statistical approach**
- ex-ante expected utility of data for an investor

$$U(\mathcal{I}_{it}) = \underbrace{\mathbb{E}\left[R_t\right]' \hat{\mathbb{V}}_i^{-1} \mathbb{E}\left[R_t\right]}_{\text{(sharp ratio)}^2} + \text{Tr}\left[\underbrace{\left(\mathbb{V}[R_t] - \mathbb{V}[R_t \mid \mathcal{I}_{it}]\right)}_{\text{data:variance reduction}} \hat{\mathbb{V}}_i^{-1}\right] + r\rho_i \bar{w}_{it}$$

expectations and variances:
complex functions of parameters, easy to measure!

Value of Financial Data

Dollar value of data (in $1000) affected by

- investor wealth
- investment style
- **market liquidity**

	Portfolio Type					
	Small	Large	Growth	Value	S&P500	All
Panel A: Perfect Competition						
Investor with $500,000 Wealth	0.00	1.67	2.49	0.49	1.90	3.50
Investor with $250m Wealth	0.00	566.41	844.09	164.71	643.62	1188.50
Panel B: With Price Impact						
Investor with $500,000 Wealth	0.00	1.65	2.48	0.41	1.83	1.38
Investor with $250m Wealth:)	0.00	23.93	57.00	1.45	15.98	253.62
Time Periods	31	31	31	31	31	31

Outline

Data as a Traded Product

- market for data is on the rise
- regulation is also on the rise!
- **data intermediaries**
  - data brokers sell consumer data to firms
  - open banking
  - firms buy transaction data statistics from data intermediaries such as Amazon
- price elasticity of demand: marginal value of consumer data to firms

Data Processing Technology as a Product

- market for digital services is also growing
- **digital intermediaries:** large tech firm like Amazon, Google and Microsoft
  - large investment in digital infrastructure
  - rent out cloud storage and computing to other firms
  - build an ecosystem
- how does it affect entry? reduces barriers to entry for startups by transforming fixed entry cost to variable cost
- non-competitive behavior

Concluding Remarks. The Big Data Research Agenda

- numerous shifts in the financial and real sector are a logical consequence of improvement in data processing
- data is changing how firms operate: "Data Is the New Oil"
- data measurement is far from obvious

Technology is transforming markets. We need theory and measurement to make sense of a constantly evolving landscape!