Data and the Aggregate Economy

Laura Veldkamp

Columbia, NBER, and CEPR

lv2405@columbia.edu

2019 Meeting of Society for Economic Dynamics

Data in the Aggregate Economy

- Data can be encoded as binary sequence of zeroes and ones.
  - This includes music, poetry, patents ... not my focus
  - My focus: statistics and *records of transactions*.
- Entering a new era: Big data is the industrialization of knowledge production
  - Replaces artisanal production and mitigates diminishing returns to data.
- When we transitioned from agrarian to industrial, we replaced land with capital in production functions.

  As we transition again, *what if production includes data?*

Agenda

1. **Data is changing every corner of macroeconomics**
    - A quick tour through many facts/questions.
2. **Existing theoretical tools to understand data**
    - How data is similar to and different from technology?
    - Relevant tools from information frictions literature
3. **Framework to explore effect of data on aggregate economic activity**

GDP Measurement

- Measured data GDP is large and growing: information and communications tech (ICT) $\equiv$ 6.5% of global GDP, $\sim$ 100 million workers (UN (2017))
- Most data is not in GDP because it's bartered.
  - Consumers give data in exchange for a service.
  - Not priced $\rightarrow$ Not captured by GDP

Data in International Economics

- Will the data economy negate geography?

  "... [W]e are determined to take full advantage of the digital revolution. This revolution is summed up by the fact that it no longer is of utmost importance where you are but rather what you can do ..."

  — *Rwandan President Paul Kagame (2010)*

- Control of data is a significant issue in trade negotiations. Are we exporting data assets at zero price to foreign entities?

Data Creates Price Flexibility

- Digital economy changes how firms adjust prices.
- Price changes occur more 4-5 times more frequently in online stores than in regular stores (Gorodichenko and Talavera (2017))
- Might more price flexibility reduce monetary policy effectiveness?

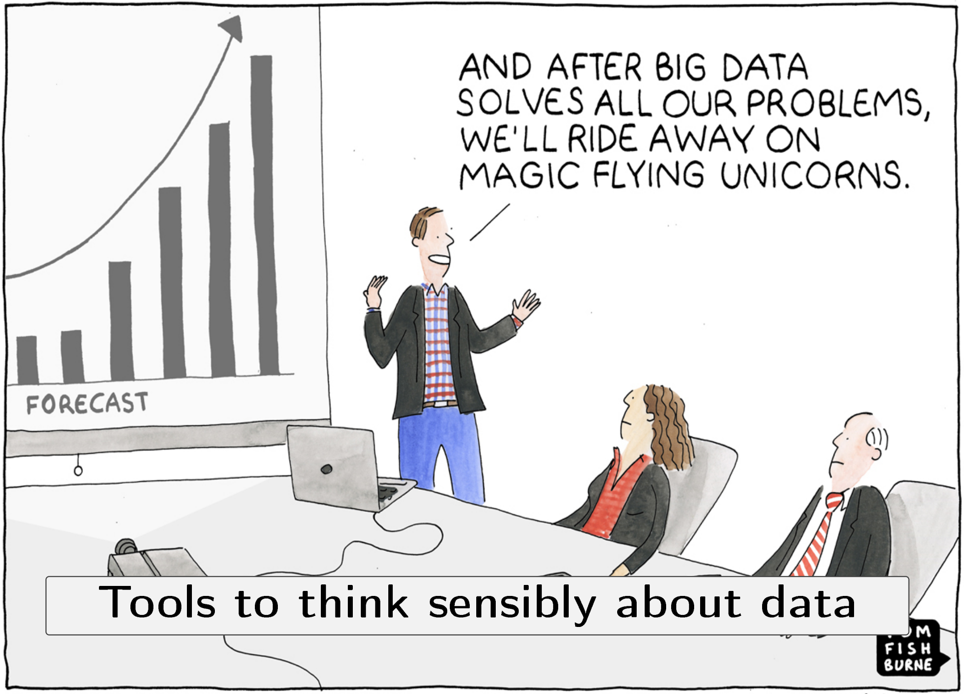
Data and Labor

- Will artificial intelligence displace labor?
  (Acemoglu and Restrepo (2018))

- Will machine learning reduce search frictions?

Data Changes Firm Dynamics

- IT investment boosts firms' efficiency (Aral and Weill (2007)), especially benefits large firms
  (Tambe and Hitt (2012), Brynjolfsson and McElheran (2016))
- Data in finance also benefits large firms
  - Big data in finance lowers the cost of capital for large firms most (Begenau, Farboodi, Veldkamp (2018))
- Market power: Is data a barrier to entry?



Tools to think sensibly about data

© marketoonist.com

Employing Ideas from Growth Theory

- Data enhances productivity, is non-rival, has returns to scale.
  (Jones and Tonetti (2018), Agrawal et al. (2018), Lu (2019), Aghion et al. (2017))
  Current thinking: This makes data just like ideas.

- Data can also be traded like ideas/patents.

- But ... While technologies leak (Easterly (2002)), data is not embodied in human capital. Less likely to diffuse.

Ideas From Information Frictions: Where does data come from?

- Data is a by-product of economic activity
  Veldkamp (2005), VNV (2006), Ordonez (2013), Fajgelbaum, Schaal, T-D (2017)
  - had aggregate productivity data.
    I'll adapt these to make data good-specific.
  - Unlike tech, incentives/rents not required to produce.
- Big data (ML, AI) is used for prediction / forecasts.
  - Incentives to acquire information for prediction:
    Grossman-Stiglitz (1980), Sims (2003), Hellwig-Veldkamp (2009)
- Talking about data overcomes 3 barriers to information literature:
  1. We can't measure information. Bytes are countable.
  2. We can't integrated it into standard recursive dynamic models.
  3. What important macro question can't be answered without it?

Modeling Data

- Why model data?
  - Macroeconomists should have something to say about the information economy. It's an aggregate phenomenon.
  - Answer policy questions: regulation, tariffs.
  - Learn how to value data.
- What features does a macro-data theory need?
  - Data is a by-product of economic activity.
  - Firms with data make better forecasts, are more productive.
  - Data is a valuable asset.

Main results

1. Data has diminishing returns.
   Data cannot sustain growth without technological progress.
2. Data barter.
3. Cross-firm inequality rises, then falls.
4. Some firms specialize in data production.

A Model of Data as a Productive Asset

A Macro Model of Data

Farboodi & Veldkamp (WP 2019)

- A continuum of competitive firms i
- Each firm produces $k_{i,t}^{\alpha}$ units of goods with $k_{i,t}$ units of capital
- 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: How Data Improves Goods Quality

- Quality depends on chosen production technique $a_{i,t}$
- Firm has one optimal technique: $\theta_{i,t} + \varepsilon_{a,i,t}$
  - $\theta_{i,t}$ is an AR(1), with innovation $\eta_{i,t} \sim N(\mu, \sigma_\theta^2)$.
  - $\varepsilon_{a,i,t} \sim N(0, \sigma_a^2)$ is unlearnable and i.i.d.

$$A_{i,t} = \widehat{A} - (a_{i,t} - (\theta_{i,t} + \varepsilon_{a,i,t}))^2$$

- Data is a by-product of $t-1$ production: $n_{i,t} = z_i k_{i,t-1}^\alpha$
  z is data-mining ability.

- Each data point $m \in [1 : n_{i,t}]$ reveals
  $s_{i,t,m} = \theta_{i,t} + \epsilon_{i,t,m}$ where $\epsilon_{i,t,m} \sim N(0, \sigma_\epsilon^2)$,
- Data adjustment cost: $\Psi(\cdot)$.

A Simple Recursive Solution

- **State variable: Stock of knowledge**

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

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

[redacted]

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

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

where $n_{i,t+1} = z_i k_{i,t}^{\alpha}$ and (Kalman filter)

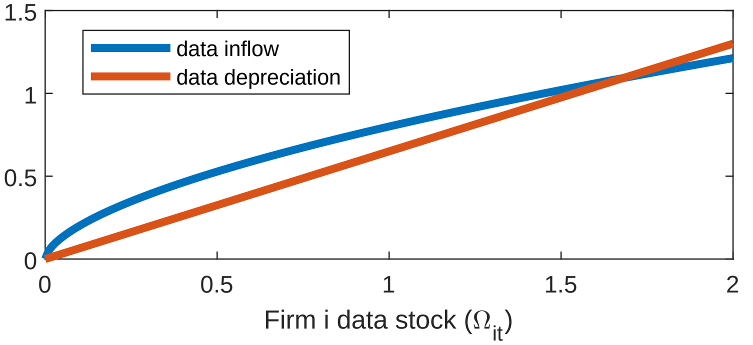
$$\Omega_{i,t+1} = \left[\rho^2(\Omega_{i,t} + \sigma_a^{-2})^{-1} + \sigma_\theta^2\right]^{-1} + n_{i,t+1}\,\sigma_\epsilon^{-2}$$

Pricing and Valuing Data

- Firm's stock of knowledge, $\Omega_{i,t}$, is a discounted sum of data "investments."
- $\partial V_t / \partial \Omega_{i,t} =$ Marginal value of one additional piece of data (additional unit of precision).
  - $=$ price of data
- The price of goods may be zero (data barter)
  - Firms still produce to generate transactions and accumulate data: $V'(\Omega_{t+1}) > 0$.

Data Growth, Solow-Style

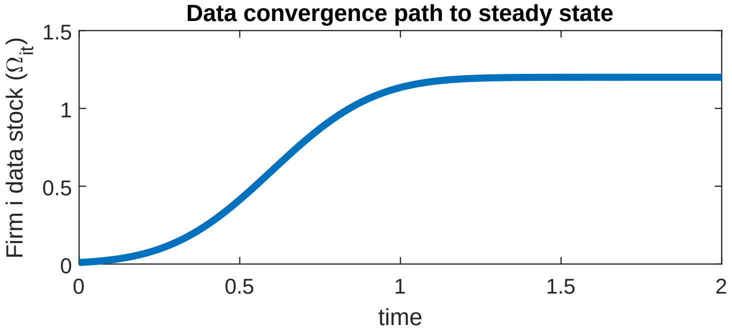
- Inflows: $z_i k_{it}^{\alpha} \sigma_{\theta}^{-2}$ (# of data points $\times$ precision)
- Outflows: data depreciation
- Steady state: inflows $=$ outflows



To overturn concavity you'd need: Infinite data $\rightarrow$ certainty about tomorrow's state $\rightarrow$ infinite production.

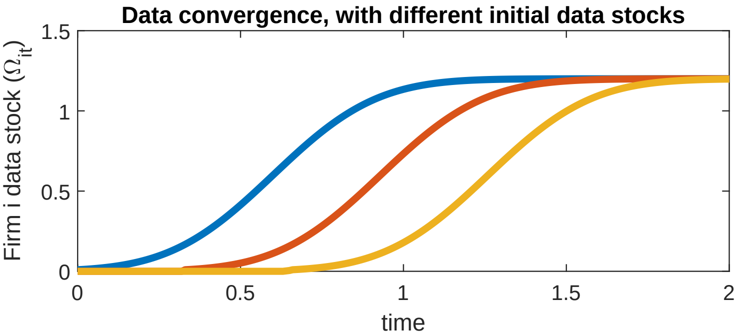
How Firm's Data Stock Grows Over Time

- **Takeaway:** Data has increasing, then diminishing returns (Bajari et al. (2018)).
  Data → more output → more data, until forecast error → 0.



How Firms' Data Stock Grows Over Time

- Three firms with different initial data stocks.



Divergence, then convergence. Inequality is transitory?

Tradeable Data (one firm, exclusive use, quadratic adj cost)

- Use precision ω at price π:
  $\omega < (>) \, n$ is selling (buying) data.

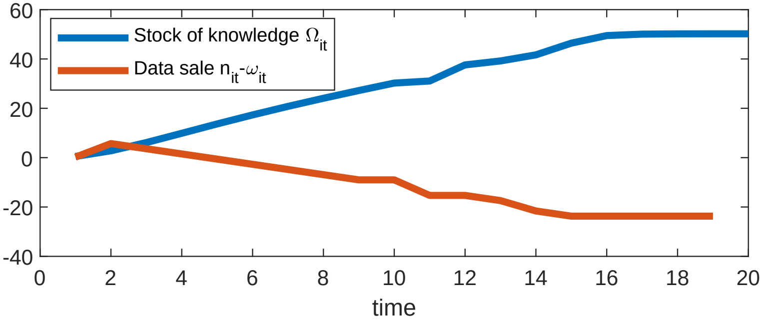
$$V(\Omega_{i,t}) = \max_{\{k_{i,t},\omega_{it}\}} P_t\left(\overline{A} - \Omega_{i,t}^{-1} - \sigma_a^2\right) k_{i,t}^{\alpha} + \pi\left(n_{i,t+1} - \omega_{it}\right)$$

$$- \Psi(\Delta\Omega_{i,t+1}) - rk_{i,t} + \beta V\left(\Omega_{i,t+1}\right)$$

$$\Omega_{i,t+1} = \left[\rho^2(\Omega_{i,t} + \sigma_a^{-2})^{-1} + \sigma_\theta^2\right]^{-1} + \omega_{it}$$

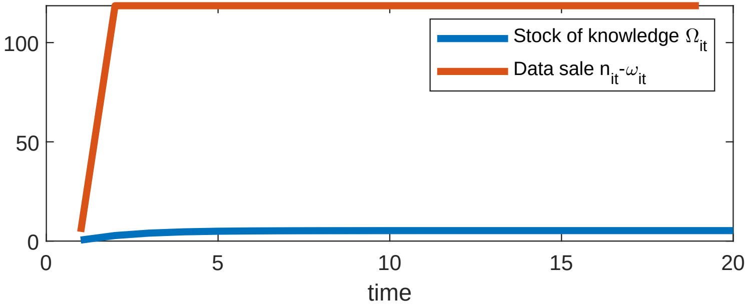
- Without adjustment frictions ($\Psi = 0$), immediate convergence.

Tradeable Data:
Unproductive data miner ($z_i = 1$) sells then buys data



Early data sales avoid data adjustment cost. Firm doesn't have systems in place to use their own data.

Tradeable Data:
Data miners ($z_i = 15$) sell most of their data



Data miners don't use much data. They don't need high-quality goods. Most revenue from data sales.

Data as a Business Stealing Technology

- Lots of data used for advertising. Maybe not socially productive. Does this overturn results?
- Suppose data processing helps the firm that uses it, but has no social value:

  (Morris-Shin (2002))

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

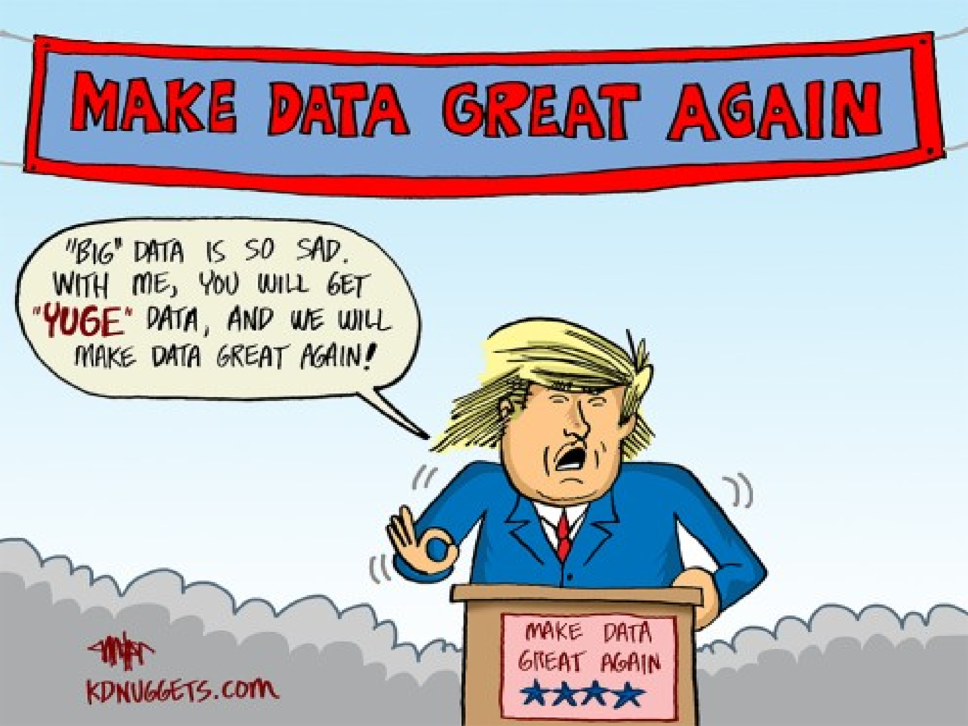
- Result: Same results as figures above. Logic holds for either case. This doesn't change firms' choices or firm dynamics. Does affect aggregate quality and welfare.

Many ways to enrich this model

- Non-rival data requires imperfect competition.
  (See Jones-Tonetti, 2019; Argente-Moreira-Oberfield-Venkateswaran, 2019.)
- Data markets: high-demand data has lower price (Veldkamp 2006) or data provider chooses what to provide (Nimark, 2014)
  $\rightarrow$ Realistic business cycle variance and covariance.
- Choice of what data to purchase / process (θ vector):
  Rational inattention tools originally developed for data allocation.
  - Sims (2003), Mackowiak-Wiederholt (2009), Woodford(2009), Matejka-McKay(2015), Kacperczyk-VNV (2016)...
- Agents use data to learn about the distribution of shocks (Kozlowski-Veldkamp- Venkateswaran, 2019)
  $\rightarrow$ persistent responses to unusual events.
- Firm entry / exit ... and many others

Conclusions

- Data is an important asset in the modern economy ... like capital.
- Has some features of tech: non-rival, raises productivity.
    - Key difference: As forecast errors get smaller, most information has decreasing returns.
- Data may be changing the allocation of rents in the economy ... but not its fundamental economics.
- Open questions for theory and data:
    - How much of the superstar firm phenomenon is due to data economics?
    - Has data changed the returns to capital / labor?
    - How do we value firms whose main asset is data?
    - ... and many more


MAKE DATA GREAT AGAIN
"BIG" DATA IS SO SAD. WITH ME, YOU WILL GET "YUGE" DATA, AND WE WILL MAKE DATA GREAT AGAIN!
MAKE DATA GREAT AGAIN
KDNUGGETS.com

Questions & Comments