A Model of the Data Economy

by M. Farboodi and L. Veldkamp

Discussion by Edouard Schaal

CREI, ICREA, UPF and BGSE

ECB Annual Research Conference - Sep. 12-13 2022

Quick Summary

- This paper asks an important question:
  - How is Big Data transforming our economy?

- The paper attempts to provide a basic framework to understand short- and long-run issues
  - Considers Big Data as information used in forecasting
    - Key idea: transactions generate data that can raise productivity
    - Allows firms to trade data
  - Very tractable model that can be easily embedded in general DSGE model
  - Delivers a number of predictions broadly in line with the data:
    - Changes in firm dynamics (increasing then decreasing returns, firm entry...)
    - Negative profits (IRS), possibility of data barter
    - Issues of measurement (value of data, missing GDP)
    - Long-run: Can data generate long-run growth?
    - Efficient benchmark, welfare implications with business-stealing externality

- Thought-provoking paper:
  - ▶ Nice and clean tractable model to formalize discussions on the topic
  - ▶ Not exhaustive, may be missing important elements but easy to extend the model
  - ▶ Great tool for other researchers to build upon
- The paper has already been around for a few years
  - ▶ My comments are merely suggestions...

Comments

- The paper is quite theory-driven and predictions are quite model-dependent
    - ▶ Natural modeling choices for economists working on information
    - ▶ ex.1: why do data affect the *level* of TFP, rather than its *growth*?
    - ▶ ex.2: is transaction-based data feedback strong in practice?
    - ▶ ex.3: do data face strong diminishing returns? (Gaussian vs. fat-tail)
    - ⇒ These modeling choices should be guided by empirics
- In practice:
    - ▶ Which firms/sectors use Big Data?
    - ▶ From what sources?
    - ▶ How do firms actually use data?
    - ▶ Strength of data feedback, IRS/DRS?
- Data is scarce and anecdotal
    - ▶ Some suggestive evidence from Rabano-Suarez (2022)

- Sources:
  - France: INSEE TIC 2016 survey of 12,700 businesses with 10+ employees
  - US: 2019 Annual Business Survey from Census Bureau
- Big Data: "*massive datasets whose large volume, flow and complex structure cannot be dealt with by traditional data-processing softwares and require the use of specific techniques*"

Businesses Using Big Data in 2015 and Data Sources		Data sources			
Sector	% Adoption of Big Data	Geolocalization	Social Networks	Connected objects/sensors	Other
Transport	25	92	7	20	4
Information and communication	23	33	64	38	26
Administrative and support services; real estate activities	13	75	34	14	8
Construction	11	89	13	14	4
Specialized scientific and technical activities	11	43	43	36	18
Commerce	9	50	34	35	8
Accommodation and catering	9	49	76	25	16
Manufacturing	8	51	25	46	10
Total (France)	**11**	**62**	**32**	**29**	**10**
Source : Insee, enquête TIC 2016.					

- Adoption of Big Data varies a lot by sector:
  - ▶ transport and IT are the main users (23-25%), manufacturing (8%)
- Big Data comes from varied sources:
  - ▶ geolocalization is the main source (62%), then social networks
  - ▶ not all sources are necessarily transaction based...

- Different sectors use different kinds of data

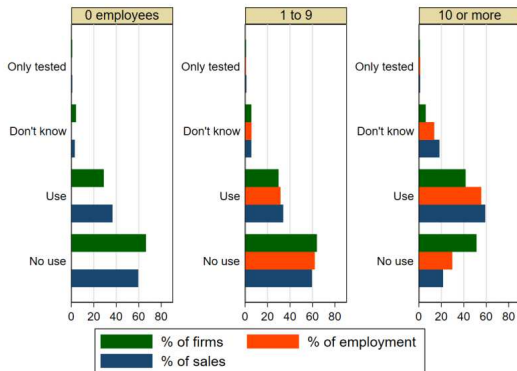- Large firms are more likely to adopt Big Data:

Figure: Adoption of Big Data by number of employess in the US (Rabano-Suarez, 2022)

- In line with paper's conclusion? Also suggestive of large setup costs

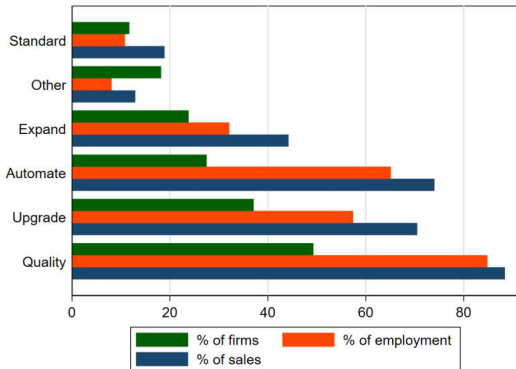Figure: Motivation to adopt AI in the US (Rabano-Suarez, 2022)

Comments: Uses of Big Data

Main uses of Big Data:

- **Marketing and advertising** ("Expand" ∼40%)
  - ▶ e.g. targeted advertising through social media
  - ⇒ Big Data as improving targeted search (Wu, 2020)? product awareness (Perla, 2019)? or advertising (Cavenaile and Roldan-Blanco, 2021)

- **Production process improvement** ("Automate", "Upgrade" ∼60%):
  - ▶ ex.1: geolocalizing drivers to streamline delivery process
  - ▶ ex.2: feedback from customer services for spare part needs and better product design
  - ⇒ Big Data as R&D, possibly growth effect (Lucas and Moll, 2014;...)

- **Quality upgrade** ("Quality" ∼80%):
  - ▶ Improve product design to fit customer's tastes
  - ▶ e.g. Netflix "We know what people watch on Netflix and we're able with a high degree of confidence to understand how big a likely audience is for a given show based on people's viewing habits," (WIRED, 2013)
  - ▶ e.g. Cheesecake Factory's guest forecasting (New Yorker, 8/6/12)
  - ⇒ what this paper does

- Adoption of Big Data is large and does seem to affect businesses importantly
  - ▶ But the phenomenon is varied and could be modeled in different ways
    - The welfare impact will depend on the specific modeling choice
  - ▶ Not all data is necessary linked to past transactions
    - weaker feedback from economic activity to data (IRS)?
    - two-stage process: data-barter, then monetization?
- The modeling approach is closer to the quality upgrade channel
  - ▶ Perhaps be more specific about the mechanism
  - ▶ Perfect competition model with TFP effects remains a little abstract

  $$A_{it} = g\left(\left(a_{it} - \theta_t - \varepsilon_{ait}\right)^2\right)$$

    - Where does g come from? What should it look like? $\Rightarrow$ empirics
  - ▶ Perhaps use a fully micro-founded model with differentiated products and quality differentiation?
    - Full welfare analysis, different nature of competition between varieties
- Paper discusses the value of data and missing GDP
  - ▶ But how can we use this model in practice for the measurement? Implementation?

Nerdy Comments

- Market for data:
  - ▶ Can firms buy and sell data at the same time?
    - Without the ι cost, could firms buy and sell all available data to learn the truth?
  - ▶ Should firms behave strategically in selling data to competitors?
  - ▶ Why model sale of data as a decrease in own-data stock?
    - Could also appear discounted by others without having kink if value function?
  - ▶ If same signal can be traded multiple times, idiosyncratic noise could create correlation across beliefs and not vanish at the aggregate
    - Should be accounted for in equilibrium definition or ruled out
    - Possibility of herding if firms fail to control for this correlation?
    - ⇒ A model where same info can be traded multiple times needs to keep track of what info is being traded...
    - Need to be more explicit about assumptions that underlie equilibrim definition

- Inspiring paper to lay out foundations for the debate!
  - ▶ Already an important milestone in this new agenda
- Framework that can be extended in multiple ways:
  - ▶ Other impacts of info (R&D, advertising...)
  - ▶ Firm entry/exit and market concentration
  - ▶ Market power (Eeckhout and Veldkamp, 2022)
  - ▶ Labor market impact (skill premium, sorting)
  - ▶ Property rights, privacy and regulation (Jones and Tonetti, 2020)
  - ▶ Other externalities...