The Changing Economics of Knowledge Production

Simona Abis and Laura Veldkamp

解读：雷印如

2021 年 11 月 28 日

Motivation

- Big data technologies change the way in which data and human labor combine to create knowledge.
- This nature of this technological shift is similar to industrialization, but the process of data knowledge is ambiguous.
- How valuable are the data and the AI technologies?

Literatures-AI

- A handful of recent working papers investigate how machine learning and artificial intelligence are affecting labor demand.
  - Acemoglu and Restrepo (2018), Babina et al.(2020) and Deming and Noray (2018)
  - Cockburn et al. (2018) and Alekseeva et al. (2020)
- Others examine the productivity gains or potential discrimination costs that follow the adoption of AI
  - Credit (Fuster et al. (2018))
  - Equity analysis (Grennan and Michaely (2018))

Literatures-Data models

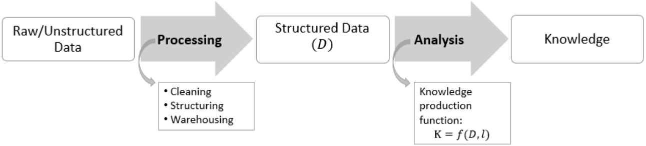
- Models of the role of data in the process of economic growth equate data and knowledge, without unpacking how raw data is transformed into that valuable output-enhancing knowledge.
    - Jones and Tonetti (2018), Agrawal et al. (2018a), Aghion et al. (2017) and Farboodi and Veldkamp (2019))
- Berg et al. (2018) explore models with different elasticities of substitution between robots and manual workers, focus on a different topic.

Contribution

- This paper contributes a structural, production function approach.
- Our emphasis, on how inputs combine to create knowledge, is complementary to such studies that examine the outputs and effects of machine learning.
- This study unpacks how raw data is transformed into that valuable output-enhancing knowledge.

A Model for Measurement

- There are three types of workers: AI analysts, old technology (OT) analysts, and data managers.
- It needs to relate hiring to labor as well as quantities and prices of labor to data stocks and knowledge production.



A Model for Measurement

- The new technology knowledge production function is:

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

  where L_{it} is labor input for data analysts with machine learning skills

- The old technology knowledge production function is:

$$K_{it}^{OT} = A_t^{OT} D_{it}^{\gamma} l_{it}^{1-\gamma} \tag{2}$$

  where l_{it} is labor input for data analysts with traditional analysis skills

A Model for Measurement

- We use a Cobb-Douglas production function for knowledge because
  - It offers a clear mapping between incomes shares and the production function parameters
  - It facilitates our comparison between new data technologies and the changes induced by industrialization.

Data management and Data Stocks

- Data inputs for analysis are not raw data. They need to be structured, cleaned and machine-readable. This requires labor.

$$
\begin{aligned}
D_{i(t+1)} &= (1-\delta)D_{it} + A^{DM}\lambda_{it}^{1-\phi} \\
&= D_{i0}(1-\delta)^t + \sum_{s=0}^{t}(1-\delta)^{t-s}A^{DM}\lambda_{is}^{1-\phi}
\end{aligned} \tag{3}
$$

where λ_{it} is labor input for data managers and A^{DM} is the productivity of data manager (DM) labor.

Equilibrium

- We are interested in a competitive market equilibrium where all firms choose the three types of labor to maximize firm value.

$$
\begin{aligned}
v(D_{it}) = \max_{\lambda_{it}, L_{it}, l_{it}} & A_t^{AI} D_{it}^{\alpha} L_{it}^{1-\alpha} + A_t^{OT} D_{it}^{\gamma} l_{it}^{1-\gamma} \\
& - w_{L,t} L_{it} - w_{L,t} l_{it} - w_{\lambda,t} \lambda_{it} + \frac{1}{r} v(D_{i(t+1)})
\end{aligned}
\tag{4}
$$

where $D_{i(t+1)} = (1-\delta) D_{it} + A^{DM} \lambda_{it}^{1-\phi}$

Optimal firm hiring and wages

- F.O.C with respect to AI analyst labor is

$$(1-\alpha)K_{it}^{AI} - w_{L,t}L_{it} = 0$$

- The F.O.C with respect to old tech analyst labor is

$$(1-\gamma)K_{it}^{OT} - w_{l,t}l_{it} = 0$$

- Taking the ratio of F.O.C implies that

$$\frac{(1-\alpha)K_{it}^{AI}}{(1-\gamma)K_{it}^{OT}} = \frac{w_{L,t}L_{it}}{w_{l,t}l_{it}}$$

Optimal firm hiring and wages

- F.O.C with respect to data management labor is

$$\frac{1}{r}v^{'}(D_{i(t+1)})(1-\phi)A^{DM}\lambda_{it}^{-\phi}=w_{\lambda,t}$$

- The F.O.C with respect to old tech analyst labor is

$$\frac{(\alpha K_{it}^{AI}+\gamma K_{it}^{OT})(1-\phi)}{r-(1-\phi)}\frac{D_{i(t+1)}-(1-\delta)D_{it}}{D_{it}}-w_{\lambda,t}\lambda_{it}=0$$

- This yields an expression of the data stock

$$D_{it}-\frac{(\frac{\alpha}{1-\alpha}w_{L,t}L_{i,t}+\frac{\gamma}{1-\gamma}w_{l,t}l_{i,t})(1-\phi)}{r-(1-\gamma)}\frac{A^{DM}\lambda_{it}^{-\phi}}{w_{\lambda,t}}=0$$

Data and Estimation

Our model is about knowledge production generally, in any industry. But we use asset management industry labor and data estimates.

- Why look at the investment management industry?
  1. The investment management industry is primarily a knowledge industry
  2. Finance is an early adopter of AI and big data technology
  3. The financial industry is a useful laboratory because finance jobs are typically filled.

Labor demand

Job postings data set collected by Burning Glass, from January 2010 through December 2018.

- Subset the BG data to jobs in the financial industry
  - Jobs' NAICS, O*NET and proprietary BG codes to restrict in the financial industry.
- Identify investment management skills
  - Construct a list of investment management skills
- Assign all jobs to unique employers
- Keep only job postings from employers that hire in investment management.

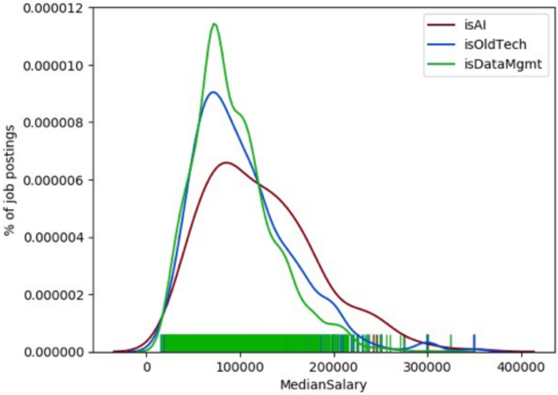
The total number of employer-month observations is 33,610

Labor demand

- There are three types of workers: AI analysts, old technology (OT) analysts, and data managers.
- The keywords specific to the assigned category have a significantly higher relevance.

Wages

- We typically use the median of the salary range listed as the salary for that job.
- AI jobs clearly pay more than traditional analyst jobs.



Cumulating hiring to get labor

- Each month, the BLS reports the job posting, job filling and separation rate for each occupation.
- We multiply each job posting number by the fraction of job postings that results in a new hire (h)

$$L_{it} = (1 - s_t^{AI})L_{i(t-1)} + j_{it}^{AI} h_t^{AI}$$

$$l_{it} = (1 - s_t^{OT})l_{i(t-1)} + j_{it}^{OT} h_t^{OT}$$

$$\lambda_{it} = (1 - s_t^{DM})\lambda_{i(t-1)} + j_{it}^{DM} h_t^{DM}$$

we start the initialization from zero for all job types and we use the first 5 years of data [2010-2014] as a burn-in period

Get structured data stocks

- We measure each firm's stock of data in each period

$$D_{i(t+1)} = (1-\delta)^t D_{i0} + \sum_{s=0}^{t} (1-\delta)^{t-s} \lambda_{is}^{1-\phi}$$

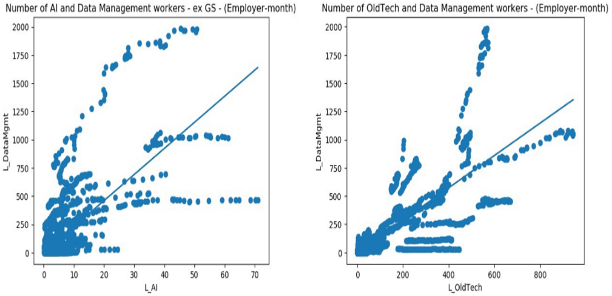
  We fix the depreciation rate of data at 2.5% per month.

- the average initial data stock is: $\frac{1}{N}\sum_i \iota A^{DM} \lambda_{2015,i}^{1-\phi} = \hat{D}_0$

$$D_{it} = (1-\delta)^t \iota A^{DM} \lambda_{2015,i}^{1-\phi} + \sum_{s=0}^{t} (1-\delta)^{t-s} A^{DM} \lambda_{is}^{1-\phi}$$

Results

- Do firms that hire more data management workers, and thus presumably have larger structured data sets, also hire more analysis workers?
- YES



Results

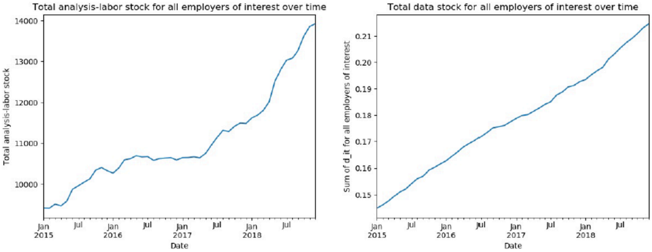
- What are the production function exponents from each technology?

		$\delta = 1\%$	$\delta = 2.5\%$	$\delta = 10\%$
Data Management	ϕ	0.172	0.190	0.144
		(0.0025)	(0.0019)	(0.0022)
AI Analysis	α	0.806	0.734	0.613
		(0.0013)	(0.0026)	(0.0038)
Old Technology Analysis	γ	0.458	0.560	0.567
		(0.0024)	(0.0017)	(0.0006)

- $\alpha > \gamma$ means that the rate of diminishing returns to data is less with the new AI technology.

Data Stocks and Labor Stocks

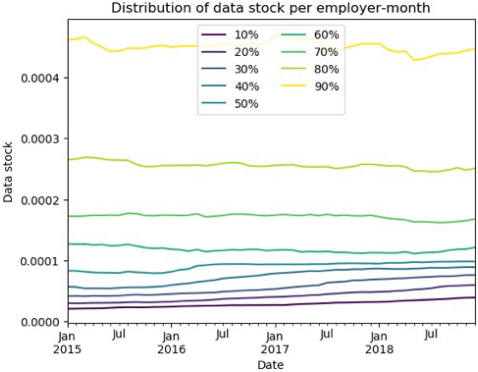
- One of the concerns people have with new data technologies is that they might be labor replacing.



- AI technology seems to be increasing not decreasing labor demand.
- The old technology analysts who are also made more productive by the abundance of structured data.

Data in the Cross-Section of Firms

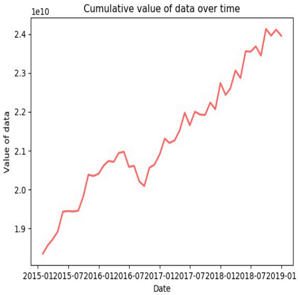
- The distribution of data is skewed. A few firms have enormous troves of data and many have very little.

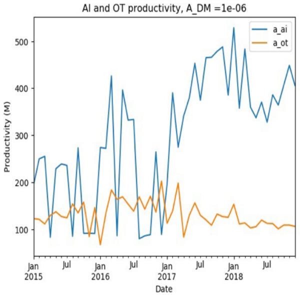


- The aggregate stock of data is growing rapidly, and the stock of data is quite stable.

Estimating the Value of Data

- One of the big questions in economics and finance today is how to value firms' data stocks.
- Where does this increase in value come from?
  - The accumulation of data
  - The increase in financial analysts that work with data
  - Firms are becoming more productive at using data.





Conclusion

- The key feature of industrialization is that factor shares changed. Thus if big data technologies are the industrialization of knowledge production, they should offer less diminishing returns to data.
  - We explored this hypothesis by modeling the production of knowledge
  - We described how labor and data can be mixed with a Cobb-Douglas production function to produce knowledge.
  - The firms with more data are more prone to hire more big-data or AI workers.

Consideration

- Monopoly & Steady state
- Job posting data predicts the future returns