Laura Veldkamp

The Missing Value of Firms' Data: Measurement in an Era of AI

On Thursday, February 20, Laura Veldkamp joined Markus' Academy for a conversation. Veldkamp is the Leon G. Cooperman Professor of Finance & Economics at Columbia University's Graduate School of Business.

A few highlights from the discussion.[1]

A summary in four bullets:

- Data is one of the most valuable assets in the modern economy, but it is also one of the hardest to measure
- Drawing from her research and recent book with Isaac Baley, *The Data Economy*, Veldkamp discussed the reasons and implications of these measurement problems, along with potential solutions
- The value of data is hard to measure because it is exchanged through barter and bundled within traditional transactions. When a customer buys a product with money they also provide data in the process. Firms get paid both with money and data
- The literature has followed three approaches to value data: (1) seeing data as a driver of firm productivity and valuing data through value functions, (2) seeing data as a complementary input, and (3) valuing data based on the present discounted value of the revenues it generates

[00:00] Markus' introduction

- Solow (1987) famously pointed that we "can see the computer age everywhere but in the productivity statistics"
- The OECD/IMF have estimated that counting free social media (funded by advertising) as household spending would raise the GDP growth rate by 0.07%. Correcting for the measurement of ICT equipment and software in the price index would raise the growth rate by 0.1% (Ahmad et al. 2017)
- However, accounting for the value of data assets does not have a large effect because these depreciate quickly, and their ownership rights are hard to enforce

[08:09] Data barter and bundling

- Why does economic measurement struggle to capture the value of data?
- Economists' tools are from an industrial era. In our models firms combine capital and labor at a location to produce widgets which only one person can consume at a time. Our frameworks need updating for the modern economy
- Data is the fuel for AI. AI requires big data sets and is a prediction tool.
- Big data is a byproduct of economic activity (search history, purchases, foot traffic…)

[1] Summary produced by Pablo Balsinde (PhD student, Stockholm School of Economics)

- More fuel for AI increases its prediction accuracy. Thus data is a tool to reduce prediction errors. In this sense it is different from technologies or patents (concepts) or learning-by-doing (human capital)
- The mismeasurement of data in GDP happens because a barter trade occurs when data, a valuable asset, is exchanged for a good without money changing hands
- Partial data barter is widespread: a customer buys a product with money but also provides data in the process (e.g., payment method, ZIP code)
- Firms receive compensation in money and data, while consumers pay only the net value of the good. Examples are supermarket loyalty cards or frequent flier programs
- In a perfectly competitive economy, all the gains from data collection should be passed back to consumers in the form of lower prices. However, in reality, firms often retain the surplus, in part because consumers have trouble seeing the value of data (to them it is like a foreign currency)
- We already have competition tools to prevent transaction bundling. Firms could be required to offer a regular price that allows for data collection and a privacy-protected price that mandates data deletion
- Clearer pricing structures would help consumers recognize data as an asset, leading to more informed choices about how and when to share it

[27:17] Data dynamics: feedback loops and depreciation

- More transactions (customers) leads to more data, which improves efficiency and profits. More profitable firms tend to grow faster, so they will be able to generate even more data. This creates an increasing returns data feedback loop
- If this is true, data is already counted in GDP measurements through the increased efficiency and profits. But this has two problems:
- (1) Firms' future output will also be priced below its actual value, since it is also paid for partly with data. As a result, we do not capture the full value of the additional future productivity
- (2) The value of data would be recorded at the point where it is used by the firm, rather than its inception. The timing will be misaligned, leading to an appearance of a productivity slowdown
- We can understand how data depreciates through a simple model (Veldkamp, 2023). Suppose we want to predict a variable that follows an AR(1) process. Define the stock of data as the precision (inverse variance) of our forecasts of that variable
- Through Bayesian updating, one can arrive at a law of motion for data (precision) similar to the ones we tend to write for capital: the stock of data in a period will depend on the prior period stock, depreciation, and new data inflows (investment)
- However, unlike capital, data depreciates faster when it is abundant and when the environment is volatile, as shocks make old information obsolete more quickly

[35:40] 1st approach to measure and value data: value functions

- If data generates payoffs over multiple periods, we can value it in the same way that macroeconomists value capital—with a value function (Bellman eq.) for the data stock (Farboodi and Veldkamp, 2022)
- Consider a firm with a Cobb-Douglas function over capital and labor. Data is not an input in production, but it drives the firm's productivity parameter (the A). It follows a law of motion with depreciation and new data inflows

- Feedback loops emerge if we add model ingredients like the number of transactions determining data inflows. Even with constant returns in capital and labor, firms will exhibit increasing returns to scale due to how new data is generated

[41:39] 2nd approach: data as a complementary input

- Consider a model where actionable insights (or knowledge) are produced using structured data and "analyst" labor. Structured data needs to be created by specialized "data management" labor, and depreciates over time
- We are ultimately decomposing the knowledge value chain into raw data, structured data and knowledge. Although this decomposition is helpful for economists, it means that the natural units to measure the amount of data (bits/bytes) are not helpful to us. Measuring the value and amount of data become intrinsically linked questions
- As researchers, we observe firms' choices of both types of labor and their wages. The key idea of the second approach is to infer how much data a firm must have for its observed labor decisions to be optimal
- In this line, Bresnahan et al. (2002) consider IT capital as another complementary input and try to infer firms' stock of data.
- Also with this approach, Abis and Veldkamp (2023) find that the value of structured data grew by 30% between 2015-2018. Value is growing for three reasons roughly equally: (1) firms accumulate more data, (2) more analysts make each data point more valuable, (3) firms are becoming more productive at using AI
- To many, the growth in data's value seems surprisingly low, given the exponential rise in data storage and cloud computing. However, data has decreasing marginal returns. Data is a tool for prediction, and standard errors shrink far less with the millionth data observation

[49:18] 3rd approach: revenues

- The value of data should be the present discounted value of the revenue it generates, but how can we isolate data revenues from other revenue?
- One can take an econometric approach: Kumar et al. (2023) implement an RCT treating firms with data
- If one knows how data generates revenue one can build a model. In a finance setting, one can quantify portfolio choice models to estimate investors' willingness to pay for data (Manela and Kada, 2021; Davila and Parlatore, 2024, Cong et al. 2021)
- Although many have assumed that the value of data is common across investors, Farboodi et al. (2025) finds that data has a large private value component depending on the investor's wealth, style, price impact or trading frequency. The value of the same data can vary from $10 to $1.2 million
- What if we don't know how firms extract value from data? We can measure the stock of data with forecast errors
- Consider a model where a firm's profits decline with (squared) forecast errors. Firms' precision depend on their prior beliefs, their own data, and the data they purchase
- We can observe how much forecast errors hurt profitability (from Asriyan and Kohlas, 2025), the value of the data firms directly purchase, and with some assumptions we can estimate their priors
- This allows us to back out firms' stock of data. In preliminary work, Ordonez and Veldkamp estimate that the average firm has a stock of data worth $43 million, equivalent to 14% of annual revenues

- They also find that, while very small firms make enormous mistakes, there has been a convergence in the errors made by the rest of the firms. That is, there might be a convergence of data stocks across firms

[1:01:10] Conclusions

- Future research should explore firm data heterogeneity, its impact on competition and market power, how AI-driven gains are distributed, and the tension between the private and social value of data: do privacy laws help or harm consumers?
- The EU is trying to restrict data sharing, while the US allows for free and open trade of data. The key is that some consumers strongly prefer privacy, and we should have structures in place to provide that value to them
- Many consumers worry about being price-discriminated against based on data. However the solution need not be to restrict data, but instead to promote competition so that the gains from data flow back to consumers

Timestamps: