

6 approaches to valuing data

By Laura Veldkamp



Data drives our modern economy yet remains one of the hardest assets to observe, measure and put a price on. Can the financial tools we use to measure physical capital be applied to data? Let's do the math and find out.





Think about some of the most valuable firms in the world today – Amazon, Microsoft and Google. What are they most valued for? Their buildings? Their workers? Their technologies? Primarily, it's their data. And that raises lots of questions that touch on every aspect of finance, from asset pricing (If data is a new kind of asset, how do we value it?) to corporate financing (Do data-intensive firms have valuations that are realistic?) to entrepreneurship (If data serves as an entry barrier for new firms, what does market power look like if digital services are being given away for free?).



The data economy is not the same as the old economy. We need to update our tools

Answering these questions isn't easy, given our reliance on industrial-era economics tools based on using labor and capital for production. The data economy is not the same as the old economy. We need to update our tools. And while some existing financial tools can be used to value data, we must acknowledge there are more open questions than resolved ones, given that data is one of the hardest assets to observe, measure and put a price on. Yet, given its central importance to our modern economy, we must at least try.

In this article, I will suggest six approaches to valuing data and how we might make progress, even with imperfect measurement tools.

The economic nature of data

Data information has been around for centuries, but what's different now is that data has become digitized. Yet, it's not just the digitization of data that's interesting; it's the rise of big data, such as that generated by AI algorithms based on your search history, your car traffic patterns, your purchases and so on, which exists as a by-product and as a predictor of economic activity, rather than being a mere input for invention or research. This is a key difference. I've never engaged in economic activity and had a research paper appear as a byproduct. Conversely, data *does* materialize as a consequence of my activities. Moreover, that data asset can be bought and sold in ways that I can't likewise trade the knowledge assets in my own head.

Another key difference between big data and other forms of intellectual property is that it may arise from a barter trade, i.e., you trade your personal information to obtain something for "free" but then the company is able to monetize that information. This is the business model of countless apps. The more transactions there are, the more data is generated and the more money is made, or the higher quality, efficiency or productivity a firm is able to achieve, thanks to your data.

How many red versus purple sweaters will be demanded? How many should I load onto the truck or container ship? How many do I need to keep in stock? Whom should I hire? Which firms should I invest in? These economic decisions are being informed by data. And the more transactions I have and the more efficient or productive I become, the more I can grow and the more profits I can make, in a self-reinforcing feedback loop of increasing returns.



This data economy can be represented in a simple model (see figure). Taking our initial asset (in this case, data), we can presume some depreciation of that data over time, and we can calculate a depreciation value. To offset that depreciation, we must generate more data, so we add some new inflow – more transactions and/or customers (or new capital or labor) and calculate a new asset value. We can also use efficiency ratios to measure our ability to generate income with our data assets in a certain period of time. We can enrich these formulas any way we want, but with simple equations similar to those used to measure a capital stock, this spiral of increasing returns can be calculated mathematically.

There's a second piece to this, which I alluded to before: how data creates value and raises not just current profits (e.g., through a data-led decision to produce red sweaters instead of purple ones) but future profits, too. Firms with more data get more customers; more customers, more data – until they become the superstar firm that nobody can compete with. As such, data acts as an entry barrier, creating value by keeping competitors out.

A third way that I want us to think about data creating value is that it reduces risk. Remember that data is information. And having information essentially helps you predict something better or more accurately, which resolves uncertainty and randomness, and thereby reduces risk. This is where we sometimes miss the boat. CFOs may think about data in terms of the first piece:

how to raise current profits. Other managers and strategists may focus on data as market power. But few think about how data reduces risk and how to put a price on that.

With this in mind, let's now look at six approaches to measuring data, derived from various academic work I have done with Simona Abis (Columbia), Juliane Begenau (Stanford), Cindy Chung (Stanford), Jan Eeckhout (Pompeu Fabra), Maryam Farboodi (MIT Sloan), Adrien Matray (Princeton), Roxana Mihet (HEC Lausanne), Thomas Philippon (NYU Stern), Dhruv Singal (Columbia) and Venky Venkateswaran (NYU Stern).

1. Cost accounting

The first approach is similar to book value. When we value a physical asset, like an office, we take account of all the costly investments that went into building it and maintaining it, and factor in some depreciation or discount rate. Could we do the same with data?

The challenge goes back to what I said before: most data is a byproduct of economic activity, and no one explicitly paid anything for it, or if they did, it was probably as a barter trade.

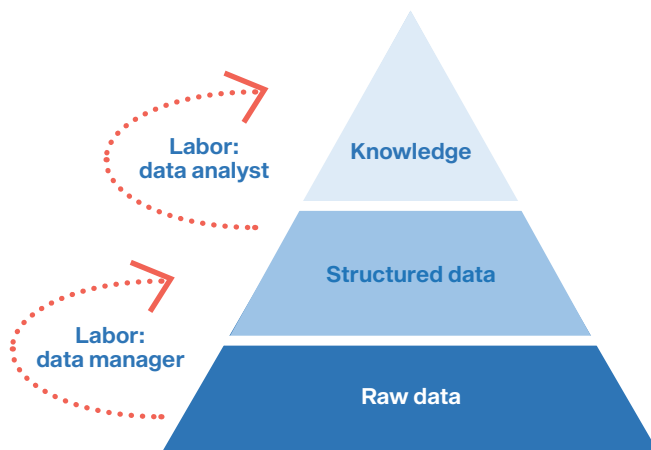
Even so, we should be able to put a price on it. Consider this everyday example: a store offers you a 10% discount if you sign up to their membership scheme, which involves you giving them your data. The store lowering its price in order to get your data acts as a kind of payment for your

Knowing the quantity of labor and wages, we can infer how much data these firms must have and how valuable it may be to them

data, in which case we could say that the cost of that trade was the 10% amount of your purchase.

The same goes for startups. New firms starting out tend to lower their prices to attract more transactions and more customers, in order to grow their database. That data is extremely valuable to them – indeed, their entire business model may rest on it – in which case, the amount they lowered their prices by could be counted as payment for your data. If we can impute the discount, then a cost accounting approach might work to calculate the value of data in such contexts.

Another example would be if we just bought the datasets from someone; then, of course, we could simply add up those costs. However, most cases aren't so nice and clean, where we know exactly how much we are being compensated with data.



2. Complementary inputs

Consider this pyramid. Firms collect lots of raw data, but they need to structure it for some meaningful use. For that, firms need data managers. They also need analysts to transform that structured data into actionable knowledge. The number of data managers and analysts that a firm employs can be found in its hiring records.

Using a mathematical model of how firms combine labor and data to create knowledge, we can relate the number of data managers and analysts a firm hired, as well as the wages of those workers, to an amount of data. The idea is that we answer the question: How much data must this firm have for hiring that many workers of each type, at these salaries, to make sense as an optimal hiring strategy?

Thus, even though no firm is going to reveal what data they have on their servers, we can observe firms' hiring and wages. Knowing the quantity of labor and wages, we can infer how much data these firms must have and how valuable it may be to them.

Another observable complementary input would be a firm's IT. Knowledge of how many computers a firm has, and how much was paid for each one, can reveal the value that the firm's data must have to justify this IT investment.

3. Value functions

Another approach to valuing data is to treat data like physical capital. When macroeconomists value physical capital, they use a *recursive* approach. That is to say, they ask: How much value will a firm get from its capital today, and then what will the remaining value of the capital be tomorrow? This procedure of calculating one-period-ahead benefits, and then using them to build up the entire value of a long-lived asset, can be applied to data.

Data is similar in that it doesn't just have a value today; it's a long-lived asset that we evaluate both for what it's worth to us today as well as in the future. In order to determine the future value, we need to model how data depreciates over time. We determine how the depreciation rate of data depends on the volatility of the market environment, as well as the amount of data a firm has.

In research with colleagues, we combined the labor complementary input approach described earlier, with the recursive valuation, to estimate not just the quantity but also the value of financial data for firms in the financial analysis sector.

We found that the value of the data in the group of firms we studied grew 33% over three years (2015-18). We identified three reasons for such an enormous rate of growth:

- **The accumulation of more data.** This accumulation effect goes back to what I mentioned at the beginning of this article, where more begets more.
- **More analysis workers.** More labor working with a given unit of data means the marginal product of data goes up and its return increases as well.
- **More productivity using AI.** Looking at the job postings for these firms, we noted they required skills that would imply AI and machine learning. These skills boost the ability to create new knowledge, further contributing to the growth and value of data.

4. Revenue

Can we value data, like any other asset, as the present discounted value of the revenue it generates? The complication here is how to isolate data revenue from other sources of revenue. And even among firms for whom data is their primary asset, a lot of data-intensive firms operate at a loss. So, if you look at their revenue today and try to project forward, you're going to get a negative.

That said, using a revenue approach is doable, but it requires some counterfactual thinking. Because when we say, "What's the value of the data?" we're basically saying, "What's the value of the firm *with* the data versus *without*?"

This kind of exercise has been done specifically for data that is used to manage a portfolio of risky assets. The problem with using financial assets is, if you put a price of \$203 on a share of General Motors, then I should put approximately the same price on it as well, and data is not like that; data has a big *private value* piece.

Using a revenue approach is doable, but the complication is how to isolate data revenue from other sources of revenue

To probe this, colleagues and I worked out the value of data as expected utility *with* the data and expected utility *without*, for a firm that uses the data to craft an optimal portfolio of risk assets to invest in. We measured the value of data as being related to one's absolute risk aversion, multiplied by a term that had to do with three moments: *expected returns*, the *variance of returns* and the *conditional variance of returns* (i.e., the degree of uncertainty about our return forecast, with *high* conditional variance meaning that we believed our forecast was likely to be far from the truth).

As noted above, we found that the same piece of data had vastly different values for different traders, ranging from \$10 to as much as \$1.2 million, depending on the investor's wealth, investment style, price impact or trading frequency.

The takeaway here is that we must be very careful whenever trying to calculate what data is worth. The answer could be anything from \$10 to \$1.2 million, and both are correct, because ultimately the value of data to *you* depends on how *you* want to use it.

5. Choice covariance

At its heart, data allows agents to make better choices, which is why firms value it: They can get better signals and achieve a better match with something they are trying to forecast. In statistics, we use covariance to determine the relationship or correlation between the movements of two random variables. A better choice, then, means agents take an action that covaries with some payoff according to their expectation or objective.

An example could be: What color sweaters are we going to produce? And what's the profit or markup for each of these kinds of sweaters? The firm would like their production to covary with the profitability of each sweater type. In other words,

Production-markup covariance is a useful measure of the amount of data a firm has

they want to produce more of the sweaters that will yield high profits and less of the sweaters that turn out to be less profitable.

It's important to note that agents cannot achieve a high covariance without knowing something about what the profit will be – and that requires data. Without data, you cannot systematically choose an action that covaries with something you know nothing about – it's not a feasible strategy.

Regarding the product markup example, a firm would need to have some data about what the future markup will be, in order to systematically choose to produce and sell more products that will subsequently have higher markups. Colleagues and I theorized that high-data firms skew their product mix toward high-markup products, and our research fit the empirical markup facts well. In such an environment, the production-markup covariance is a useful measure of the amount of data a firm has.

6. Intangible asset

You might say data is an intangible asset, so why not use a typical intangible asset approach, like book-to-market value, or rather market-to-book value?

First, you have to make sure data is not in a firm's book value. Data might indeed show up on your book if you purchased it, if you acquired a firm that was valued largely for its data, or if you generated data as a byproduct of economic activity via a barter trade, as discussed earlier.

Other research has delved into measuring the market value of intangible assets such as branding, patents and organizational capital. But data is not the same as any of these. Moreover, data might contribute to some intangible assets, such as a good dataset helping us organize our firm better for more valuable organizational capital. Yet, how do we tease apart the book value of the data we acquired from the part that data played in the intangible asset which might be captured in the firm's market value?

Even if we could split out the value of data from the value of these other types of intangible assets, I remain skeptical about using market equity values to give us the answer to how much data is worth, because it presumes that equity market participants know exactly how to value that data. And as I think we can all agree, we haven't totally figured that out yet.

Many questions to tackle

Even though these six approaches each have their flaws, we must not give up. We can't pretend that data doesn't exist or that it isn't producing or worth anything of value. We're going to have to update our industrial-era tools, and try to make progress with these new methods, as imperfect as they may be.

In this effort, theory and measurement are going to have to work together. To understand data value, we must learn more about data supply: from data markets, platforms and data ownership. And we should intersect supply and demand, and try to understand what the equilibrium prices of

The author



Laura Veldkamp is a professor of finance at Columbia University's Graduate School of Business and a former editor of the *Journal of Economic Theory*.



data should be. Then, we should see if these prices fluctuate. Do data prices reflect risk premia? What do they covary with? What are the characteristics of this asset and its market price? What does the competition look like? How easy is it to enter the data market?

The number of questions to tackle is enormous. Here's one more: Are you prepared to help your colleagues in the field work out the answers? ■

SOURCE: This article is based on Laura Veldkamp's keynote talk on "Valuing Data and Data-Intensive Firms" delivered during the 49th Annual Meeting of the European Finance Association, which was organized by IESE on its Barcelona campus in August 2022. <https://efa2022.efa-meetings.org/>

Laura Veldkamp also co-authored, with Xavier Vives (IESE), Darrell Duffie (Stanford) and Thierry Foucault (HEC Paris), the *Technology and Finance* report (2022), the fourth in the Future of Banking series, which is part of IESE's Banking Initiative. <https://media.iese.edu/research/pdfs/75956.pdf>