Valuing Financial Data

Maryam Farboodi
MIT

Dhruv Singal
Columbia

Laura Veldkamp
Columbia

Venky Venkateswaran
NYU

2023

Value of Financial Data to Investors

- Data is a valuable asset for investors. How valuable?

  What is an investor's willingness to pay? This is a demand, not an equilibrium transactions price.

- Data valuation is not easy
  - How much you can profit from data depends on who else knows that data, who knows similar data, and how aggressively they will trade on it.

    $\Rightarrow$ impossible data requirements: Everyone else's data sets, preferences, price impact, investment mandates ...

- Our contribution: sufficient statistics that bypass the need to know others' information sets and characteristics.

$\Rightarrow$ A tool to put a dollar value on a piece of data.
It depends only on returns and *your* characteristics

Findings

- Investor characteristics change the value of data by orders of magnitude
  Data is not a common value asset: same data valued very differently by different investors.

- Demand elasticity of data tied to asset market elasticity.

- Related work puts one value on each piece of data
  Manela & Kadan ('20, '21), Glode, Green, Lowrey ('12), Kacperczyk & Sundaresan ('19), Davila & Parlatore ('20), Farboodi, Matray, Veldkamp, Venkateswaran ('21)

Measuring different data values for different traders is new.
It is necessary to trace out data a demand curve.

Outline

A Measurement Tool

Estimation

Results

A Kitchen-Sink Model for Measurement

A noisy rational expectations model with:

- Utility with wealth effects
- Long-lived asset
- Correlated asset payoffs
- Different trading styles (mandates)
- Data correlated across assets
- Price impact
- Public and private data

We do not need these assumptions.

The point: Our sufficient static survives all these complications.

Equilibrium Solution

- **Equilibrium**
  - Investors learn from prices and data. Update beliefs with Bayes' law.
  - Choose portfolios q_{it} to max EU, accounting for price impact dp/dq_i.
  - Price p_t equates demand and supply.
- A second-order approximation to expected utility.
- Profits $(\Pi_t) \rightarrow$ excess returns (R_t) for measurement
- Substitute optimal portfolio q_{it}, equilibrium price, price information and take expectations over realization of random outcomes and signals

$$\tilde{U}(\mathcal{I}_{it}) \approx \underbrace{\mathbb{E}\left[R_t\right]' \hat{\hat{V}}_{it}^{-1} \mathbb{E}\left[R_t\right]}_{(\text{Sharpe ratio})^2} + \text{Tr}\left[\underbrace{\left(\mathbb{V}[R_t] - \mathbb{V}[R_t \mid \mathcal{I}_{it}]\right)}_{\text{variance reduction}} \hat{\hat{V}}_{it}^{-1}\right] + r\rho_i \bar{w}_{it}$$

  - R_t: returns for i's investable assets based on his investment style
  - $\hat{\hat{V}}_{it}$: conditional variance of this return, adj for price impact $\frac{dp}{dq_i}$

The Insight: Others' Info Disappeared!

- **Dollar value of data:** investor indifferent between
  having the data $\equiv$ no data+ additional riskless wealth

$$\$\text{Value of Data}_i = \frac{1}{r\rho_i}\left(\tilde{U}(\mathcal{I}_{it} + \text{data}) - \tilde{U}(\mathcal{I}_{it})\right)$$

  We can estimate this with your info and public info!

  Whether data is public, private or correlated with what others know is crucial, but it matters through conditional variances

- Utility looks like in many REE models. What's new?
  - Mapping many models into these sufficient stats is new!
    Models with: heterogeneous investors, style constraints, investment in many assets, data that is private, partially public or correlated with what others know ...
  - Return-based sufficient stats are a crucial recent step forward for NREE.

Outline

A Measurement Tool

Estimation

Results

Estimating Conditional Variances

- Small insight: For linear normals, Bayes law and OLS coincide. $\mathbb{V}[R_{t+1} \mid \mathcal{I}_{it}]$ is the expected squared residual from OLS regression.
- Data to be valued X_t, and existing data Z_t

$$R_{t+1} = \gamma_2 Z_t + \varepsilon_t^Z$$
$$R_{t+1} = \beta_1 X_t + \beta_2 Z_t + \varepsilon_t^{XZ}$$

- Conditional variance without data we're valuing

$$\mathbb{V}[R_{t+1} \mid \mathcal{I}_{it}] \approx \widehat{\mathrm{Cov}}[\varepsilon_t^Z] = \frac{1}{T - |Z|} \sum_{t=1}^{T} \varepsilon_t^Z \varepsilon_t^{Z\prime}$$

- Conditional variance with data

$$\mathbb{V}[R_{t+1} \mid \mathcal{I}_{it} + \mathrm{data}] \approx \widehat{\mathrm{Cov}}[\varepsilon_t^{XZ}] = \frac{1}{T - |Z| - |X|} \sum_{t=1}^{T} \varepsilon_t^{XZ} \varepsilon_t^{XZ\prime}$$

- Plug these in equilibrium expected utility to get data value.

Estimation: Data Sources

- A proof of concept: Value the same data for many different investors.
- The data we value: Institutional Brokers Estimate System (I/B/E/S) earnings forecasts, 1985-2015
- Problem: There are too many covariances to estimate.
  1. Group assets into portfolios,
  2. Use value-weighted means.
- **Key:** methodology can be easily adapted.

Outline

A Measurement Tool

Estimation

Results

Different Values for the Same Data

How much are this year's IBES forecasts worth, to an investor who only knows D_t / P_t (learning from prices)? A take-it-or-leave-it offer.

	Investment Style				
	Small	Large	Growth	Value	All
Perfect Competition					
Investor with $500,000 Wealth	0.00	$1.7k	$2.5k	$490	$3.5k
Investor with $250m Wealth	0.00	$566k	$844k	$164k	$1.2m
With Price Impact					
Investor with $500,000 Wealth	0.00	$1.6k	$2.5k	$410	$1.4k
Investor with $250m Wealth	0.00	$24k	$57k	$1.5k	$253k

Purple: Richer investors value data much more than poorer ones.

Yellow: Investment style matters enormously.

Red: Price impact reduces the value of data - a little or a lot.

The dispersion of valuations for the same data is immense!

Effects of Investor Heterogeneity

	Effect on Data Value	
More wealth	$\uparrow$	
Price impact	$\downarrow$	
Investment style	it matters	
Previously purchased data	$\downarrow$ modestly	in the paper
Trading horizon	modest effect that varies	in the paper

Directional effects intuitive but effects often compete.

Magnitudes would be tough to guess without our tool.

Conclusion

- Data is one of the most valuable assets in the modern economy.
- Data has enormously variable private values.

  The same data is worth vastly different amounts to investors with different wealth and style, with and without price impacts, ...
- Next steps to understand data markets:
  - Estimate distributions of investor characteristics to produce a demand curve.
  - Understand the data supply side.

Then we can do asset pricing theory ... for data!