Valuing Financial Data

by Farboodi, Singal, Veldkamp, and Venkateswaran

Discussant: Asaf Manela

Reichman University and Washington University in St. Louis

September 2022

Question

- What is the value of financial data to an investor?
- Data = information
- Large theoretical literature on value of information
  - Grossman-Stiglitz (1980 AER)
  - ...
  - Farboodi-Veldkamp (2020 AER)
- Few empirical estimates
  - Manela (2014 JFE)
  - Kadan-Manela (2019 RFS)
  - Farboodi-Matray-Veldkamp-Venkateswaran (2021 RFS)
  - Kadan-Manela (2020 WP)

The value of information/data

General definition from Kadan-Manela (2019 RFS)

- An information structure α generates a stream of signals s_t, s_{t+1}, ...
- $V(a_t, z_t, s_t) =$ value function of an informed investor with wealth a_t in state z_t
- $V(a_t, z_t; \alpha_0) =$ value attained without observing the signals

The value of information/data

General definition from Kadan-Manela (2019 RFS)

- An information structure α generates a stream of signals s_t, s_{t+1}, ...
- $V(a_t, z_t, s_t)$ = value function of an informed investor with wealth a_t in state z_t
- $V(a_t, z_t; \alpha_0)$ = value attained without observing the signals

Definition

The value of information structure α in state z_t is the fraction of wealth Ω the agent would give up to observe a stream of signals s_t, s_{t+1}, ..., each generated by α

$$\mu\left[V\left(a_t\left(1-\Omega\right), z_t, s_t; \alpha\right) | z_t; \alpha\right] = V\left(a_t, z_t; \alpha_0\right)$$

where $\mu[\cdot]$ is the certainty equivalent over the signal s_t

What Farboodi-Singal-Veldkamp-Venkateswaran (2022 WP) does

Sufficient statistic approach that uses equilibrium asset return moments to summarize the value of data

- Elaborate noisy rational expectations model with many realistic features:
  - n heterogeneous risk averse informed investors
  - N assets
  - H data sources
- Magically, as number of informed investors grows large

$$\tilde{U}\left(\mathcal{I}_{it}\right) \approx \frac{1}{2}\mathbb{E}\left[R_{it}\right]' \mathbb{E}\left[\mathbb{V}\left[R_{it}|\mathcal{I}_{it}\right]^{-1}\right] \mathbb{E}\left[R_{it}\right] + \frac{1}{2}\mathsf{Tr}\left[\mathbb{V}\left[R_{it}\right]\mathbb{V}\left[R_{it}|\mathcal{I}_{it}\right]^{-1} - I\right] + r\bar{w}_{it}\rho_i$$

- Dollar value of data

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

- Beauty is that most elements can be easily estimated with data!

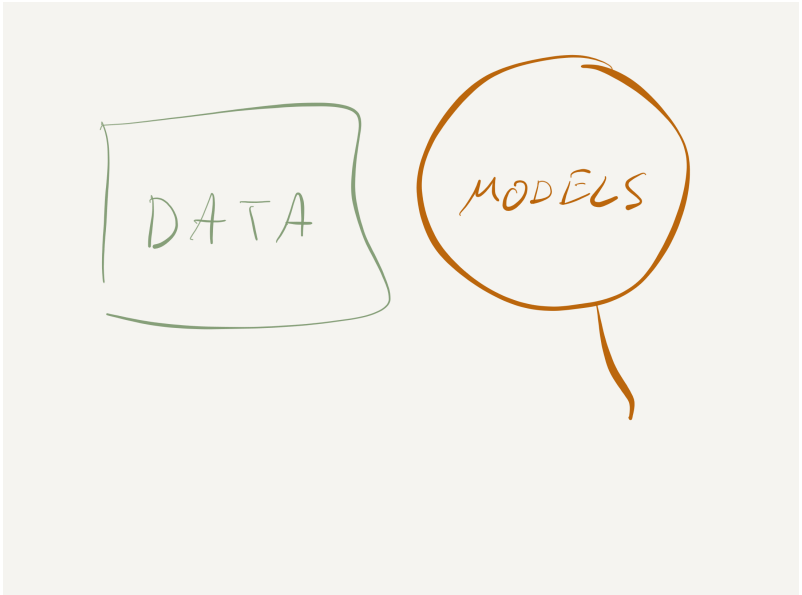
Contribution

	Preference	Fundamental	Asset	Period	Informed	Signal	Estimation	Price impact
Manela (2014 JFE)	CARA	Normal	1	3	Mass	1	Stocks around drug approvals	0
Kadan-Manela (2019 RFS)	Recursive	Any	N	∞	1	1	Index options around macro news	0
Farboodi-Matray-Veldkamp-Venkateswaran (2021 RFS)	CARA	Normal	N	1	n	1	Decade-freq for stock portfolios (Initial/marginal value)	0
Kadan-Manela (2020 WP)	Risk-neutral	Log normal	1	∞	n	n	Daily-freq for individual stocks	Estimated per day-stock
Farboodi-Singal-Veldkamp-Venkateswaran (2022 WP)	Mean-var.	Normal	N	∞	n	H	Decade-freq for stock portfolios	Constant from Hasbrouck

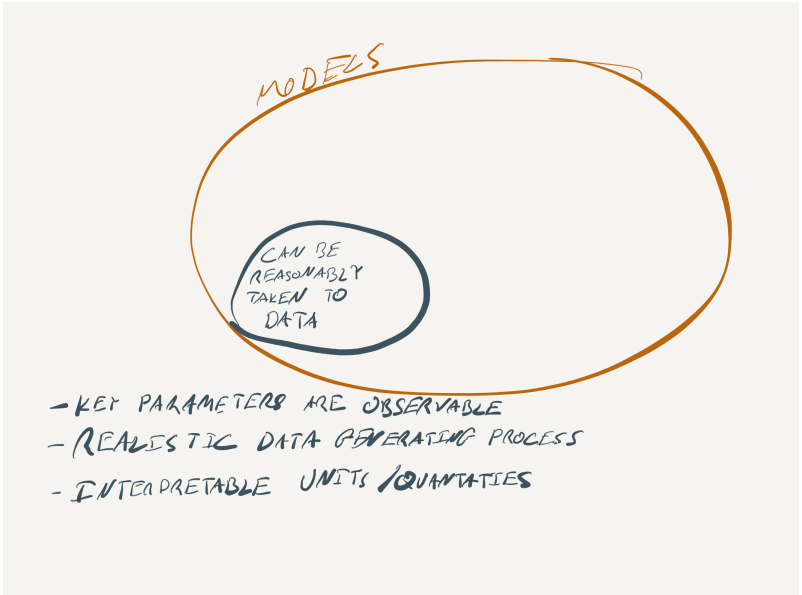
Suggestion: Embrace strong assumptions

- Examples:
  - Prices and dividends are not stationary or Gaussian
  - Risk aversion is correlated with wealth
  - Investors are not that myopic
  - Price impact is not constant over time or across assets and is correlated with volatility
- But sometimes empirical progress requires squinting
- Future work should try to relax these assumptions

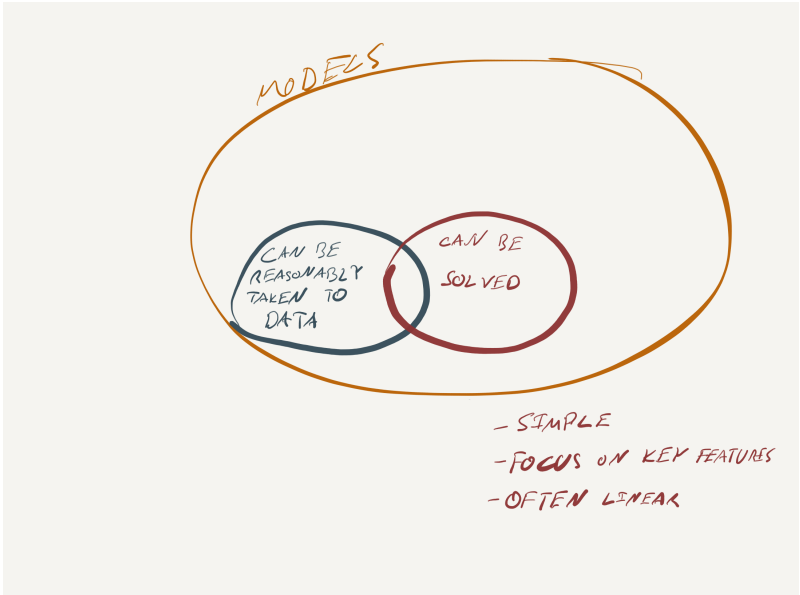
Big picture



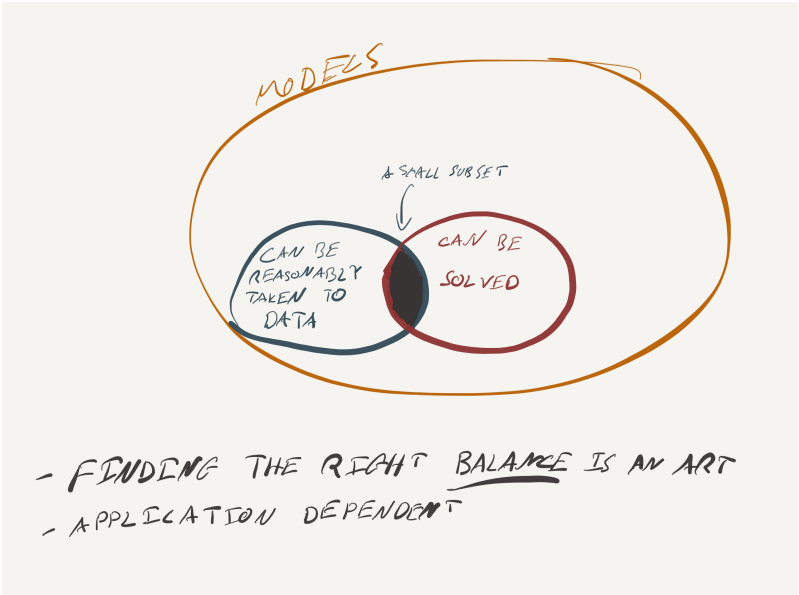
Big picture


MODELS
CAN BE REASONABLY TAKEN TO DATA
- KEY PARAMETERS ARE OBSERVABLE
- REALISTIC DATA GENERATING PROCESS
- INTERPRETABLE UNITS / QUANTITIES

Big picture


MODELS
CAN BE REASONABLY TAKEN TO DATA
CAN BE SOLVED
- SIMPLE
- FOCUS ON KEY FEATURES
- OFTEN LINEAR

Big picture


MODELS
A SMALL SUBSET
CAN BE REASONABLY TAKEN TO DATA
CAN BE SOLVED
- FINDING THE RIGHT BALANCE IS AN ART
- APPLICATION DEPENDENT

An alternative setup (Kadan-Manela, 2020 WP)

- In strategic trading models à la Kyle (1985):

$$\text{Value of information} = \frac{\text{Reduction in variance}}{\text{Price impact}}$$

- This simple closed-form expression for the value of information
  - Can be extended to the log-normal values case of Back (1992)
  - Can be estimated using TAQ at high frequencies
  - Provides a lower bound for cases with many investors and signals
- A reasonable approximation for idiosyncratic information

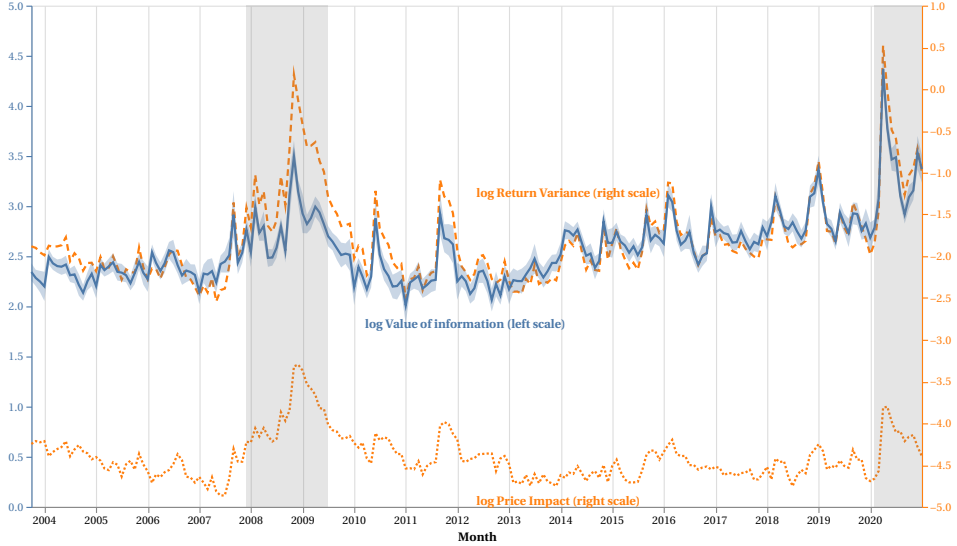
Estimation (Kadan-Manela, 2020 WP)

- For each common stock in the TAQ dataset (2003...) we calculate

$$\$\text{Value of information} = \frac{\sigma_v^2}{\lambda} \times P_0$$

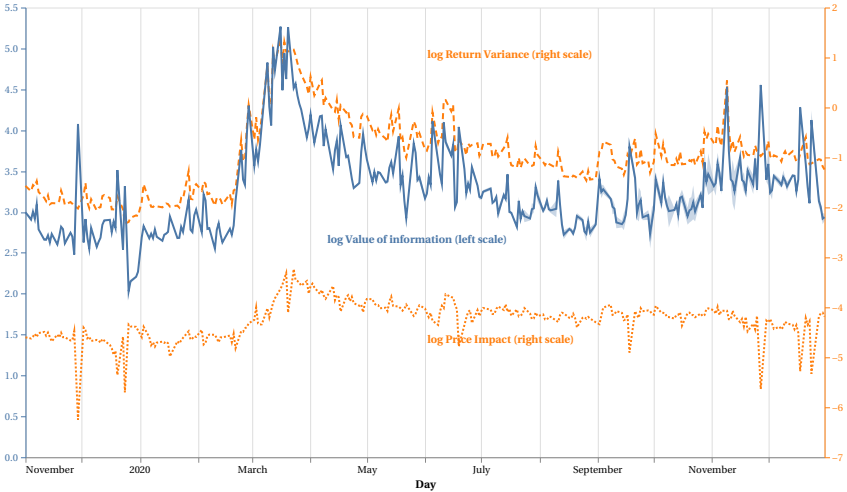
- $\hat{\sigma}_v^2$ is daily log return variance based on 1 min intraday + overnight returns
- $\hat{\lambda}$ estimated daily by regressing 1 min intraday log returns on signed order flow
- $\hat{P}_0$ is previous day closing price

Value of information rises in turbulent times



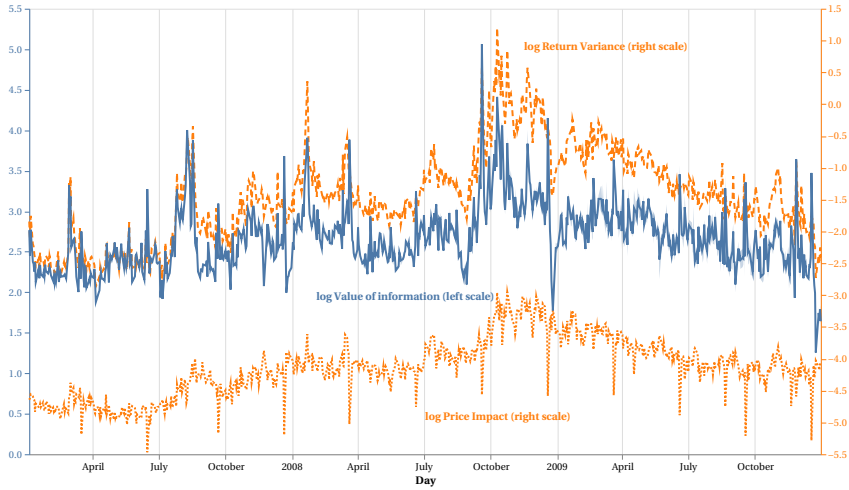
Start of the Covid-19 pandemic

Peaks at $20M on Mar 20 (highest in our sample).
9 of 10 highest daily values occur in Mar 2020.



Financial crisis of 2007–2009

Value peaks on Sep 19, 2008, then approaches zero as volatility declines and liquidity worsens



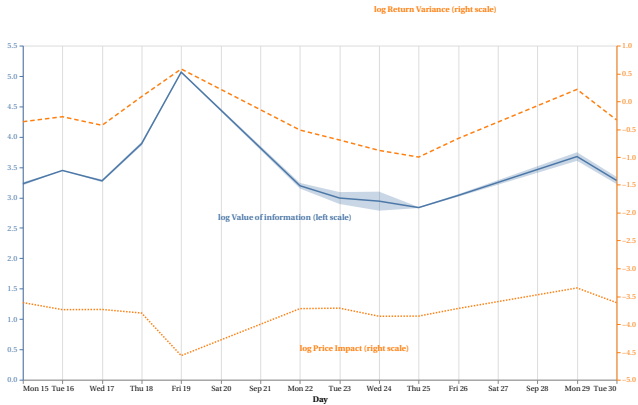
Financial crisis of 2007–2009 - zoom in on Sep 19

Value peaks on Sep 19, 2008, then approaches zero as volatility declines and liquidity worsens

9/15 Lehman fails; BofA acquires Merrill

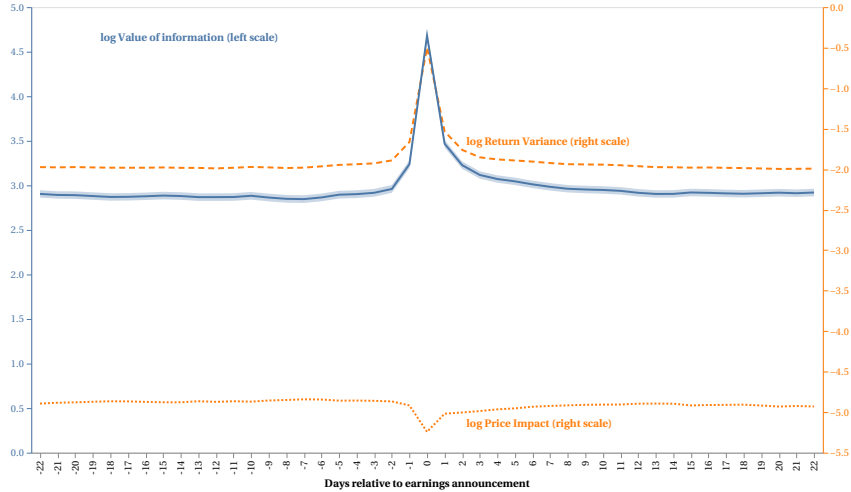
9/16 AIG bailout; Reserve Funds; FOMC maintains 2% rate

9/19 Fed suspend bank capital regulation; creates AMLF



Value of information rises when firms report earnings

Explains the previously documented fact that investor attention spikes around earnings



What if more than one investor has private information?

Kadan-Manela (2020 WP) building on Back-Cao-Willard (2000 JF)

- n informed investors
- Each gets a signal about the fundamental
- $\rho =$ Signal correlation across investors
- The value of information to each informed trader becomes

$$\$\text{Value of information} = \frac{c\left(n,\rho\right)}{n} \times \frac{\sigma_v^2}{\lambda_0}$$

- We show that for any $n \geq 1$ and $\rho \geq 0$, the value of information to each informed trader becomes

$$\$\text{Value of information} \geq \frac{0.92}{n} \times \frac{\sigma_v^2}{\lambda_0}$$

Concluding thoughts

- Data is increasingly available and useful
- Estimating investor demand for information is important
- Lots of work remains to find models:
  1. Can be solved
  2. Can be reasonably taken to data