Market efficiency in the age of machine learning

Talis Putnins

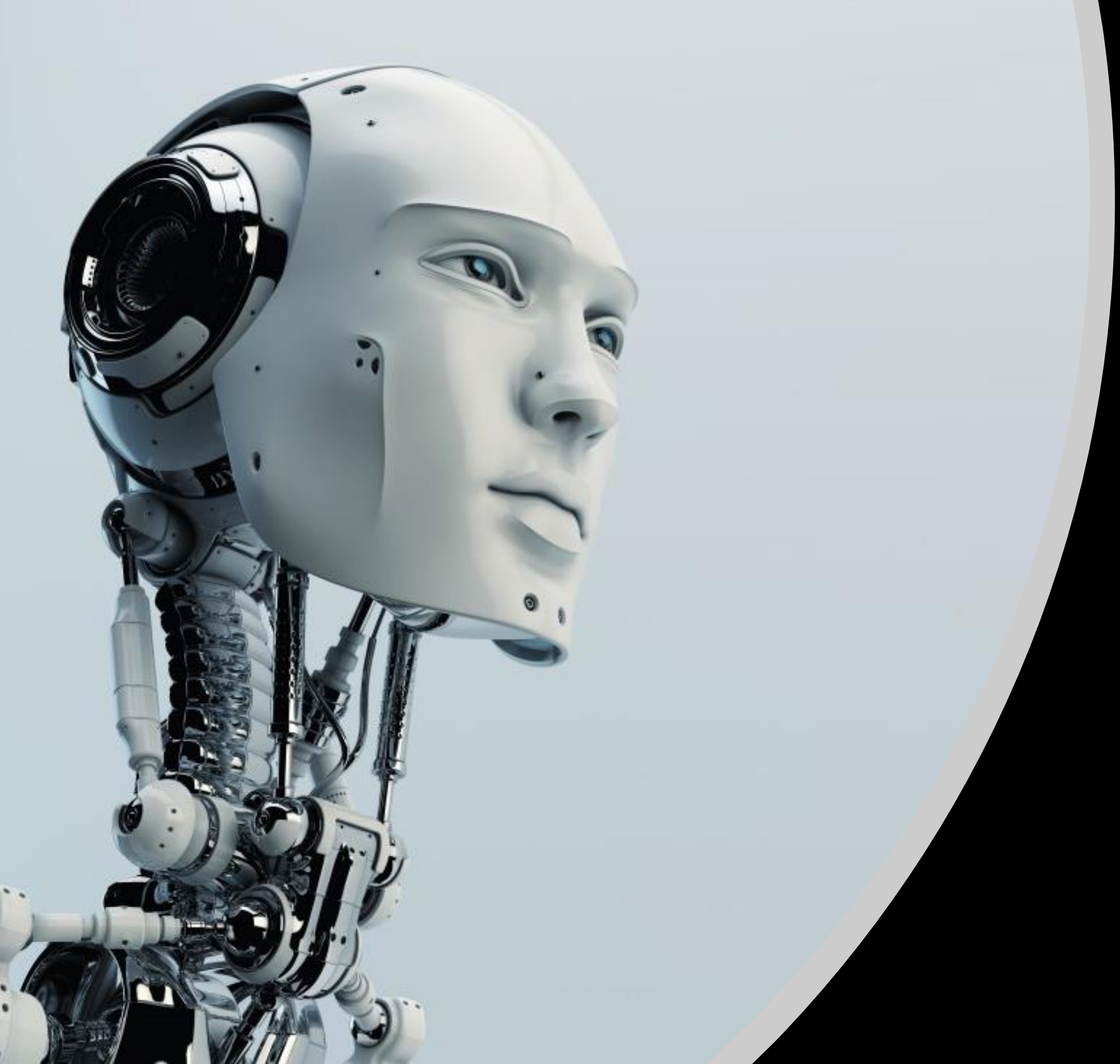
University of Technology Sydney and University of Edinburgh

Leonidas Barbopoulos, Rui Dai, Anthony Saunders

Disclaimer

This presentation was produced solely by Talis Putnins . The opinions and statements expressed herein are those of Talis Putnins are not necessarily the opinions of any other entity, including UBS AG and its affiliates. UBS AG and its affiliates accept no responsibility whatsoever for the accuracy, reliability or completeness of the information, statements or opinions contained in this presentation and will not be liable either directly or indirectly for any consequences, including any loss or damage, arising out of the use of or reliance on this presentation or any part thereof.

Reproduced with permission.



The question:

As machines replace humans, how is market (informational) efficiency impacted?

It’s not obvious …

- Market efficiency could **improve**:
  - ➔ Machines can process *more data* than a human
  - ➔ Machines can interpret data *faster* than a human
  - ➔ Machines are less susceptible to *emotions and bias*

… but …

- Market efficiency could **deteriorate**:
  - ➔ Machines cannot deal with *“soft” information* (can’t be quantified)
  - ➔ *Overfitting* ML models can result in trading on *spurious* correlations
  - ➔ Machines learn to predict the actions of humans – *front-run* them, reduce human profits and info gathering, without bringing new info
  - ➔ Trade-off between speed and accuracy in decision-making, squeezing out slow, deliberating humans could come at a *cost of accuracy*

Measuring the machines

Unique data from US SEC’s EDGAR servers

- Focus on 8-K filings of US stocks:
  - **Report of *unscheduled*, *material* events or corporate changes** at a company that are deemed important to shareholders or the SEC
    - Changes to a material agreements and contracts,
    - Certain financial information,
    - Mergers / Acquisitions / Disposals,
    - Substantial impairments / loan defaults,
    - Change in directors/officers
    - Change in control
    - Results of shareholder votes
    - Other material information, including Reg FD disclosures and press releases
  - Legally required ➔ they provide a complete record of certain unscheduled info types
  - Filed with the SEC and made public via the SEC’s EDGAR server

Example: Tesla 2 July 2021

Inline Viewer

UNITED STATES
SECURITIES AND EXCHANGE COMMISSION
WASHINGTON, D.C. 20549

FORM 8-K

CURRENT REPORT

Pursuant to Section 13 or 15(d) of the Securities Exchange Act of 1934

Date of Report (Date of earliest event reported): July 02, 2021

Tesla, Inc.

(Exact name of Registrant as Specified in Its Charter)

Delaware	**001-34756**	**91-2197729**
(State or Other Jurisdiction of Incorporation)	(Commission File Number)	(IRS Employer Identification No.)

3500 Deer Creek Road Palo Alto, California	**94304**
(Address of Principal Executive Offices)	(Zip Code)

EX-99.1 2 tsla-20210702ex99_1.htm EX-99.1

Exhibit 99.1

Tesla Q2 2021 Vehicle Production & Deliveries

In the second quarter, we produced and delivered over 200,000 vehicles. Our teams have done an outstanding job navigating through global supply chain and logistics challenges.

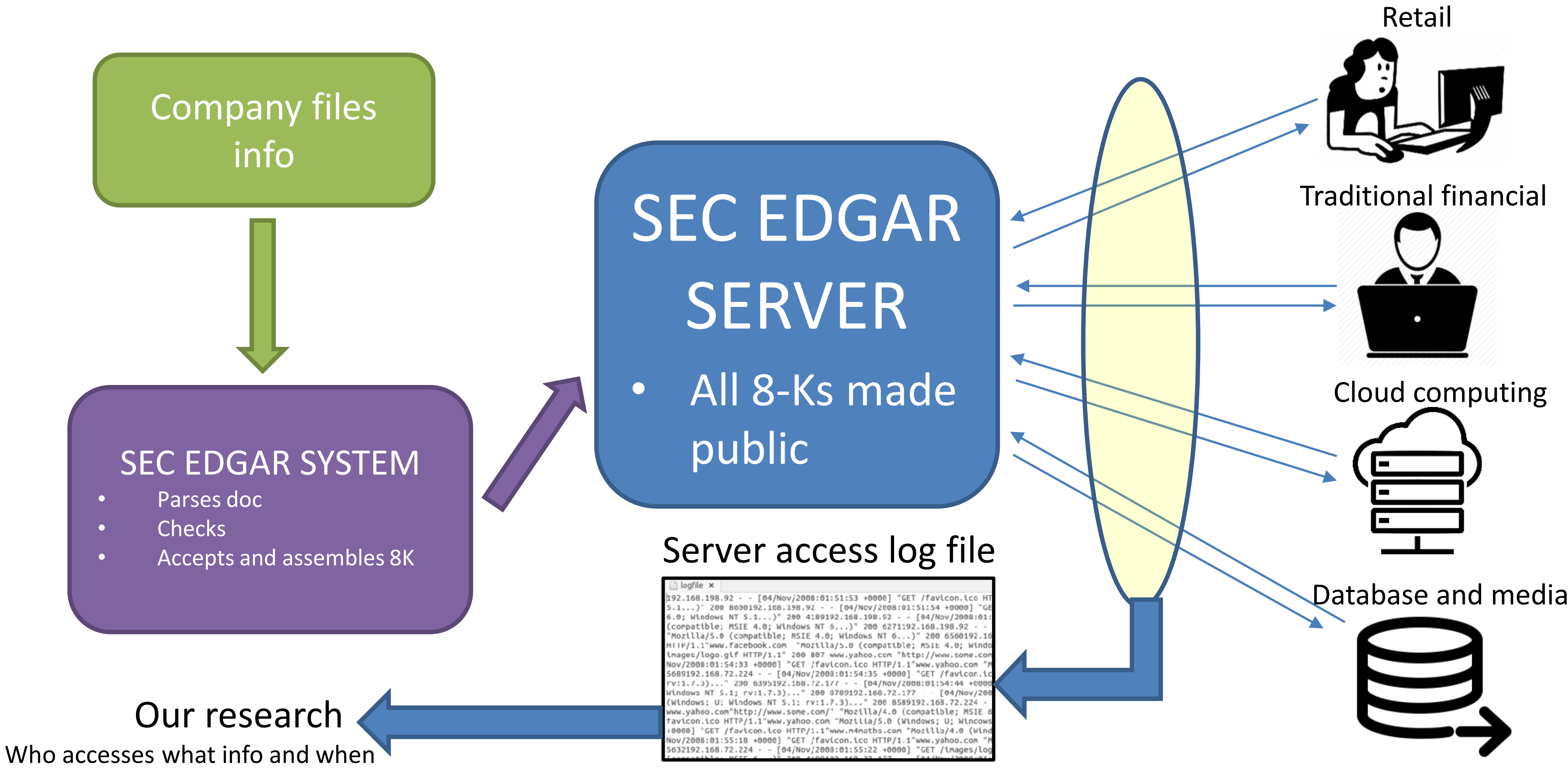
	Production	Deliveries	Subject to operating lease accounting
Model S/X	2,340	1,890	18%
Model 3/Y	204,081	199,360	7%
Total	**206,421**	**201,250**	**7%**

****

Our net income and cash flow results will be announced along with the rest of our financial performance when we announce Q2 earnings. Our delivery count should be viewed as slightly conservative, as we only count a car as delivered if it is transferred to the customer and all paperwork is correct. Final numbers could vary by up to 0.5% or more. Tesla vehicle deliveries represent only one measure of the company's financial performance and should not be relied on as an indicator of quarterly financial results, which depend on a variety of factors, including the cost of sales, foreign exchange movements and mix of directly leased vehicles.

- Mix of hard (numbers) and soft (language) info
- Range from completely unstructured/unstandardised to fairly standardised

Filing and accessing 8-K information



The SEC’s EDGAR Server log file

- 14 year period: 2003 to 2016
- Data on each “viewing” (referred to as a “visit”) of an 8-K
- 4 billion visits, multiple terabytes
- time stamp, HTTP status codes, IP address (partial redaction), crawler flag, …

Use IP addresses (MaxMind IP lookup + TR Ownership + Capital IQ) to classify users:

- Cloud computing users
- Traditional financial institutions
- Database / media
- Internet service providers (retail)
- Regulators and education / other

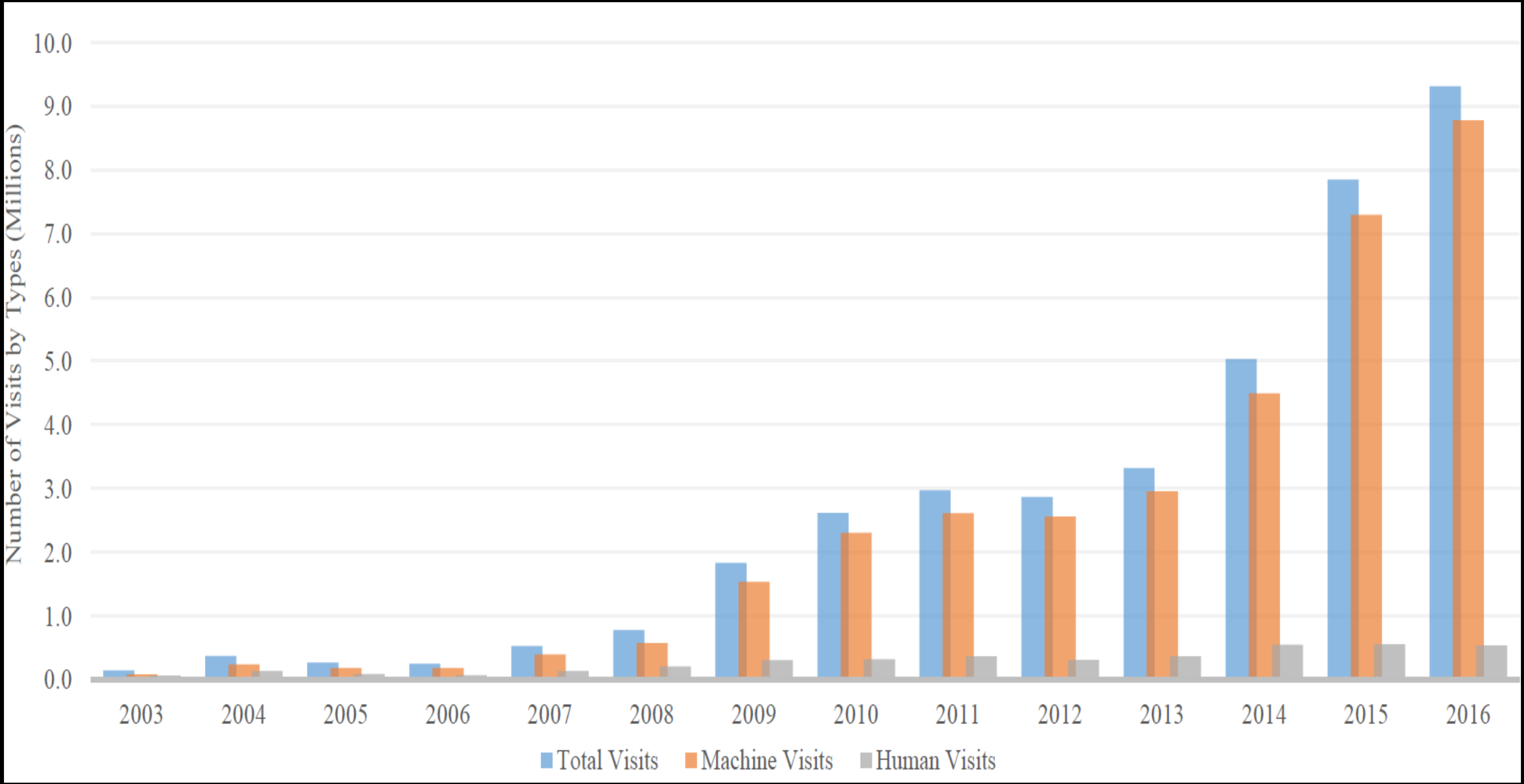
Use access patterns and reaction times to classify:

- Humans vs Machines
  - e.g., >5 downloads per minute or >1,000 per day,
  - + 2 other methods … 96% agreeance

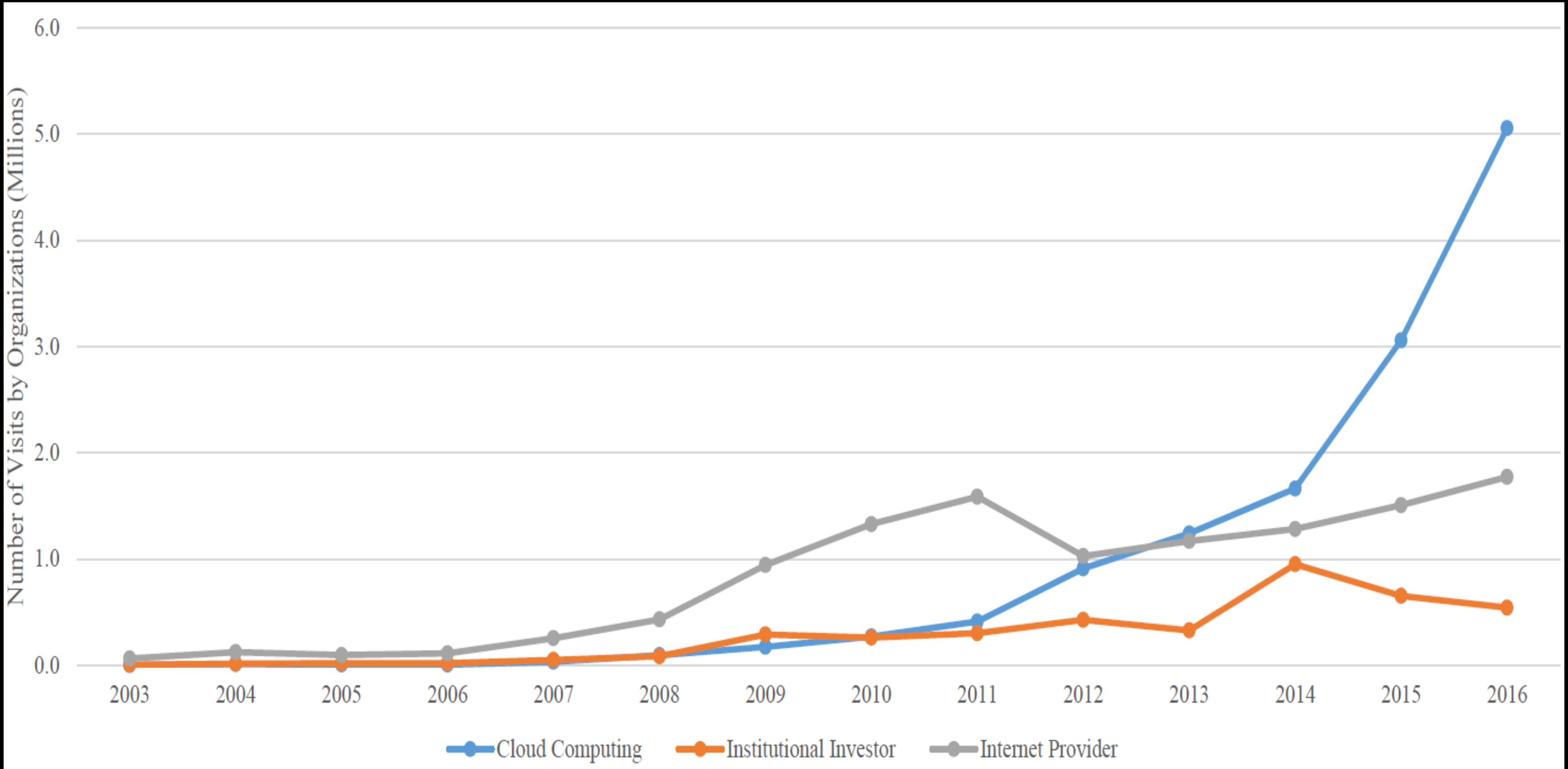
Link all that with

CRSP
Compustat
IBES
Refinitiv
NYSE TAQ
SEC MIDAS

Humans vs machines through time



The growing importance of cloud computing machines



What information do machines access? How does that compare to humans?

	$TotalVisits$	$Machine$	$Human$
	(1)	(2)	(3)
$FinNeg$	0.984*	-0.028	4.756***
	(1.76)	(-0.05)	(9.02)
FOG	0.000	0.000	-0.002
	(0.31)	(0.49)	(-1.21)
$WordCount$	0.050***	0.051**	0.060***
	(3.21)	(2.99)	(5.85)
$DayRelease$	-0.021***	-0.014***	-0.055***
	(-7.09)	(-5.00)	(-7.87)
$\#Item$	0.098***	0.081***	0.150***
	(11.58)	(8.41)	(16.56)
BM	0.009**	0.001	0.037***
	(2.28)	(0.22)	(5.99)
$SIZE$	0.012	0.007	0.033***
	(1.37)	(0.89)	(3.08)
$InstOwn$	0.004	-0.007	-0.148***
	(0.09)	(-0.17)	(-3.12)
$Analysts$	0.000	-0.002	0.009
	(0.04)	(-0.19)	(0.72)
Firm FE	Yes	Yes	Yes
Year FE	Yes	Yes	Yes
Nobs	556,283	556,283	556,283
Adj. R^2	0.845	0.348	0.404

Humans drawn to negative sentiment news / machines are sentiment neutral

Both pay more attention to bigger 8-Ks that likely contain more info

Humans pay more attention to bigger stocks and value stocks / machines care uniformly about the cross-section

➔ Limited capacity of humans ➔ must allocate scarce attention

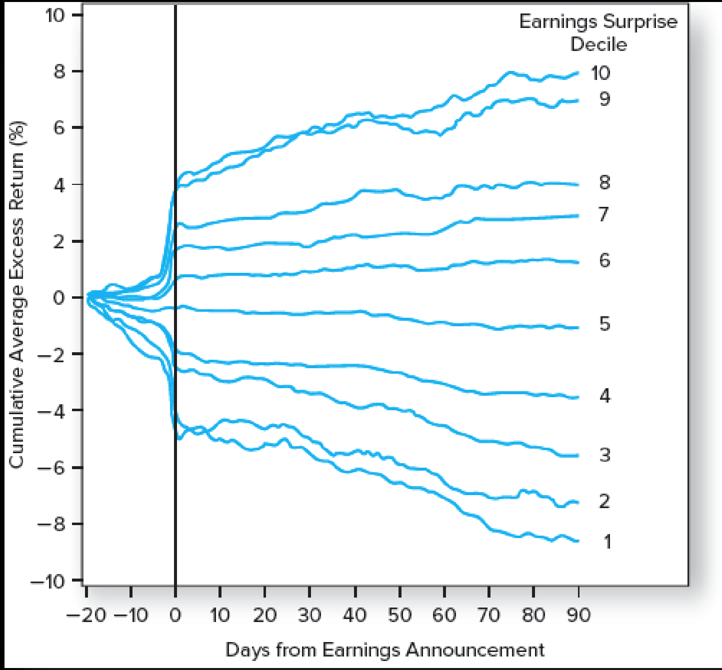
---

By 8-K content, humans drawn to more specific, anticipated info, machines not

Impact on market efficiency?

Measure drift (underreaction) + overshoots

- Recall PEAD → drift:



Measure the (inefficient) drift but also capture overreaction (if present):

$$CAR_{i,t}^{0,T} = \sum_{t=0}^{T}\left(r_{i,t} - \alpha_i - \sum_{k=1}^{k}\beta_{i,k} f_{m,t}\right) = \sum_{t=0}^{T}\varepsilon_{i,t}$$

$$DRIFT(2,T) = \left|CAR_{i,t}^{0,T} - CAR_{i,t}^{0,1}\right|$$

Second measure:
Separate information + noise with variance decomposition

- Brogaard, Nguyen, Putnins, Wu (2021):

$$r_t = \underbrace{\mu}_{\text{discount rate}} + \underbrace{\theta_{r_m}\varepsilon_{r_m,t}}_{\text{market-wide info}} + \underbrace{\theta_x \varepsilon_{x,t}}_{\text{private info}} + \underbrace{\theta_r \varepsilon_{r,t}}_{\text{public info}} + \underbrace{\Delta s_t}_{\text{noise}} \; ,$$

$$r_{m,t} = \sum_{l=1}^{5} a_{1,l} r_{m,t-l} + \sum_{l=1}^{5} a_{2,l} x_{t-l} + \sum_{l=1}^{5} a_{3,l} r_{t-l} + \varepsilon_{r_m,t}$$

$$x_t = \sum_{l=0}^{5} b_{1,l} r_{m,t-l} + \sum_{l=1}^{5} b_{2,l} x_{t-l} + \sum_{l=1}^{5} b_{3,l} r_{t-l} + \varepsilon_{x,t}$$

$$r_t = \sum_{l=0}^{5} c_{1,l} r_{m,t-l} + \sum_{l=0}^{5} c_{2,l} x_{t-l} + \sum_{l=1}^{5} c_{3,l} r_{t-l} + \varepsilon_{r,t},$$

$$MktInfoShare = \theta_{r_m}^2 \sigma_{\varepsilon_{r_m}}^2 / (\sigma_w^2 + \sigma_s^2)$$

$$PrivateInfoShare = \theta_x^2 \sigma_{\varepsilon_x}^2 / (\sigma_w^2 + \sigma_s^2)$$

$$PublicInfoShare = \theta_r^2 \sigma_{\varepsilon_r}^2 / (\sigma_w^2 + \sigma_s^2)$$

$$NoiseShare = \sigma_s^2 / (\sigma_w^2 + \sigma_s^2).$$

Overall impacts on efficiency

$$DRIFT(2,T)_{i,t} = \alpha_0 + \sum_{j=1}^{2} \beta_j V_{i,t}^{M,H} + \sum_{j=1}^{k} \varphi_j \Gamma_{i,j,t} + \tilde{f} + \tilde{\tau} + \varepsilon_{i,t}$$

	$DRIFT(2,10)$			
	(1)	(2)	(3)	(4)
$CloudMachine$	-0.002***			
	(-3.68)			
$InstMachine$		0.002		
		(1.24)		
$DataMachine$			-0.001	
			(-0.39)	
$OtherMachine$				0.002**
				(2.68)
Controls	Yes	Yes	Yes	Yes
Firm FE	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes
Nobs	517,848	517,848	517,848	517,848
Adj. R^2	0.180	0.180	0.180	0.180

1) *Cloud computing machines* consistently *improve* efficiency

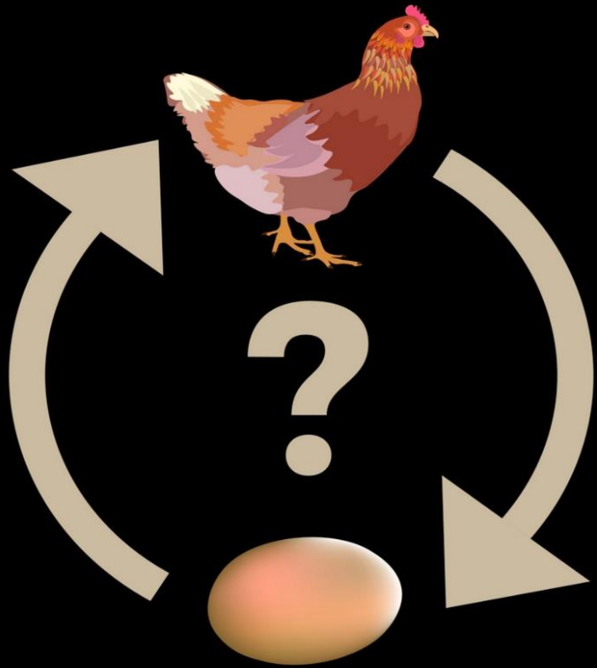
   ➔ impact is significant out to 20 days post 8-K

2) *Machines collectively* have an *insignificant* impact, because of the heterogeneity in machine types

   ➔ not all machines are equal!

3) More *humans* accessing information does *not* help efficiency (can harm)

Which way does causality run?



Identifying causality

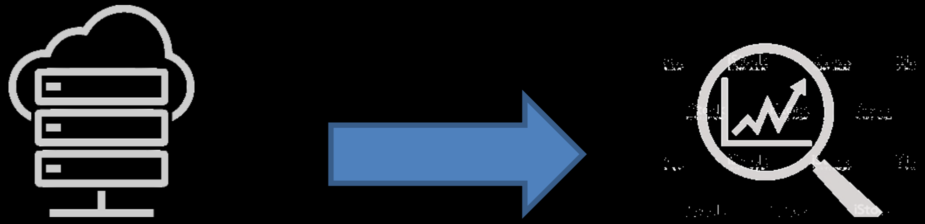
- Three identification strategies:

1) Exploit exogenous cloud computing server outages
   - Cloud servers are fairly robust, but they do go down! (hundreds of times in our sample)
   - Also exploit electricity outages that affect humans

2) Instrumental variables, exploiting the fact that human viewership is constrained on high macro news days, concentrated in certain stocks, influenced by sentiment, but machines are not

3) Index additions/deletions ➔ disproportionately impact attention-constrained humans compared to unconstrained machines

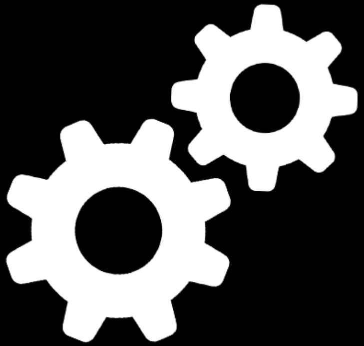
➔ All three suggest causality is:



Cloud machines

Efficiency

Mechanism?
Machine views → informed trades → price discovery?



Machine viewership ➔ informed trading … but human viewership does not

- PIN = Probability of Informed Trading (Easley and O'Hara)

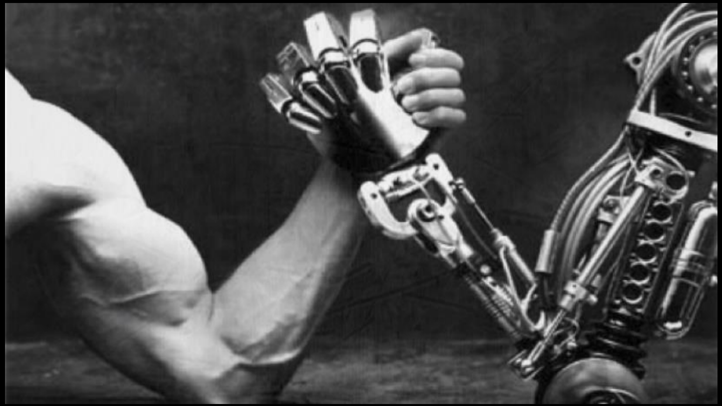
	Daily Average $PIN(0,1)$				Daily Average $PIN(0,5)$			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
$Machine$	0.002**				0.002***			
	(2.61)				(3.06)			
$Human$		-0.002***				-0.002***		
		(-5.86)				(-4.59)		
$CloudMachine$			0.001*				0.002**	
			(1.88)				(2.40)	
$InstMachine$				-0.000				0.000
				(-0.02)				(0.11)
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Firm FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Nobs	500,817	500,817	471,747	471,747	500,817	500,817	471,747	471,747
Adj. R^2	0.591	0.591	0.597	0.597	0.674	0.674	0.679	0.679

Cloud machine viewership ➔ algo trading post 8-K … human viewership does not

	OddLotRatio(0,1)				*TradeSize(0,1)*			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Machine	0.000				0.000			
	(0.19)				(0.45)			
Human		-0.009**				0.005***		
		(-2.93)				(7.47)		
CloudMachine			0.005**				-0.002**	
			(3.92)				(-2.78)	
InstMachine				-0.002				0.002
				(-0.57)				(1.67)
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Firm FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Nobs	165,821	165,821	165,821	165,821	165,821	165,821	165,821	165,821
Adj. R^2	0.101	0.101	0.101	0.101	0.534	0.534	0.534	0.534

When do machines have an edge?

When do humans have an edge?



#1: Readability of the info

- Known: Companies increasingly cater to machines by preparing filings with higher machine readability (Cao et al., 2020) ➔ adds to machine advantage
- Known: Companies with disclosures that humans find difficult to read trade at significant discount to fundamentals (Hwang and Kim, 2017) ➔ induces uncertainty and distrust among ***humans*** (not machines)

➔ Machines might have advantage when info is difficult for humans

➔ Confirmed in the data: Machines have stronger positive impact in *linguistically complex 8-Ks (Gunning FOG + Flesch-Kincaid measures)*

#2: Sentiment and bias

- Known: Emotion interferes with decision-making
- Known: Humans struggle to process bad news rationally and tend to overreact (Tetlock, 2007) ➔ excessive pessimism
  - An asymmetric bias. Machines should not be affected

➔ Machines might contribute more to efficiency in high-sentiment settings and high pessimism bias setting

➔ Confirmed in the data: Machines have stronger impact on efficiency in filings with many negative words

#3: Sequential/repeat information

- Known: Trade-off between fast+noisy and slow+accurate processing of information (Dugast and Foucault, 2018) ➔ many machines are trained to respond quickly to information in isolation
    - e.g., predict whether a given announcement => P↑ or P↓
  - ➔ Combining sequential, incremental information is how humans can make slow, but good decisions (difficult for a machine)
  - ➔ Repeat information can be incorrectly interpreted by a machine as a new signal

➔ Expect humans will have an edge when information is sequential with some repetition and some incremental element

➔ Confirmed in the data: In Item 2.02 (largely repeats previously disclosed financial info), humans are more effective than machines

Summary

- The *most sophisticated machines* (cloud computing users) consistently drive ↑ efficient price discovery around info events

- Not all machines are *beneficial* to market efficiency

- Channel: Cloud machines ➔ Informed and algo trades ➔ efficiency

- Machines excel:
  - Where humans are most prone to bias (e.g., pessimism),
  - When humans are constrained (small stocks, busy days)
  - Linguistic complexity

- Humans have an edge:
  - Combining sequential information that overlaps
  - Soft information

Thank you

Contact: Talis Putnins

talis.putnins@uts.edu.au

Further details in the paper: https://ssrn.com/abstract=3783221