Market Efficiency in the Age of Machine Learning

Leonidas Barbopoulos
U. of Edinburgh

Rui Dai
WRDS

Tālis Putniņš
UTS / SSE @ Riga

Anthony Saunders
NYU Stern

AI & Machine Learning in Finance
Stockholm, Sweden
August 22 – 23, 2022

Introduction

- As machines become more prevalent in financial markets, *how is (informational) efficiency impacted?*

It's not obvious whether...

- Market efficiency will be **improved**:
  - Machines can process **more data** than a human.
  - Machines can interpret data **faster** than a human.
  - Machines are less susceptible to **emotions and bias**.
- Market efficiency will be **deteriorated**:
  - Machines cannot deal with **"soft" information** (...it can't be quantified).
  - **Overfitting** ML models can result in trading on **spurious correlations**.
  - Machines learn to predict the actions of humans, i.e., **front-run** humans (Van Kervel and Menkveld, 2019).
    - However, they do not bring new information.
  - Trade-off between speed and accuracy in decision-making, squeezing out slow, deliberating humans, could come at a **cost of accuracy**.

Theoretical background

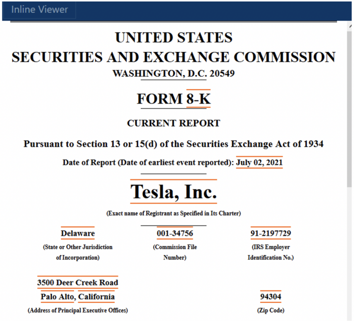
- Information acquisition costs and processing (Grossman and Stiglitz, 1980; Kyle, 1985, 1989; Admati and Pfleiderer, 1988).
- The cost of gathering and analyzing information has declined as machines replace humans (Chen et al., 2020).
- Increased computational power and the use of ML/AI algorithms help investors to:
  - Outperform traditional empirical asset pricing models (Gu et al., 2020).
  - Make use of highly unstructured textual data (Bybee et al., 2020).
  - Front-run slower traders (Van Kervel and Menkveld, 2019).

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 $\rightarrow$ they provide a complete record of certain unscheduled info types.
- Filed with the SEC (within 4 days from the event date) 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

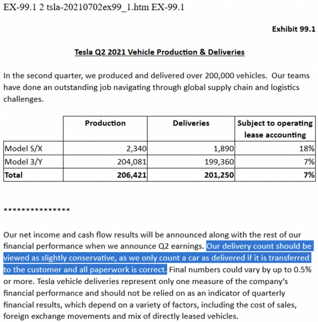
Tesla, Inc.

(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
(Address of Principal Executive Offices)

94304
(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.

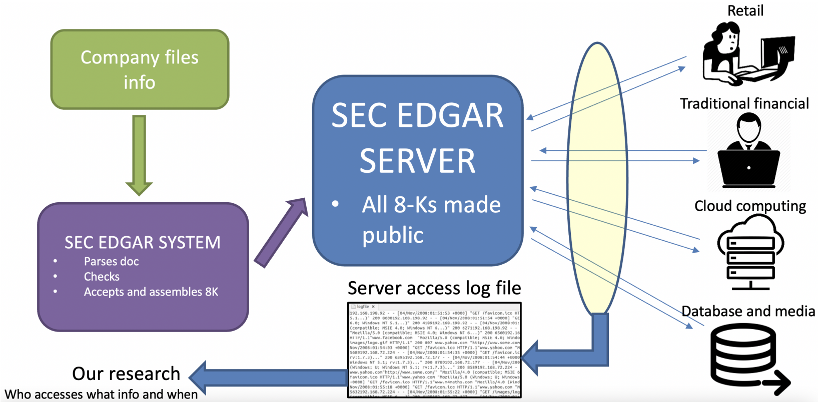
	Production	Deliveries	Subject to operating lease accounting
Model S/X	2,340	1,890	18%
Model 3/Y	204,081	199,360	7%
Total	**206,421**	**201,250**	**7%**

****

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 "download" (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,...

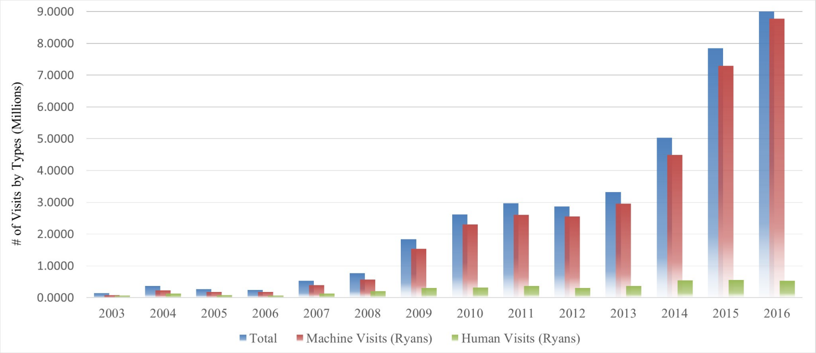
Use IP addresses (MaxMind IP lookup (as in Chen et al. (2020)) + TR Ownership + Capital IQ) to classify users/organisations:

Organizations of 8-K viewers	and	**Link all that with**
Cloud computing users		CRSP
Traditional financial institutions		Compustat
Database / media		IBES
Internet service providers (retail)		Refinitiv
Regulators and education		NYSE TAQ
Other		SEC MIDAS

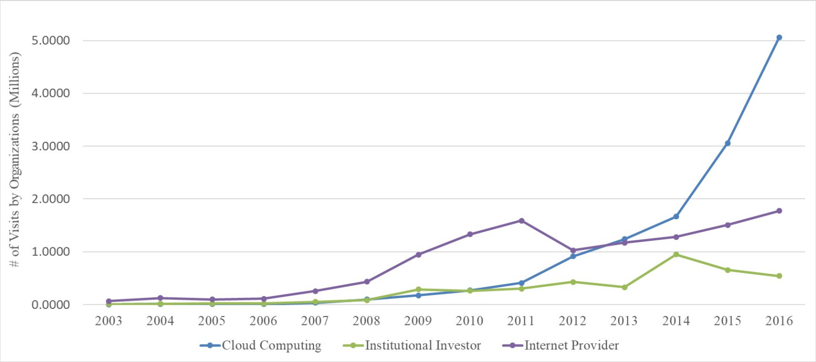
Use access patterns and reaction times to classify:
Humans vs Machines:

- e.g., >5 downloads per minute or >1,000 per day (Drake et al., 2015).
- e.g., >25 downloads per minute or >500 per day (Dechow et al., 2015).

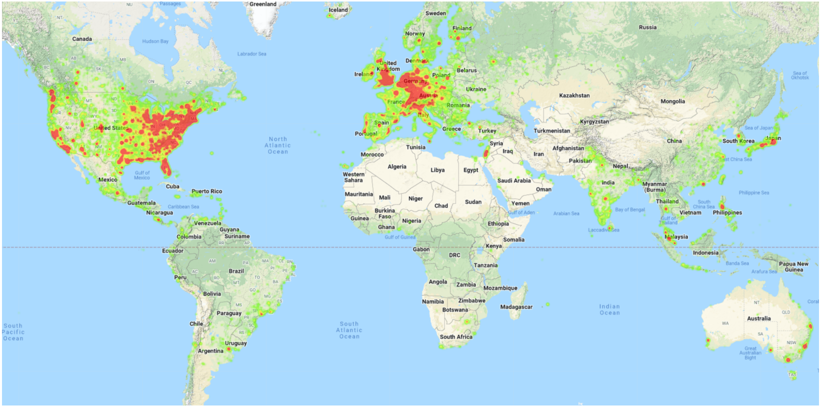
Humans vs machines through time



The growing importance of cloud computing machines



The growing importance of cloud computing machines



Housekeeping

- What information (8-Ks) do machines or humans access?
- Impact on market efficiency?
- Identifying causality.
- Mechanisms?
- When do machines or humans have an edge?

What info do machines access? Compare to humans?

	DLR				DRT		
	Total Visits	*Human*	*Machine*	*Cloud Machine*	*Human*	*Machine*	*Cloud Machine*
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
FinNeg	1.072*	5.404***	0.040	-1.445**	6.082***	0.229	-1.415**
	(1.99)	(12.75)	(0.08)	(-2.73)	(11.84)	(0.43)	(-2.65)
#Words	0.048***	0.048***	0.048**	0.054**	0.039***	0.051***	0.054***
	(3.06)	(6.29)	(2.77)	(2.94)	(5.31)	(3.08)	(3.05)
Gap	-0.014***	-0.042***	-0.009***	-0.011**	-0.034***	-0.011***	-0.011**
	(-5.35)	(-7.38)	(-3.49)	(-2.55)	(-7.34)	(-4.09)	(-2.54)
#Items	0.108***	0.160***	0.089***	0.049***	0.111***	0.101***	0.053***
	(11.89)	(19.04)	(8.48)	(4.11)	(13.10)	(9.82)	(4.15)
#News	0.006	0.053***	-0.000	-0.009	0.028***	0.004	-0.009
	(1.16)	(11.03)	(-0.05)	(-1.69)	(7.66)	(0.81)	(-1.58)
BM	0.016**	0.057***	0.002	-0.007	0.058***	0.006	-0.006
	(2.90)	(6.93)	(0.31)	(-1.06)	(7.55)	(1.18)	(-0.95)
SIZE	0.015	0.031**	0.009	-0.009	0.036***	0.010	-0.008
	(1.70)	(2.96)	(1.07)	(-1.51)	(3.76)	(1.14)	(-1.37)
Cld.Outage	-0.045**	0.009	-0.048**	-0.070***	-0.007	-0.048**	-0.071***
	(-2.42)	(0.68)	(-2.23)	(-4.38)	(-0.48)	(-2.43)	(-4.71)
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Firm FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Nobs	514,731	514,731	514,731	514,731	514,731	514,731	514,731
Adj. R^2	0.854	0.400	0.879	0.912	0.345	0.867	0.907

- Humans (H) drawn to negative sentiment news / machines (M) are sentiment-neutral.
- Larger 8-Ks (word count and # of items) receive more attention by both M & H.
- All viewers (H; M) are less attentive to 8-Ks published later in the 4-day window.
- H pay more attention to bigger stocks and value stocks / M care uniformly about the cross-section $\rightarrow$ limited capacity of humans $\rightarrow$ must allocate limited attention.
- By 8-K content (unreported), H drawn to more specific, anticipated info, M not.

Impact on market efficiency?

Price drift measure (primary measure)

- The ACAR of Ball and Brown (1968) and Fama et al. (1969) is a standard measure of the information aggregation into prices.
- It measures the post-event price drift, net of a predicted returns from a factor model.
- It measures the (inefficient) drift but also capture overreaction (if present):

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

- The cumulative abnormal return (CAR) cumulates the abnormal return from announcement date $t = 0$ to T.
- We use the absolute difference between post-announcement price variation and variation over a short window right after the announcement:

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

with $T > 1$ to capture post announcement drift.

Alternative measures

For robustness, we:

- Examine the *scaled* version of the drift measure, normalizing it by the total information content of the information release.
- Separate information + noise with variance decomposition (Brogaard et al., 2022).

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

$$r_{m,t} = \sum_{l=1}^{5} \alpha_{1,l} r_{m,t-l} + \sum_{l=1}^{5} \alpha_{2,l} x_{t-l} + \sum_{l=1}^{5} \alpha_{3,l} r_{t-l} + \varepsilon_{r_m,t} \qquad MktInfoShare = \frac{\theta_{r_m}^2 \sigma_{\varepsilon_{r_m}}^2}{(\sigma_w^2 + \sigma_s^2)}$$

$$x_t = \sum_{l=1}^{5} b_{1,l} r_{m,t-l} + \sum_{l=1}^{5} b_{2,l} x_{t-l} + \sum_{l=1}^{5} b_{3,l} r_{t-l} + \varepsilon_{x,t} \qquad PrvInfoShare = \frac{\theta_x^2 \sigma_{\varepsilon_x}^2}{(\sigma_w^2 + \sigma_s^2)}$$

$$r_t = \sum_{l=1}^{5} c_{1,l} r_{m,t-l} + \sum_{l=1}^{5} c_{2,l} x_{t-l} + \sum_{l=1}^{5} c_{3,l} r_{t-l} + \varepsilon_{r,t} \qquad PubInfoShare = \frac{\theta_r^2 \sigma_{\varepsilon_r}^2}{(\sigma_w^2 + \sigma_s^2)}$$

$$NoiseShare = \frac{\sigma_s^2}{(\sigma_w^2 + \sigma_s^2)}$$

- We use the *NoiseShare*, which is based on a variance decomposition model that **partitions stock return variance into information of various types**, and **noise** (temporary price deviations from efficient prices, driven by under or over reaction).

Impact of machine and human viewership on price drift

	$DRIFT(2,10)$			$DRIFT(2,20)$		
	(1)	(2)	(3)	(4)	(5)	(6)
Machine	0.003		0.002	0.005		0.003
	(1.09)		(0.68)	(1.23)		(0.79)
Human		0.002***	0.002***		0.003***	0.003***
		(4.27)	(5.87)		(4.26)	(5.55)
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Firm FE	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
NObs	295,142	295,142	295,142	295,142	295,142	295,142
Adj. R^2	0.221	0.222	0.222	0.243	0.243	0.243

- More downloads by human viewers are associated with a larger magnitude of price drift.
- A one standard deviation increase in human viewership is associated with a 13.23% aggregate increase in the price drift over the post-publication month $[DRIFT(2,20)]$.

Machine types and price drift

	DRIFT(2, 10)			*DRIFT*(2, 20)		
	(1)	(2)	(3)	(4)	(5)	(6)
CloudMachine	-0.002***			-0.002**		
	(-3.33)			(-2.34)		
InstInvMachine		0.001			0.002	
		(1.24)			(1.29)	
DataVendMachine			-0.001			-0.001
			(-0.44)			(-0.37)
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Firm FE	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
NObs	295,142	295,142	295,142	295,142	295,142	295,142
Adj. R^2	0.221	0.221	0.221	0.242	0.242	0.242

- Increased downloads of 8-K filings by cloud-computing services (IT firms) are associated with a smaller magnitude of price drift.
- A one standard deviation increase in the machine-based viewership of 8-K from cloud computing services is associated with a 10.49% aggregate decrease in the price drift over the post-filing month [$DRIFT(2, 20)$].

Alternative measures (*scale drift*); the impact of viewership on price variance due to *noise* trading

	Scaled *DRIFT*(2, 10)			Scaled *DRIFT*(2, 20)			*Noise*		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Machine	-0.024			0.005			0.114		
	(-1.38)			(0.46)			(0.90)		
CloudMachine		-0.037**			-0.026***			-0.089**	
		(-2.34)			(-3.64)			(-2.57)	
Human			0.001			0.028***			0.062**
			(0.13)			(3.63)			(2.68)
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Firm FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
NObs	286,852	286,852	286,852	286,852	286,852	286,852	259,478	259,478	259,478
Adj. R^2	0.033	0.033	0.033	0.037	0.037	0.037	0.523	0.523	0.523

- Increased downloads of 8-K filings by cloud-computing services are associated with a smaller variance due to noise trading.

Identifying causality

Three identification strategies:

- **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/power outages that affect humans.
- **Index additions/deletions** → disproportionately impact attention-constrained humans compared to unconstrained machines.
- **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.

All three suggest causality is: Cloud machines → Efficiency.

Evidence from cloud and power outages: 2SLS

	First Stage			Second Stage					
	Machine	*CloudMachine*	*Human*	$DRIFT(2,10)$			$DRIFT(2,20)$		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Machine				-0.050*			-0.002		
				(-2.03)			(-0.06)		
CloudMachine					-0.036**			-0.001	
					(-2.67)			(-0.06)	
Human						0.158			0.006
						(1.19)			(0.06)
CloudOutage	-0.050**	-0.068***	0.016						
	(-2.10)	(-3.63)	(1.57)						
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Firm FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
NObs	295,142	295,142	295,142	295,142	295,142	295,142	295,142	295,142	295,142
Weak ID F-Value	4.361	13.157	2.661	-	-	-	-	-	-

- During cloud service outage days, we find that **the relationship between cloud machine viewership and price drift is less negative**.

Evidence from cloud and power outages: Outage dummies

	Cloud Outage Day						*Power Outage Day*					
	$NDRIFT(2,10)$			$NDRIFT(2,20)$			$NDRIFT(2,10)$			$NDRIFT(2,20)$		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
CloudMachine	-0.002***			-0.002**			-0.002***			-0.002**		
	(-4.11)			(-2.40)			(-3.70)			(-2.26)		
CloudMachine × *OutageDay*	0.004***			0.004**			0.000			-0.001		
	(3.63)			(2.50)			(0.32)			(-0.40)		
Machine		0.003			0.005			0.003			0.005	
		(0.95)			(1.06)			(0.97)			(1.09)	
Machine × *OutageDay*		0.002***			0.002**			0.002**			0.002	
		(3.20)			(2.22)			(2.35)			(1.12)	
Human			0.003***			0.004***			0.003***			0.004***
			(4.35)			(4.12)			(4.84)			(4.42)
Human × *OutageDay*			0.001			0.001			0.004			0.003
			(1.21)			(0.70)			(1.48)			(1.06)
OutageDay	-0.008**	-0.006*	0.000	-0.010*	-0.009	-0.001	0.001	-0.006*	-0.003*	0.003	-0.005	-0.002
	(-2.83)	(-1.86)	(0.19)	(-2.02)	(-1.70)	(-0.20)	(0.34)	(-2.01)	(-1.93)	(0.41)	(-1.04)	(-0.88)
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Firm FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Nobs	517,739	517,739	517,739	517,739	517,739	517,739	517,739	517,739	517,739	517,739	517,739	517,739
Adj. R^2	0.206	0.206	0.207	0.235	0.235	0.236	0.205	0.206	0.207	0.235	0.235	0.236

- During cloud service outage days, we find that **the relationship between cloud machine viewership and price drift is less negative**.

The effects of S&P500 inclusion on machines and humans

	DRIFT(2, 10)			*DRIFT*(2, 20)		
	(1)	(2)	(3)	(4)	(5)	(6)
Machine	0.003			0.005		
	(1.09)			(1.24)		
Machine × *InIndex*	-0.000			0.000		
	(-0.24)			(0.06)		
CloudMachine		-0.002***			-0.002**	
		(-3.35)			(-2.32)	
CloudMachine × *InIndex*		0.000			0.000	
		(1.39)			(0.95)	
Human			0.002***			0.004***
			(4.98)			(5.14)
Human × *InIndex*			-0.001**			-0.001*
			(-2.18)			(-1.82)
InIndex	0.003	0.002	0.004**	0.006**	0.005**	0.008***
	(1.58)	(0.99)	(2.46)	(2.56)	(2.51)	(3.39)
Control	Yes	Yes	Yes	Yes	Yes	Yes
Firm FE	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
NObs	295,142	295,142	295,142	295,142	295,142	295,142
Adj. R^2	0.221	0.221	0.222	0.243	0.243	0.243

- Index inclusion attracts (more) sophisticated and skilled investors, whose viewership is more likely to improve price discovery than that of average human viewers.
- After index inclusion, we find that the **relationship between human viewership and price drift is less positive**.

Instrumental variables tests

	First Stage			Second Stage	
	Machine	*CloudMachine*	*Human*	$DRIFT(2,10)$	$DRIFT(2,20)$
	(1)	(2)	(3)	(4)	(5)
Human				0.077**	0.096**
				(2.83)	(2.60)
InvSent	-0.306*	0.411***	-0.269***		
	(-1.70)	(3.19)	(-3.31)		
MacroNews	0.110**	0.028	0.112**		
	(2.34)	(1.10)	(2.08)		
Controls	Yes	Yes	Yes	Yes	Yes
Firm FE	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes
NObs	295,142	295,142	295,142	295,142	295,142
Weak ID F-Value	4.304	5.179	28.883	-	-
Hansen J-Stats	-	-	-	0.339	0.310

Our IVs are:

1. Investor sentiment index (Baker and Wurgler, 2006)
2. A dummy variable for major macro news announcement (Flannery and Protopapadakis, 2002; Savor and Wilson, 2013)

Mechanism?

Machine views $\rightarrow$ informed trades $\rightarrow$ price discovery?

Machine viewership → informed trading... but human viewership does not

PIN = Probability of Informed Trading (Easley and O'Hara, 2004)

	$PIN(0,1)$			$PIN(0,5)$		
	(1)	(2)	(3)	(4)	(5)	(6)
Machine	0.002***			0.002***		
	(3.27)			(3.12)		
CloudMachine		0.002**			0.002**	
		(2.52)			(2.46)	
Human			-0.002***			-0.002***
			(-5.24)			(-3.67)
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Firm FE	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
NObs	271,079	271,079	271,079	277,526	277,526	277,526
Adj. R^2	0.633	0.633	0.633	0.686	0.686	0.686

Cloud machine viewership → algorithmic trading post 8-K... human viewership does not

	OddLotRatio(0, 1)			*TradeSize*(0, 1)		
	(1)	(2)	(3)	(4)	(5)	(6)
Machine	0.004**			-0.003		
	(2.61)			(-1.91)		
CloudMachine		0.006*			-0.002*	
		(2.44)			(-2.26)	
Human			-0.005**			0.003**
			(-3.45)			(3.69)
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Firm FE	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
NObs	101,368	101,368	101,368	101,368	101,368	101,368
Adj. R^2	0.203	0.203	0.203	0.586	0.586	0.586

High-Frequency Traders (HFT) and price drift

	DRIFT(2, 10)						*DRIFT*(2, 20)					
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Machine	-0.001	-0.001					-0.001	-0.001				
	(-1.37)	(-1.38)					(-0.92)	(-1.02)				
CloudMachine			-0.002**	-0.002**					-0.002*	-0.002*		
			(-3.02)	(-2.80)					(-2.27)	(-2.15)		
Human					0.001**	0.001**					0.001**	0.001**
					(3.69)	(2.71)					(2.90)	(2.57)
Cancel-to-Trade	0.000		0.000		0.000		0.000		0.000		0.000	
	(0.18)		(0.16)		(0.15)		(0.69)		(0.67)		(0.66)	
Trade-to-Order		0.001**		0.001**		0.001**		0.001**		0.001**		0.001**
		(2.61)		(2.70)		(2.72)		(3.99)		(3.96)		(3.88)
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Firm FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
NObs	101,368	101,368	101,368	101,368	101,368	101,368	101,368	101,368	101,368	101,368	101,368	101,368
Adj. R^2	0.250	0.250	0.249	0.250	0.250	0.250	0.279	0.279	0.279	0.279	0.279	0.279

- Earlier studies show that HFTs, which are also automated machine-based traders, improve price discovery at intra-day horizons (Brogaard et al., 2014).
- We use the *Cancel-to-Trade* ratio and the *Trade-to-Order* ratio (an inverse measure of HFT activity).
- The positive impact of cloud machine access of 8-Ks on info efficiency *is not* diminished by controlling for HFT activity.
- Hence, our paper and the HFT-literature suggest that machines can benefit info efficiency in multiple ways and at different horizons.

When do machines have an edge?

When do humans have an edge?

Does numerical information increase the impact of machines?

	DRIFT(2, 10)			*DRIFT*(2, 20)		
	(1)	(2)	(3)	(4)	(5)	(6)
Machine	0.003			0.005		
	(1.24)			(1.45)		
Machine $\times$ *NER*	-0.006			-0.012		
	(-1.27)			(-1.22)		
CloudMachine		-0.002**			-0.001	
		(-2.87)			(-1.66)	
CloudMachine $\times$ *NER*		-0.011**			-0.015**	
		(-2.53)			(-2.21)	
Human			0.002***			0.004***
			(5.02)			(5.72)
Human $\times$ *NER*			-0.008			-0.015
			(-1.19)			(-1.42)
NER	-0.022	-0.022	-0.039**	-0.053	-0.064*	-0.084**
	(-0.94)	(-1.48)	(-2.60)	(-1.04)	(-2.01)	(-2.54)
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Firm FE	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
NObs	291,848	291,848	291,848	291,848	291,848	291,848
Adj. R^2	0.222	0.221	0.222	0.243	0.243	0.243

- We test whether machines have a comparative advantage in handling hard, numeric or quantitative information.
- We find that cloud machine 8-K viewership reduces the absolute price drift following 8-K filings with significantly more numerical content.

Negative sentiment

Known: Emotion broadly interferes with decision-making.

Known: Humans struggle to process bad news rationally and tend to overreact (Tetlock, 2007) $\rightarrow$ excessive pessimism.

- There is therefore an asymmetric bias to which **machines should be neutral**.

- Hence, machines might contribute more to price discovery in high-sentiment settings, as well as high pessimism-bias settings.

Negative sentiment and price drift

- Confirmed in the data: Machines have stronger impact on efficiency in filings with many/more negative words (Loughran and Mcdonald, 2011).

	DRIFT(2, 10)			*DRIFT*(2, 20)		
	(1)	(2)	(3)	(4)	(5)	(6)
Machine	0.003			0.005		
	(1.19)			(1.29)		
Machine × *FinNeg*	-0.076**			-0.105*		
	(-2.24)			(-1.88)		
CloudMachine		-0.002**			-0.002*	
		(-2.93)			(-1.89)	
CloudMachine × *FinNeg*		-0.045**			-0.079**	
		(-2.71)			(-2.37)	
Human			0.002***			0.004***
			(3.77)			(4.03)
Human × *FinNeg*			-0.029			-0.055
			(-1.00)			(-1.48)
FinNeg	0.316**	0.148**	0.075	0.470*	0.279*	0.162
	(2.18)	(2.27)	(1.07)	(1.88)	(2.10)	(1.48)
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Firm FE	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
NObs	295,142	295,142	295,142	295,142	295,142	295,142
Adj. R^2	0.222	0.221	0.222	0.243	0.243	0.243

Sequential/repeated information

Known: Trade-off between fast+noisy and slow+accurate processing of information (Dugast and Foucault, 2018) $\rightarrow$ many machines are trained to respond quickly to information in isolation.

  - E.g., predict whether a given announcement leads to price $\uparrow$ or $\downarrow$.

- Combining sequential, incremental information is how humans can make slow, but accurate decisions (difficult for machines).
- Repeated information can be incorrectly interpreted by a machine as a new signal.
- We expect **humans will have an edge** when information is sequential with some repetition and some incremental element.

Sequential/repeated information and price drift

- Confirmed in the data: In *Item2.02* (largely repeats previously disclosed financial info), **humans are more effective than machines in improving price discovery**.

	$DRIFT(2,10)$			$DRIFT(2,20)$		
	(1)	(2)	(3)	(4)	(5)	(6)
Machine	0.003			0.004		
	(1.12)			(1.23)		
Machine × *Item*2.02	-0.000			0.000		
	(-0.15)			(0.34)		
CloudMachine		-0.002***			-0.002**	
		(-3.18)			(-2.39)	
CloudMachine × *Item*2.02		-0.000			-0.000	
		(-1.17)			(-0.20)	
Human			0.003***			0.004***
			(5.15)			(5.20)
Human × *Item*2.02			-0.001***			-0.001**
			(-3.03)			(-2.48)
*Item*2.02	-0.003	-0.002*	-0.002**	-0.006*	-0.006**	-0.005**
	(-1.53)	(-1.96)	(-2.80)	(-1.82)	(-2.43)	(-2.84)
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Firm FE	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
NObs	295,142	295,142	295,142	295,142	295,142	295,142
Adj. R^2	0.222	0.222	0.222	0.243	0.243	0.244

Additional test: Readability and price drift

- Do machines have an edge over humans **when information is harder for humans to read**?

	Gunning FOG				*Flesch-Kincaid*			
	$DRIFT(2,10)$		$DRIFT(2,20)$		$DRIFT(2,10)$		$DRIFT(2,20)$	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Machine	0.004		0.005		0.004		0.005	
	(1.24)		(1.10)		(1.20)		(1.12)	
Machine × *DifficultToRead*	-0.006**		-0.003		-0.006**		-0.004	
	(-2.48)		(-0.75)		(-2.52)		(-0.96)	
Human		0.004***		0.005***		0.004***		0.005***
		(3.95)		(3.79)		(4.05)		(3.90)
Human × *DifficultToRead*		-0.003		-0.003		-0.004		-0.004
		(-1.37)		(-1.29)		(-1.47)		(-1.51)
DifficultToRead	0.009	-0.001	0.006	0.004	0.010*	-0.001	0.007	0.004
	(1.76)	(-0.19)	(0.69)	(0.84)	(1.91)	(-0.32)	(0.85)	(0.89)
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	517,848	517,848	517,848	517,848	517,848	517,848	517,848	517,848
Adj. R^2	0.206	0.207	0.235	0.236	0.205	0.206	0.237	0.237

- Machine 8-K viewership lessens the price drift following 8-Ks by a larger amount for harder-to-read 8-Ks, which is info harder for human readers to interpret.
- Consistent with evidence suggesting that the increasing machine downloading activity motivates firms to prepare filings according to machines' readership, processing capacity, and capability (Cao et al., 2020).

Conclusion

- The most sophisticated machines (cloud computers/IT firms) consistently improve price discovery around information events (8-K publication day).
- Not all machines are beneficial to price discovery.

Mechanism: Cloud machines → Informed and algorithmic trades → price discovery.

- **Machines excel:**
  – High fraction of numerical information (NER) in the 8-K.
  – Less biased by negative sentiment.
  – Less capacity constrained.
  – Difficult-to-Read filing.
- **Humans have an edge:**
  – Combining sequential information that overlaps (i.e., sequential info).
  – Soft information.

References

Admati, A. R., Pfleiderer, P., 1988. A Theory of Intraday Patterns: Volume and Price Variability. The Review of Financial Studies 1, 3–40.

Baker, M., Wurgler, J., 2006. Investor sentiment and the cross-section of stock returns. The Journal of Finance 61, 1645–1680.

Ball, R., Brown, P., 1968. An empirical evaluation of accounting income numbers. Journal of Accounting Research 6, 159–178.

Brogaard, J., Hendershott, T., Riordan, R., 2014. High-frequency trading and price discovery. The Review of Financial Studies 27, 2267–2306.

Brogaard, J., Nguyen, T. H., Putniņš, T. J., Wu, E., 2022. What moves stock prices? the role of news, noise, and information. Review of Financial Studies pp. 1–65.

Bybee, L., Kelly, B. T., Manela, A., Xiu, D., 2020. The structure of economic news. Tech. rep., National Bureau of Economic Research.

Cao, S., Jiang, W., Yang, B., Zhang, A. L., 2020. How to talk when a machine is listening: Corporate disclosure in the age of ai. Working Paper 27950, National Bureau of Economic Research.

Chen, H., Cohen, L., Gurun, U., Lou, D., Malloy, C., 2020. Iq from ip: Simplifying search in portfolio choice. Journal of Financial Economics 138, 118 – 137.

Dechow, P. M., Lawrence, A., Ryans, J. P., 2015. SEC Comment Letters and Insider Sales. The Accounting Review 91, 401–439.

Drake, M. S., Roulstone, D. T., Thornock, J. R., 2015. The determinants and consequences of information acquisition via edgar. Contemporary Accounting Research 32, 1128–1161.

Dugast, J., Foucault, T., 2018. Data abundance and asset price informativeness. Journal of Financial Economics 130, 367 – 391.

Easley, D., O'Hara, M., 2004. Information and the cost of capital. The Journal of Finance 59, 1553–1583.

Fama, E. F., Fisher, L., Jensen, M. C., Roll, R., 1969. The adjustment of stock prices to new information. International Economic Review 10, 1–21.

Flannery, M. J., Protopapadakis, A. A., 2002. Macroeconomic Factors Do Influence Aggregate Stock Returns. The Review of Financial Studies 15, 751–782.

References (Cont'd)

Grossman, S. J., Stiglitz, J. E., 1980. On the impossibility of informationally efficient markets. American Economic Review 70, 393–408.

Gu, S., Kelly, B., Xiu, D., 2020. Empirical Asset Pricing via Machine Learning. The Review of Financial Studies 33, 2223–2273.

Kyle, A. S., 1985. Continuous auctions and insider trading. Econometrica 53, 1315–1335.

Kyle, A. S., 1989. Informed speculation with imperfect competition. The Review of Economic Studies 56, 317–355.

Loughran, T., Mcdonald, B., 2011. When is a liability not a liability? textual analysis, dictionaries, and 10-ks. The Journal of Finance 66, 35–65.

Savor, P., Wilson, M., 2013. How Much Do Investors Care About Macroeconomic Risk? Evidence from Scheduled Economic Announcements. Journal of Financial and Quantitative Analysis 48, 343–375.

Tetlock, P. C., 2007. Giving content to investor sentiment: The role of media in the stock market. The Journal of finance 62, 1139–1168.

Van Kervel, V., Menkveld, A. J., 2019. High-frequency trading around large institutional orders. The Journal of Finance 74, 1091–1137.