

Market Efficiency in the Age of Machine Learning

(discussed by A. Gargano, University of Houston)

L.B. Barbopoulos¹ R. Dai² T. Putnis³ A. Saunders⁴

¹Edinburgh Business School ²WRDS ³University of Technology Sydney ⁴Stern School of Business

AI & Machine Learning in Finance, SSH

Summary of the Paper

- **Research Question:** How does information acquisition by machines affect market efficiency?
 - ▶ *Relevant*
 - ★ New technologies oftentimes produce nuanced effects
 - ★ Efficient Market Hypothesis is (still) one of the most debated topics in finance
 - ▶ *Novel*
 - ★ Many papers on the effect of information acquisition by humans, not much on machines
 - ▶ *Not obvious*
 - ★ Positive, if machines can process a larger amount of info, faster, and in an unbiased fashion
 - ★ Negative, if machines are unable to understand the context
 - ▶ *Stimulates thinking* on the differences between humans and machines
 - ★ We are heading towards a world where humans will interact more with machines
- The paper has clearly already incorporated feedback from many conferences and seminar presentations
 - ▶ Extensive data work

Summary of the Paper

- **Setup:** EDGAR requests of 8-K forms by humans and machines
 - ▶ Electronic system introduced in the '90 by the SEC to improve information dissemination
 - ▶ 8-K filed when new *material information* → Semi-strong efficiency
 - ▶ From I.P. address, 7 requester categories
 - ★ Cloud Computing Facilities, Institutional Investors, Media, Audit Firms, Internet Services, Data Vendor, Other
 - ▶ 2 types of requesters: human and machine
 - ★ Based on the volume of requests
 - ▶ 7×2 total combinations
 - ▶ Market Efficiency: Absolute Price drift in the 20 days after the release of information
 - ★ Larger drift implies lower efficiency
 - ★ Because information at time t is *slowly* incorporated over $t + N$

Main Results

- **Magnitude:** Since 2015, requests from machines have exploded, humans remained constant
- **Determinants of Info Acquisition:**
 - ▶ Similarities: Higher attention to 8-K containing more info (longer and more items) and more timely
 - ▶ Differences
 - ★ Humans: more attention to 8-K with negative news and from firms with high market cap and BM
 - ★ Humans: more attention to Item 2.02 (Earnings) while machine to item 8.01 (Other events)
- **Market Efficiency**
 - ▶ Humans decrease efficiency
 - ★ No heterogeneity: also true for CouldComputing and Institutional Investors
 - ▶ Machines have no effect
 - ★ Heterogeneity: **CouldComputing improves efficiency**, Institutional Investors no effect
- **Endogeneity**
 - ▶ Exogenous Cloud Outages
 - ▶ S&P 500 inclusion
- **Channel:** Sophisticated investors use CloudMachines to process information and trade
 - ▶ More informed trading (PIN) and More Algorithmic trading
 - ▶ Surrounding their requests
- **Humans v.s. machines:** Machines are better at processing
 - ▶ Numerical information
 - ▶ Negative news

Comment 1: Information Acquisition and Disagreement

- The world disagreement never appears in the paper...
- However, **public info can generate disagreement**
 - ▶ Theory: Harris and Raviv (1993), Kandel and Pearson (1995)
 - ▶ Intuition: Investors might interpret the **same information differently**
 - ▶ Formula:
$$\underbrace{p(\theta|data)}_{\text{posterior}} \propto \underbrace{\mathcal{L}(\theta|data)}_{\text{Likelihood / Model}} \underbrace{p(\theta)}_{\text{prior}}$$
 - ★ Same posterior (Agreement) ← uninformed prior and same likelihood
 - ★ Different posteriors (Disagreement) ← different (informed) prior and/or different Likelihood
- Extensive evidence that disagreement can exacerbate mispricing and delay its correction
 - ▶ Sadka and Scherbina (2007)
 - ▶ Gargano, Sotes-Paladino and Verwijmeren (2022)
 - ▶ And references therein
- Typically measured with
 - ▶ Abnormal Trading Volume around information events
 - ▶ Dispersion in Analysts' recommendation
- **Suggestions**
 - 1 Include as a control in regressions of Price Drift → higher disagreement makes drift worse
 - 2 Do machines disagree more than humans? Not obvious to me
 - ★ IF machines process filings in isolations, they might not have "a prior"
 - ★ However, algorithms might generate more dispersion than human

Comment 2: Price Drifts and Short Selling Constraints

- The absolute price drift measure employed does not distinguish between underreaction to
 - ▶ Good news → **positive** price drift
 - ▶ Bad news → **negative** price drift
- The presence of Short Sellers is relevant for both
- **Positive** Drift: Miller (1997) predicts that short-selling constraints
 - ▶ Prevent views from bearish investors to be incorporated
 - ▶ Measures: high shorting fees, high fee volatility, low supply
- **Negative** Drift: Higher short selling activity improves information efficiency (wrt to bad news)
 - ▶ Evidence in Bohemer and Wu (2013) based on the prediction of Diamond and Verrecchia (1987)
 - ▶ Measures: short interest, short volume
 - ▶ Also important for the regressions interacting CloudComputing with Negative sentiment
 - ★ Paper argues that CloudComputing machines are less biased than humans in processing negative news
 - ★ This is also when short sellers might be more active
- **Suggestion**
 - ① Take these effects into account

Comment 3: Tighten the Mechanism (further)

① Who are sophisticated investors behind CloudComputing-Machines?

- ▶ Heterogeneity results beg the question
 - ★ CloudComputing-Machines: improve efficiency
 - ★ CloudComputing-Humans: worsen efficiency
 - ★ Institutional Investors-Machine: no effect
 - ★ Institutional Investors-Human: worsen efficiency
- ▶ The data does not allow you to observe Institutional Investors that use cloud computing
- ▶ Can you at least show that there are sophisticated investors fading out from EDGAR?

② “Market Efficiency in the Age of Machine Learning” or “in the Age of Cloud Computing”?

- ▶ Investors might switch to cloud computing because
 - 1 Secrecy
 - 2 Cost-efficiency
 - 3 Use of Machine Learning / Big Data Analytics
- ▶ Both [1] and [2] could predict improvement in market efficiency because correlated with higher sophistication
- ▶ Can you show that those sophisticated investors who fade out from EDGAR also search for more Machine Learning related workers?
 - ★ E.g. Burning Glass, LinkUp data

Comment 4: The importance of Speed...

- When it comes to info acquisition, it is not only about whether to acquire the info, but how fast
- Sophisticated Investors compete to access information as fast as possible
 - ▶ Trade milliseconds before the public release of macro news
 - ★ Bernile, Hu, Tang (2015) and Pang Wang and Zhu (2017)
 - ▶ Use the FOIA to acquire information before it gets publicized
 - ★ Gargano, Rossi and Wermers (2015)
 - ▶ Form 4 (insider trade) filings are available to paying subscribers before posted to the EDGAR
 - ★ Rogers, Skinner and Zechman (2017)
- Given the timestamps of when the 8-K is released and acquired
 - ▶ You could plot the number of requests (y-axis) on $\Delta = \tau_{released} - \tau_{acquired}$ (x-axis)
 - ▶ It is natural to expect that machines acquire information faster than humans
 - ▶ How much heterogeneity is there across machines?
 - ▶ Are drifts different when information is acquired faster?
- You restrict to the requests on days t and $t + 1$, what happens between $t+2$ and $t+N$?
 - ▶ Do humans keep submitting requests?
 - ▶ Do Cloud machines' requests drop after $t + 1$? This would be consistent with your proposed mechanism

Comment 5:...And context

- Humans might have an advantage in that they are able to better **contextualize the info**
- Hou and Moskowitz (2005) develop an efficiency measure based on
- Delay with respect to price responds to market info
- Run $r_{i,t} = \alpha_i + \beta_i r_{m,t} + \sum_{n=1}^N \delta_{i,n} r_{m,t-n} + \epsilon_{i,t}$
 - ▶ Restricted ($\delta = 0$)
 - ▶ Unrestricted
- And compute $D_i = 1 - \frac{R^2_{[\delta_{i,n}=0, \forall n]}}{R^2}$
- Higher $D_i \rightarrow$ slower incorporation of market information \rightarrow higher inefficiency
- **This measure might capture inefficiency wrt to market information**
 - ▶ Machines might not improve this kind of efficiency if they look at 8-K in isolation

Minor Comments

- 1 Restricting the sample to filings with the same filing and event date does definitely make sense. On the other hand it might
 - ▶ Undersample bad news (literature on withholding negative news)
 - ▶ Undersample some types of items (see Ben-Rephael et. al. 2022)
- 2 Provide some info on the Supply of 8-K. Is the need of machines driven by an increase in supply?
- 3 Also try Quarterly FE rather than Annual FE
- 4 Results in Table 3 show that humans (machines) tend to pay more attention to Item 2.02 (Item 8) while results in Table 11 indicate that machines have an advantage in processing numerical information. Seems a bit of a contradiction.
- 5 Many papers that use Edgar data. Do better justice to the literature
 - ▶ Gibbons, Iliev, Kalodimos (2021); Iliev Kalodimos and Lowry (2020); Crane, Crotty Umar (2019); Li and Sun (2018); Cao, Kilic and Wang (2020) and several others

Conclusions

- 1 Very important, interesting, and carefully executed paper
- 2 Looking forward to seeing it in a top Journal!