Discussion of

Informed Trading Intensity

By Bogousslavsky, Fos, and Muravyev

@Cavalcade 2022

Discussant: “J” Jiacui Li

David Eccles, U of Utah

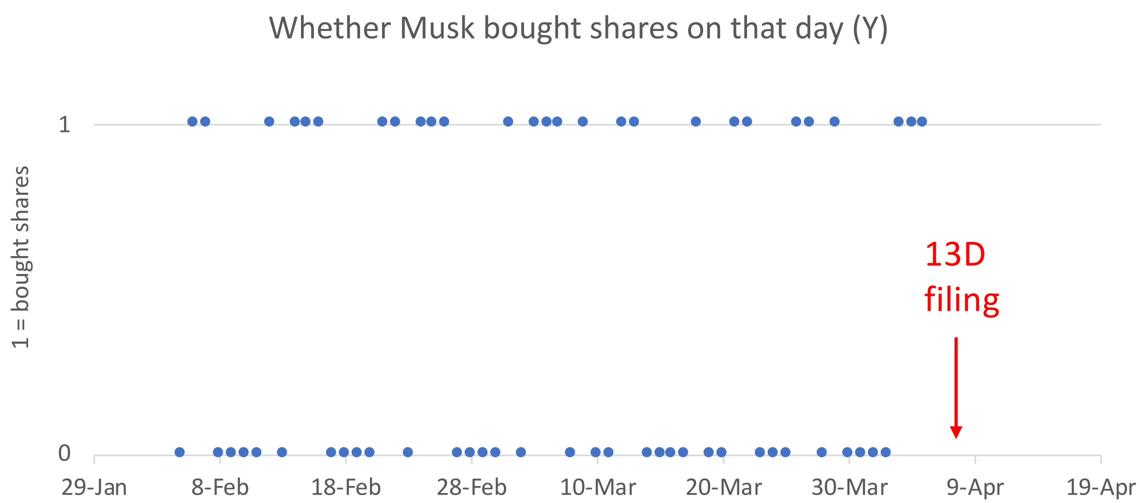
What does this paper do?

- Builds a machine learning algorithm to detect informed trading (IT)
  - Trained on "ground truth" events: 13D/insider trading/short selling

- Paper shows that $\widehat{IT}$ outperforms existing measures
  - And then applies $\widehat{IT}$ to event studies and understanding asset prices

How is the algorithm trained?

- Consider Elon Musk's 13D filing on Twitter.



- Train an algorithm: $Y_{i,t} = f(X_{i,t}, \theta)$
  - where $Y_{i,t} \in \{0,1\}$, $X_{i,t}$ includes many price/microstructure variables

Overall

- Good idea!
  - Also solid empirical execution. Results are sensible and interpretable

- My comments will focus on:
  - Conceptual interpretation
  - How well do we expect the algorithm to work out of sample

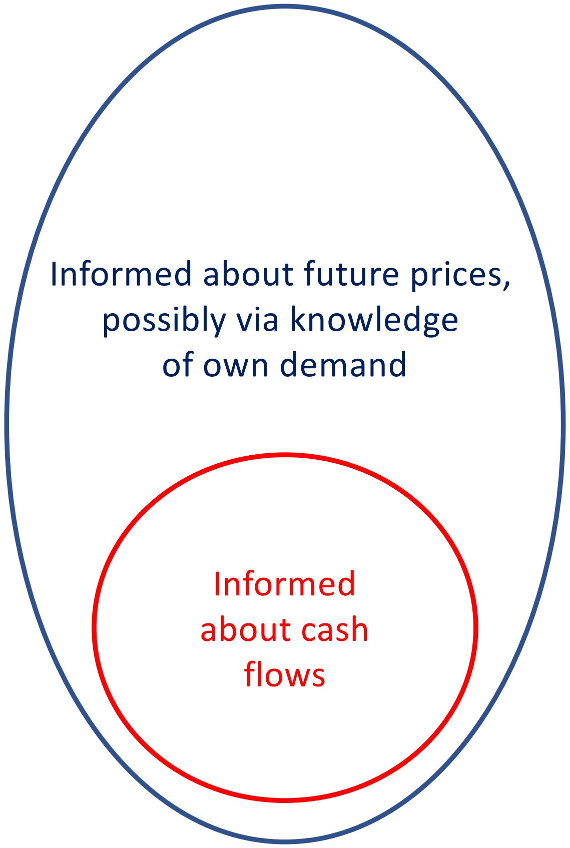
Comments

1. What is informed trading?

- Narrow definition: informed about *cash flows*

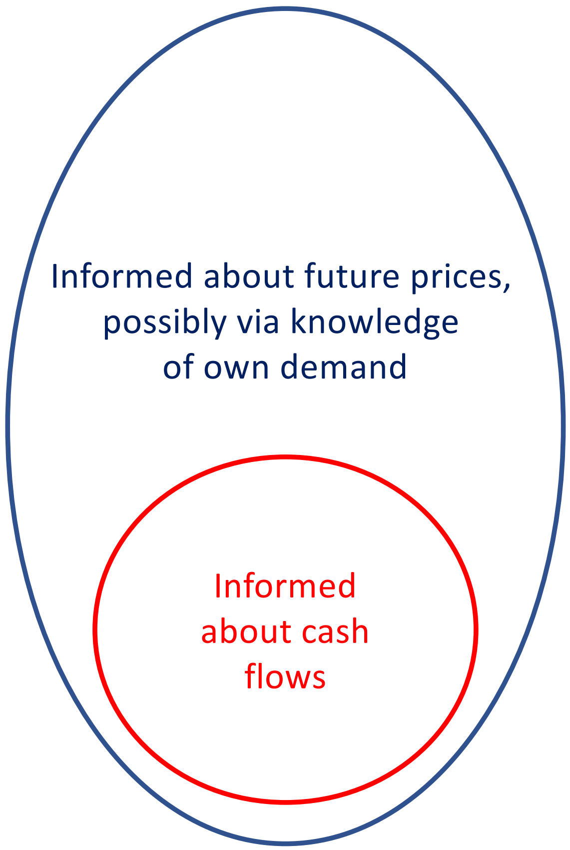
- Expansive definition: informed about *future prices*

- We use to think they are the same definition…
  - … because we assume that trades unrelated to cash flows have no price impact
  - Recent evidence does not support this view
    - See Koijen-Gabaix (2022) for a review



What are this and previous papers measuring?

- Order flow imbalances DO create somewhat persistent price impact
  - However, most imbalances are not cash flow-relevant (e.g. Li and Lin (2022))
- Recent papers show that "standard measures of IT don't work"
  - E.g. Collin-Dufresne and Fos (2015), Ahern (2020), this paper
  - *Perhaps they are measuring different types of IT?*
- One interpretation: this paper is creating a measure exclusively for the more restrictive definition of "cash flow IT"



2. How well would the algorithm capture other types of IT?

- Even with the narrower definition, there are certainly more types of IT beyond the 3 types considered (13D/insider trading/short selling)

- How well should we expect the algorithm to work for those types?

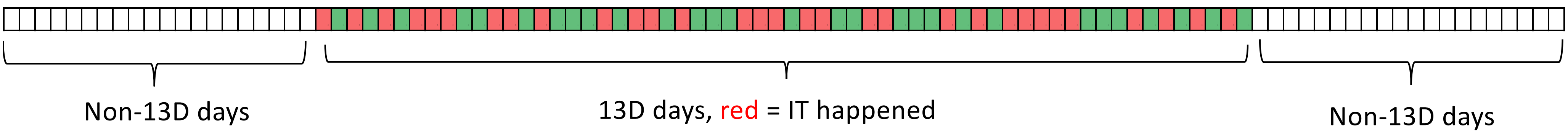
There is cross-detectability…

- … that said, the main detectability is still within the same type
  - This might reflect *fundamental differences* between the IT types

Dep. variable:	Schedule 13D trade		Insider trade		Δ Short interest	
	(1)	(2)	(3)	(4)	(5)	(6)
ITI(13D)	0.672***	0.585***	0.065***	0.060***	0.086***	0.072***
	(38.341)	(33.560)	(17.029)	(15.171)	(11.909)	(9.634)
ITI(insider)	0.049***	0.041***	0.366***	0.338***	0.026***	0.026***
	(4.017)	(3.331)	(49.036)	(44.982)	(3.715)	(3.621)
ITI(short)	0.180***	0.150***	0.185***	0.135***	0.571***	0.448***
	(4.768)	(4.034)	(21.094)	(14.491)	(34.487)	(24.408)
Controls	No	Yes	No	Yes	No	Yes
Stock/Filing FE	Yes	Yes	Yes	Yes	Yes	Yes
R^2	0.0989	0.1075	0.0079	0.0085	0.0099	0.0110
Obs.	56,979	56,979	766,156	766,156	204,237	204,237

3. Interpreting the R2 measures

- Paper shows that $\widehat{IT_{13D}}$ explains IT_{13D} with $R^2 \approx 10\%$
- What is the sample selection?



- There are two steps in predicting IT:
  - 1) Is this day part of a 13D period?
  - 2) If so, did IT happen?
- The R2 in the paper applies to the *second step*

Summary

- Conceptual: I suspect this paper creates a measure of "cash flow-IT"
  - This is a subset of expansive definition of IT (informed trading)

- Empirical:
  - There will be challenges when applying the algorithm out of sample
  - The fair comparison is not 100% R^2, but existing measures which don't work at all… at least not in detecting cash flow-IT

- I enjoyed reading the paper and look forward to the next version!