

“Trade, Quotes, and Information Shares”
by
Björn Hagströmer and Albert Menkveld

Discussant:
Pierre Collin-Dufresne
SFI@EPFL and CEPR

June 2023

- Background
 - Contribution
 - Comments
 - Conclusion
-

Fragmented Markets and Price Discovery?

- ▶ The same stock can be traded on different venues:
 - ▶ 'Lit' exchanges,
 - ▶ 'Dark' trading venues (dark pools, single dealer platforms).
- ▶ Impact of market fragmentation on price discovery?
 - ▶ Uninformed have incentive to seek off-exchange venues to separate from informed order flow (Seppi (1990))
 - ▶ Informed order flow prefers certainty of execution against limit orders to uncertain execution in Dark Pools (Zhu (2014))
 - ▶ This could stifle information production
 - ▶ Welfare implication are not straightforward however (Benveniste, Markus, Wilhelm (1992), Lee and Wang (2016))
- ▶ Empirically, can we measure how much price-discovery occurs on different ('lit' vs. 'dark') venues?

How to measure the information content of trades for one market?

- ▶ Hasbrouck (1991) estimates VAR in 'transaction time' (quote revision or trade) for
 - ▶ changes in mid-quote $m_t = \frac{1}{2}(q_t^a + q_t^b)$, and
 - ▶ cumulative signed order flow between transactions x_t

$$\begin{aligned}\Delta m_t &= \sum_{k=1}^K a_k \Delta m_{t-k} + \sum_{k=0}^K b_k x_{t-k} + v_{1,t} \\ x_t &= \sum_{k=1}^K c_k \Delta m_{t-k} + \sum_{k=1}^K d_k x_{t-k} + v_{2,t}\end{aligned}$$

- ▶ The VAR can be written (e.g., by backward iteration) as a Vector Moving Average:

$$\Delta m_t = v_{1,t} + \sum_{k=1}^{\infty} a_k^* v_{1,t-k} + \sum_{k=0}^{\infty} b_k^* v_{2,t-k}$$

- ▶ Assume $v_{1,t} = v_{2,t} = 0 \forall t > 0$, then **impact on future quote revisions** of a trade shock $v_{2,0} \neq 0$ can be inferred from VMA
($\Delta m_0 = b_0^* v_{2,0}$, $\Delta m_1 = b_1^* v_{2,0}$, $\Delta m_2 = b_2^* v_{2,0}$, ...)

- ▶ Thus the **cumulative impact** of **trade shock** $v_{2,0}$ is: $v_{2,0} \sum_{k=0}^{\infty} b_k^*$.
- ▶ Impact of **public non-trade information shock** $v_{1,0}$ is: $v_{1,0}(1 + \sum_{k=1}^{\infty} a_k^*)$.

⇒ the contribution of trades to the total variance of quote revisions is:

$$\frac{(\sum_{k=0}^{\infty} b_k^*)^2 \sigma_2^2}{(\sum_{k=0}^{\infty} b_k^*)^2 \sigma_2^2 + (1 + \sum_{k=1}^{\infty} a_k^*)^2 \sigma_1^2}$$

How to measure the information share of trades for two markets

- ▶ Hasbrouck (1995) proposes a **Vector Error Correction Model** to account for the cointegration in prices across markets.

- ▶ For example, for 2 markets L and D , define $\mathbf{z}_t = (\mathbf{m}_t^D - \mathbf{m}_t^L)$:

$$\Delta \mathbf{m}_t^L = \alpha_L \mathbf{z}_t + \sum_{k=1}^K a_k^{LL} \Delta \mathbf{m}_{t-k}^L + \sum_{k=1}^K a_k^{LD} \Delta \mathbf{m}_{t-k}^D + \sum_{k=0}^K b_k^{LL} x_{t-k}^L + \sum_{k=0}^K b_k^{LD} x_{t-k}^D + v_{1,t}^L$$

$$\Delta \mathbf{m}_t^D = \alpha_D \mathbf{z}_t + \sum_{k=0}^K a_k^{DL} \Delta \mathbf{m}_{t-k}^L + \sum_{k=1}^K a_k^{DD} \Delta \mathbf{m}_{t-k}^D + \sum_{k=0}^K b_k^{DL} x_{t-k}^L + \sum_{k=0}^K b_k^{DD} x_{t-k}^D + v_{1,t}^D$$

$$x_t^L = \beta_L \mathbf{z}_t + \sum_{k=1}^K c_k^{LL} \Delta \mathbf{m}_{t-k}^L + \sum_{k=1}^K c_k^{LD} \Delta \mathbf{m}_{t-k}^D + \sum_{k=1}^K d_k^{LL} x_{t-k}^L + \sum_{k=1}^K d_k^{LD} x_{t-k}^D + v_{2,t}^L$$

$$x_t^D = \beta_D \mathbf{z}_t + \sum_{k=1}^K c_k^{DL} \Delta \mathbf{m}_{t-k}^L + \sum_{k=1}^K c_k^{DD} \Delta \mathbf{m}_{t-k}^D + \sum_{k=1}^K d_k^{DL} x_{t-k}^L + \sum_{k=1}^K d_k^{DD} x_{t-k}^D + v_{2,t}^D$$

- ▶ The VECM can be estimated via OLS.
- ▶ The VECM can also be written as a VMA to infer the relative contributions (or upper and lower bounds thereof depending on the correlation between $v_{1,t}^j, v_{2,t}^k$) on the common 'unobservable' random walk component of the market mid-quotes of a **L-trade ($v_{2,0}^L$) or D-trade ($v_{2,0}^D$) shock** or a **non-trade information shock ($v_{1,0}^L, v_{1,0}^D$)**.

→ Using such a VECM, e.g., C-D, Junge, Trolle (2020) study the information shares of quotes vs. trades for DTD vs. DTC trading platforms for CDX.IG swap spreads (section IV and footnote 39).

- ▶ Note that some venues (e.g., DP) may not report quotes continuously, so quote feed revisions may be missing (e.g., \mathbf{m}_t^D).

This paper

- ▶ Applies this approach to 10 large-cap stocks with primary listing at the LSE (and no dual-listings outside of UK).
- ▶ Use price quotes and trades for the period April 1 to June 30 2021 from
 - ▶ LSE and
 - ▶ four largest Multilateral Trading Facilities (Aquis, CBOE BXE, CBOE CXE, LSEG Turquoise).
- ▶ Source trades also from
 - ▶ Dark pools
 - ▶ Periodic Auctions,
 - ▶ Single Dealer platforms.
- ▶ One millisecond sample frequency (> 30 million observation per stock).
- ▶ Use 100 Lags in VECM and distributed lag model to reduce number of parameters (Hasbrouck (2003,2021)).
- ▶ With 2 mid-quote series and 5 signed order flow series from 5 different venues *“it takes a full week to compute the information shares relying on parallel processing and state of the art cloud computing services”!*

Main results

- **Information shares from Lit trades and quotes dominates significantly that of Dark trades** (even when focusing on upper and lower bounds due to different decompositions of the trade and non-trade shocks correlation matrix).

	LSE quotes				LSE trades				DP trades			
	IS	GIS	Diff.		IS	GIS	Diff.		IS	GIS	Diff.	
	(%)	(%)	(%)	<i>t</i> -stat.	(%)	(%)	(%)	<i>t</i> -stat.	(%)	(%)	(%)	<i>t</i> -stat.
LBound	63.76	56.46	-7.30	-10.2**	29.24	35.52	6.27	10.7**	0.89	0.03	-0.86	-6.6**
UBound	69.63	64.33	-5.31	-9.1**	35.12	43.38	8.26	11.4**	1.23	0.21	-1.01	-6.5**

Table 3: Information Share Estimates for LSE Stocks. This table presents information share estimates of quotes and trades for a sample of LSE stocks from April 1 through June 30, 2021. The LSE midquote is used for the price quote series. Trades include LSE trades and dark pool trades. The H95 information shares (IS) estimates are based on forward-filled trade prices, while generalized information shares (GIS) use signed volume. Panel (a) presents overall averages of the information-share lower and upper bounds. Panel (b) sorts stock-days

- Perform a convincing simulation experiment that the approach that relies on the combination of signed order flow and quote revisions is superior (i.e., **unbiased**) to the approach that uses a single forward-filled (i.e., stale) price series for each venue.

'GIS' copyright?

Analytical properties of Hasbrouck and generalized information shares



Donald Lien^{a,*}, Keshab Shrestha^b, Lianne Mei Quin Lee^c

^a *The University of Texas at San Antonio, United States of America*

^b *Monash University Malaysia, Malaysia*

^c *Sunway University Malaysia, Malaysia*

ARTICLE INFO

Keywords:

Price discovery

Hasbrouck information share

Generalized information share

ABSTRACT

This paper discusses analytical properties of Hasbrouck information share (HIS) and generalized information share (GIS). We reject the conventional designation of lower and upper bounds for HIS. Moreover, we show the difference between the GIS across the two markets is larger than the corresponding difference between the average HIS. Thus, the dominance-satellite relationship is more prominent when using GIS to measure the role of price discovery. Using the data of dual-class shares of Berkshire Hathaway Inc., we find approximately 74.34% of the price discovery occurs in the B-shares market, which is consistent with the *stealth trading hypothesis*.

Comments

- ▶ Bias in measure IS when using (stale) forward filling prices is convincingly shown by simulations. Why do researchers use this approach? Hasbrouck (1995 p. 1888):

In computing the information share measures suggested in this paper, modeling centers on the bid and offer quotes. An analysis of an individual stock trading in multiple markets, if based on last sale prices, would labor under problems of autocorrelation induced by infrequent trading. In particular, a market that happened to have relatively infrequent trades would tend to have last-sale prices that were most obsolete, and therefore least informative. Quotes, on the other hand, can be updated in the absence of trades. For this reason, the present estimations are based on quotes rather than transaction prices.

As noted earlier, the quote model may be viewed as the reduced form of a richer model that involves transaction prices and signed trade volumes. These

- ▶ Information in signed order flow: Is zero trade the same as large positive and negative off-setting trades for measuring IS? (Bogousslavsky-CD (2023) suggest it may not be, but perhaps not at super high frequency...).
- ▶ Transaction time?

Conclusion

- ▶ Very impressive implementation: **computational tour de force** and convincing results.
- ▶ Findings suggest that Information shares of Lit trades and quotes dominates significantly that of Dark venues.
- ▶ This is consistent with venue selection theories (Seppi (1990), Zhu (2014), Lee and Wang (2016)).
- ▶ Implications for the welfare of market fragmentation are not obvious:
 - ▶ Uninformed versus informed agents' trade surplus,
 - ▶ Price efficiency,
 - ▶ Information production.
- ▶ When working at very high frequency, what type of information is the IS analysis capturing?
 - ▶ Fundamental vs. technical?
 - ▶ Long-term vs short-term?
 - ▶ Are the IS shares same across different types of securities (size, volatility...).