Identifying HFT Activity without Proprietary Data

Bidisha Chakrabarty
(Saint Louis University, USA)

Carole Comerton-Forde
(Univ of Melbourne, Aust.)

Roberto Pascual
(University of the Balearic Islands, Spain)

NSE NYU Conference on Indian Capital Markets
December, 2021
Identifying HFT without identifiers in the data
Identification strategies

- **Proprietary databases**  
  (e.g., Brogaard, et al., 2014; 2017; Kirilenko et al., 2017; Comerton-Forde et al., 2018; Boehmer et al., 2019)  
  - Availability; Replicability, Coverage; Based on proxies (e.g., EUROFIDAI, IIROC)

- **Latency-changing exogenous events**  
  (e.g., Hendershott et al., 2011; Riordan and Storkenmaier, 2012; Boehmer et al., 2020; Shkilko and Sokolov, 2020)  
  - Unable to directly identify HFT activity; combined with proxies

- **HFT proxies**
Proxy Metrics

- Message Traffic (Mess)
- Cancellations (Can)
- Monitoring Intensity (MonInt)
- Fleeting Orders (FleetOrd)
- Quotation intensity (QuoteInt)
- Quote flickering (Flick)
- Speed of response (SResp)
- Strategic runs (SRuns)
Our study: research questions

- Comprehensive examination of the 8 most popular HFT proxies.

- **RQ#1**: Reliability

- **RQ#2**: Type of HFT activities

- **RQ#3**: Identify HFTs but not other traders
Precise HFT identifiers – Our Data

- **National Stock Exchange (NSE) of India**: 4th (10th) largest Exchange by #trades (dollar volume)

- Ranked just below the TSX in market cap;

- TSX 2.6 trillion USD  NSE USD 2.55 trillion USD

In 2018, AT share of volume is almost 50%
## Trader types

- Two internal flags:
  - "Client account"
  - "Order entry mode"

<table>
<thead>
<tr>
<th></th>
<th>Proprietary</th>
<th>Agency</th>
</tr>
</thead>
<tbody>
<tr>
<td>Algorithmic trader (AT)</td>
<td>High-frequency traders <em>(HFTs)</em></td>
<td>Agency Algorithmic Traders <em>(AATs)</em></td>
</tr>
<tr>
<td>Non-AT</td>
<td>Non-algorithmic traders <em>(NATs)</em></td>
<td></td>
</tr>
</tbody>
</table>

- Algorithmic order entry for prop trading = SEC definition of HFT
We consider 8 metrics used to proxy for HFT activity,

- Two versions:
  - “True” HFT metrics, using the HFT messages only
  - “Proxy” HFT metrics, using all messages (ignoring trader indification)
RQ1: How reliable are the proxies?  
Correlations between true & proxy metrics

<table>
<thead>
<tr>
<th>Interval</th>
<th>Mess</th>
<th>Can</th>
<th>MonInt</th>
<th>FleetOrd</th>
<th>QuoteInt</th>
<th>Flick</th>
<th>SResp</th>
<th>SRuns</th>
</tr>
</thead>
<tbody>
<tr>
<td>30s</td>
<td>96.77***</td>
<td>78.21***</td>
<td>97.13***</td>
<td>92.28***</td>
<td>88.93***</td>
<td>67.94***</td>
<td>93.81***</td>
<td>98.02***</td>
</tr>
<tr>
<td>60s</td>
<td>96.92***</td>
<td>79.56***</td>
<td>97.40***</td>
<td>92.97***</td>
<td>89.72***</td>
<td>71.53***</td>
<td>94.84***</td>
<td>97.83***</td>
</tr>
<tr>
<td>300s</td>
<td>96.93***</td>
<td>80.39***</td>
<td>97.67***</td>
<td>93.62***</td>
<td>90.33***</td>
<td>77.70***</td>
<td>95.72***</td>
<td>98.00***</td>
</tr>
<tr>
<td>900s</td>
<td>96.72***</td>
<td>80.02***</td>
<td>97.74***</td>
<td>93.49***</td>
<td>89.71***</td>
<td>78.69***</td>
<td>95.69***</td>
<td>98.30***</td>
</tr>
<tr>
<td>1500s</td>
<td>96.60***</td>
<td>79.56***</td>
<td>97.72***</td>
<td>93.43***</td>
<td>89.36***</td>
<td>78.75***</td>
<td>95.50***</td>
<td>98.45***</td>
</tr>
</tbody>
</table>
HFTs’ contribution \((HFTCont)\) to liquidity supply/demand

- **Liquidity supply:**
  - **Trade-based metric:** % trades in which HFTs are on the passive side
  - **LOB-based metrics:** best quotes, depth, top 5 levels of LOB

- **Liquidity demand:**
  - **Trade-based contribution metric:** % trades initiated by HFTs

\[
HFTCont_{i,t} = \alpha + \beta HFTPProxy_{i,t} + \gamma HFTCont_{i,t-1} + \lambda_0 \text{Open}_{i,t} + \lambda_c \text{Close}_{i,t} + \delta_i + \varepsilon_{i,t}
\]

**Results:** At all levels of aggregation \(\beta\) is positive and highly significant
High vs low HFT liquidity demand/supply

- Unusually high (low): >75 (<25) percentile of the corresponding indicator:
  - (Dhigh, Shigh): High demand & High supply
  - (Dhigh, Slow): High demand & Low supply
  - (Dlow, Shigh): Low demand & High supply
  - (Dlow, Slow): Low demand & Low supply

\[ HFTMetric_{i,t} = \alpha + \beta_{LL} DlowSlow_{i,t} + \beta_{LH} DlowShigh_{i,t} + \beta_{HL} DhighSlow_{i,t} + \beta_{HH} DhighShigh_{i,t} + \lambda_{o} Open_{i,t} + \lambda_{c} Close_{i,t} + \delta_{i} + \varepsilon_{i,t} \]
High vs low HFT liquidity demand/supply
(proxy metrics; supply = % time at the best quotes)

<table>
<thead>
<tr>
<th></th>
<th>Mess</th>
<th>Can</th>
<th>MonInt</th>
<th>FleetOrd</th>
<th>QuotelInt</th>
<th>Flick</th>
<th>SResp</th>
<th>SRuns</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Dlow, Slow)</td>
<td>-245.98***</td>
<td>-10.55***</td>
<td>-98.83***</td>
<td>-29.46***</td>
<td>-83.43***</td>
<td>-0.18***</td>
<td>-22.10***</td>
<td>-2.67***</td>
</tr>
<tr>
<td>(Dhigh, Slow)</td>
<td>-113.00***</td>
<td>-5.74***</td>
<td>-43.60***</td>
<td>-11.56***</td>
<td>-44.15***</td>
<td>-0.07***</td>
<td>-7.26***</td>
<td>-1.32***</td>
</tr>
<tr>
<td>(Dlow, Shigh)</td>
<td>-114.26***</td>
<td>-5.94***</td>
<td>-41.27***</td>
<td>-13.36***</td>
<td>-30.18***</td>
<td>-0.07***</td>
<td>-0.45***</td>
<td>-0.96***</td>
</tr>
<tr>
<td>(Dhigh, Shigh)</td>
<td>231.16***</td>
<td>9.69***</td>
<td>92.78***</td>
<td>29.98***</td>
<td>94.96***</td>
<td>0.20***</td>
<td>33.35***</td>
<td>2.58***</td>
</tr>
</tbody>
</table>
Can the proxies isolate HFTs from others?

- **1st stage**: for each stock $i$ and “true” metric

$$HFT_{true,i,t} = \alpha + \beta_{AAT}AAT_{true,i,t} + \beta_{NAT}NAT_{true,i,t} + HFT_{true,i,t}, \forall i = \{1, \ldots, 50\}$$

- **2nd stage**: Pooled regression

$$HFT_{true,i,t} = \alpha + \beta_{HFT}proxy_{i,t} + \lambda_{o}Open_{i,t} + \lambda_{c}Close_{i,t} + \delta_{i} + \varepsilon_{i,t}$$

- We repeat the process above for AATs and NATs true metrics
## Can the proxies isolate HFTs from others?

2nd stage estimates - 30s intervals

<table>
<thead>
<tr>
<th>Trader type</th>
<th>Statistic</th>
<th>Mess</th>
<th>Can</th>
<th>MonInt</th>
<th>FleetOrd</th>
<th>QuoteInt</th>
<th>Flick</th>
<th>SResp</th>
<th>SRuns</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Coef.x100</td>
<td>35.02</td>
<td>31.05</td>
<td>51.28</td>
<td>39.66</td>
<td>52.29</td>
<td>28.05</td>
<td>34.94</td>
<td>77.59</td>
</tr>
<tr>
<td>HFT</td>
<td>t-stat</td>
<td>(22.52)</td>
<td>(13.08)</td>
<td>(31.60)</td>
<td>(14.97)</td>
<td>(19.24)</td>
<td>(4.55)</td>
<td>(9.90)</td>
<td>(40.97)</td>
</tr>
<tr>
<td></td>
<td>R^2 (2nd stage)</td>
<td>0.35</td>
<td>0.31</td>
<td>0.51</td>
<td>0.40</td>
<td>0.52</td>
<td>0.05</td>
<td>0.35</td>
<td>0.78</td>
</tr>
<tr>
<td>AAT</td>
<td>Coef.x100</td>
<td>2.42</td>
<td>33.93</td>
<td>2.25</td>
<td>6.20</td>
<td>7.54</td>
<td>22.61</td>
<td>4.69</td>
<td>0.89</td>
</tr>
<tr>
<td>t-stat</td>
<td>(11.66)</td>
<td>(10.24)</td>
<td>(10.23)</td>
<td>(12.20)</td>
<td>(7.93)</td>
<td>(14.83)</td>
<td>(5.21)</td>
<td>(7.15)</td>
<td>(7.15)</td>
</tr>
<tr>
<td></td>
<td>R^2 (2nd stage)</td>
<td>0.04</td>
<td>0.34</td>
<td>0.03</td>
<td>0.08</td>
<td>0.08</td>
<td>0.12</td>
<td>0.05</td>
<td>0.01</td>
</tr>
<tr>
<td>NAT</td>
<td>Coef.x100</td>
<td>0.32</td>
<td>1.17</td>
<td>0.17</td>
<td>1.71</td>
<td>0.27</td>
<td>13.97</td>
<td>0.37</td>
<td>0.03</td>
</tr>
<tr>
<td>t-stat</td>
<td>(6.79)</td>
<td>(5.26)</td>
<td>(4.64)</td>
<td>(6.26)</td>
<td>(5.66)</td>
<td>(8.13)</td>
<td>(2.70)</td>
<td>(7.62)</td>
<td>(7.62)</td>
</tr>
<tr>
<td></td>
<td>R^2 (2nd stage)</td>
<td>0.02</td>
<td>0.01</td>
<td>0.00</td>
<td>0.03</td>
<td>0.02</td>
<td>0.12</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>Obs.</td>
<td>2419500</td>
<td>2419500</td>
<td>2419500</td>
<td>2413936</td>
<td>2400679</td>
<td>2419500</td>
<td>2231961</td>
<td>2419500</td>
<td></td>
</tr>
</tbody>
</table>
Conclusions

Proxies of HFT activity:

a) Perform well in identifying HFT activity
b) Are highly correlated with each other
c) Are good at identifying HFT liquidity demand as well as supply, but cannot differentiate them
d) Their performance is not dependent on the level of time aggregation
e) Hasbrouck and Saar’s (2013) strategic runs outperform other proxies in capturing HFT-specific activity
Thank you!

Comments & suggestions welcome:
- bidisha.chakrabarty@slu.edu
- carole.comerton-forde@unimelb.edu.au
- rpascual@uib.es