High-Frequency Quoting: Measurement, Detection and Interpretation

Joel Hasbrouck
Outline

- Background
- Look at a data fragment
- Economic significance
- Statistical modeling
- Application to larger sample
- Open questions
Economics of high-frequency trading

- **Absolute speed**
  - In principle, faster trading leads to smaller portfolio adjustment cost and better hedging
  - For most traders, latencies are inconsequential relative to the speeds of macroeconomic processes and intensities of fundamental information.

- **Relative speed (compared to other traders)**
  - A first mover advantage is extremely valuable.
  - Low latency technology has increasing returns to scale and scope.
  - This gives rise to large firms that specialize in high-frequency trading.
Welfare: “HFT imposes costs on other players”

- They increase adverse selection costs.
- The information produced by HFT technology is simply advance knowledge of other players’ order flows.
  - Biais, Bruno, Thierry Foucault, and Sophie Moinas, 2012
Welfare: “HFT improves market quality.”

- Supported by most empirical studies that correlate HF measures/proxies with standard liquidity measures.
  - Hendershott, Terrence, Charles M. Jones, and Albert J. Menkveld, 2010
  - Hasbrouck, Joel, and Gideon Saar, 2011
  - Hendershott, Terrence J., and Ryan Riordan, 2012
“HFTs are efficient market-makers”

- Empirical studies
  - Menkveld, Albert J., 2012
  - Brogaard, Jonathan, 2010a, 2010b, 2012

- Strategy: identify a class of HFTs and analyze their trades.
- HFTs closely monitor and manage their positions.
- HFTs often trade passively (supply liquidity via bid and offer quotes)

- But ...
  - HFTs don’t maintain a continuous market presence.
  - They sometimes trade actively ("aggressively")
Positioning

- We use the term “high frequency trading” to refer to all sorts of rapid-paced market activity.
- Most empirical analysis focuses on trades.
- This study emphasizes quotes.
High-frequency quoting

- Rapid oscillations of bid and/or ask quotes.
- Example
  - AEPI is a small Nasdaq-listed manufacturing firm.
  - Market activity on April 29, 2011
  - National Best Bid and Offer (NBBO)
    - The highest bid and lowest offer (over all market centers)
National Best Bid and Offer for AEPI during regular trading hours
Caveats

- Ye & O’Hara (2011)
  - A bid or offer is not incorporated into the NBBO unless it is 100 sh or larger.
  - Trades are not reported if they are smaller than 100 sh.

- Due to random latencies, agents may perceive NBBO’s that differ from the “official” one.

- Now zoom in on one hour ...
National Best Bid and Offer for AEPI from 11:00 to 12:10
National Best Bid and Offer for AEPI from 11:15:00 to 11:16:00
National Best Bid and Offer for AEPI from 11:15:00 to 11:16:00
National Best Bid for AEPI:
11:15:21.400 to 11:15:21.800 (400 ms)
So what?

- HFQ noise degrades the informational value of the bid and ask.
- HFQ aggravates execution price uncertainty for marketable orders.
- And in US equity markets ...
  - NBBO used as reference prices for dark trades.
  - Top (and only the top) of a market’s book is protected against trade-throughs.
“Dark” Trades

- Trades that don’t execute against a visible quote.
- In many trades, price is assigned by reference to the NBBO.
  - Preferred orders are sent to wholesalers.
    - Buys filled at NBO; sells at NBB.
  - Crossing networks match buyers and sellers at the midpoint of the NBBO.
Features of the AEPI episodes

- Extremely rapid oscillations in the bid.
- Start and stop abruptly
- Possibly unconnected with fundamental news.
- Directional (activity on the ask side is much smaller)
A framework for analysis: the requirements

- Need precise resolution (the data have one ms. time-stamps)
- Low-order vector autoregression?
- Oscillations: spectral (frequency) analysis?
  - Represent a time series as a combination sine/cosine functions.
    - But the functions are recurrent over the full sample.
- AEPI episodes are localized.
Stationarity

- The oscillations are locally stationary.
- Framework must pick up stationary local variation ...
  - But not exclude random-walk components.
- Should identify long-run components as well as short-run.
Intuitively, I’d like to ...

- Use a moving average to smooth series.
  - Implicitly estimating the long-term component.
- Isolate the HF component as a residual.
Represent a time-series in terms of basis functions (wavelets)

Wavelets:
- Localized
- Oscillatory
- Use flexible (systematically varying) time-scales.

Accepted analytical tool in diverse disciplines.
- Percival and Walden; Gencay et. al.
Sample bid path
First pass (level) transform

- Original price series
- 2-period average
- 1-period detail
First pass (level) transform

Original price series

2-period average

1-period detail

1-period detail sum of squares
Second pass (level) transform

- 2-period detail
- 4-period average
- 2-period average
Third level (pass) transform

4-period average
8-period average
4-period detail
For each level $j = 1, 2, \ldots$, we have ...

- A time scale, $\tau_j = 2^{j-1}$
  - Higher level $\rightarrow$ longer time scale.
  - $\tau_j \in \{1, 2, 4, \ldots\}$
  - “the persistence of the level-\(j\) component”

- A scale-$\tau_j$ “detail” component.
  - Centered (“zero mean”) series that tracks changes in the series at scale $\tau_j$.

- A scale-$\tau_j$ sum of squares.
The full set of scale-\(\tau_j\) components decomposes the original series into sequences ranging from “very rough” to “very smooth”.

- Multi-resolution analysis.

With additional structure, the full set of scale-\(\tau_j\) sums of squares corresponds to a variance decomposition.
Multi-resolution analysis of AEPI bid

- Data time-stamped to the millisecond.
- Construct decomposition through level $J = 18$.
- For graphic clarity, aggregate the components into four groups.
- Plots focus on 11am-12pm.
Connection to standard time series analysis

- Suppose $p_t$ is a stochastic process
  - e.g., a random-walk
- The scale-$\tau_j$ sum-of-squares over the sample path (divided by $n$) defines an estimate of the wavelet variance.
- Wavelet variance (and its estimate) are well-defined and well-behaved assuming that the first differences of $p_t$ are covariance stationary.
- Wavelet decompositions are performed on the levels of $p_t$ not the first differences.
The wavelet variance of a random-walk

- $\nu^2(\tau_j) \equiv$ wavelet variance at scale $\tau_j$
- For a random-walk
  - $p_t = p_{t-1} + e_t$
    where $Ee_t = 0$ and $Ee_t^2 = \sigma_e^2$
  - $\nu^2(\tau_j) = \phi(\tau_j)\sigma_e^2$
    where scaling factor $\phi(\tau_j) = \frac{1}{6}\left(\tau_j + \frac{1}{2\tau_j}\right)$
- $\phi(\tau_j) \in \{0.25, 0.38, 0.69, 1.3, 2.7, 5.3, 10.7, \ldots\}$
\[
\lim_{j \to \infty} \nu^2(\tau_j) = \infty
\]
The wavelet variance for the AEPI bid: an economic interpretation

- Orders sent to market are subject to random delays.
  - This leads to arrival uncertainty.
  - For a market order, this corresponds to price risk.
- For a given time window, the cumulative wavelet variance measures this risk.
Timing a trade: the price path
Timing a trade: the arrival window

![Graph showing price and time]
The time-weighted average price (TWAP) benchmark
Timing a trade: TWAP Risk

Variation about time-weighted average price
Price uncertainty

- Price uncertainty at time scale $\tau_j$ is measured by the wavelet variance at time scale $\tau_j$ and all smaller (finer) time scales.
  - If I don’t know which $\frac{1}{4}$-second interval will contain my execution, I don’t know which $\frac{1}{8}$- or $\frac{1}{16}$-second interval will contain my execution.

- The “jth level wavelet rough variance” is the cumulative wavelet variance at time scale $\tau_j$ and all smaller time scales.
The wavelet variance: a comparison with realized volatility
Data sample

- 100 US firms from April 2011
- Sample stratified by equity market capitalization
- Alphabetic sorting
- Within each market cap decile, use first ten firms.
- Summary data from CRSP
- HF data from daily ("millisecond") TAQ
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Full sample</td>
<td>$13.75</td>
<td>420</td>
<td>2,140</td>
<td>1,111</td>
<td>23,347</td>
</tr>
<tr>
<td>0 (low)</td>
<td>$4.18</td>
<td>30</td>
<td>20</td>
<td>17</td>
<td>846</td>
</tr>
<tr>
<td>1</td>
<td>$3.56</td>
<td>30</td>
<td>83</td>
<td>45</td>
<td>2,275</td>
</tr>
<tr>
<td>2</td>
<td>$3.70</td>
<td>72</td>
<td>228</td>
<td>154</td>
<td>5,309</td>
</tr>
<tr>
<td>3</td>
<td>$6.43</td>
<td>236</td>
<td>771</td>
<td>1,405</td>
<td>15,093</td>
</tr>
<tr>
<td>4</td>
<td>$7.79</td>
<td>299</td>
<td>1,534</td>
<td>468</td>
<td>15,433</td>
</tr>
<tr>
<td>5</td>
<td>$17.34</td>
<td>689</td>
<td>3,077</td>
<td>1,233</td>
<td>34,924</td>
</tr>
<tr>
<td>6</td>
<td>$26.34</td>
<td>1,339</td>
<td>5,601</td>
<td>2,045</td>
<td>37,549</td>
</tr>
<tr>
<td>7</td>
<td>$28.40</td>
<td>1,863</td>
<td>13,236</td>
<td>3,219</td>
<td>52,230</td>
</tr>
<tr>
<td>8</td>
<td>$36.73</td>
<td>3,462</td>
<td>34,119</td>
<td>7,243</td>
<td>94,842</td>
</tr>
<tr>
<td>9 (high)</td>
<td>$44.58</td>
<td>18,352</td>
<td>234,483</td>
<td>25,847</td>
<td>368,579</td>
</tr>
</tbody>
</table>
Computational procedures

- $10 \text{ hrs} \times 60 \text{ min} \times 60 \text{ sec} \times 1,000 \text{ ms} = 3.6 \times 10^7$ “observations” (per series, per day)
- Analyze data in rolling windows of ten minutes
- Supplement millisecond-resolution analysis with time-scale decomposition of prices averaged over one-second.
- Use maximal overlap discrete transforms with Daubechies(4) weights.
Example: AAPL (Apple Computer)

- 20 days
- Regular trading hours are 9:30 to 16:00.
  - I restrict to 9:45 to 15:45
  - 20 days × 6 hrs × 60 min = 7,200 min
- Compute $\hat{v}^2(\tau_j)$ for $j = 1, \ldots, 18$
- Time scales: 1ms to 131,072ms (about 2.2 minutes)
- Tables report values for odd $j$ (for brevity)
\[ \sqrt{\text{Wavelet variances}} \] for AAPL (units: $0.01/\text{sh})

<table>
<thead>
<tr>
<th>Time scale</th>
<th>Median</th>
<th>99\textsuperscript{th} percentile</th>
<th>Median</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \tau_j )</td>
<td>Bid</td>
<td>Ask</td>
<td>Bid</td>
</tr>
<tr>
<td>1 ms</td>
<td>0.024</td>
<td>0.024</td>
<td>0.076</td>
</tr>
<tr>
<td>4 ms</td>
<td>0.038</td>
<td>0.038</td>
<td>0.122</td>
</tr>
<tr>
<td>16 ms</td>
<td>0.071</td>
<td>0.070</td>
<td>0.200</td>
</tr>
<tr>
<td>64 ms</td>
<td>0.145</td>
<td>0.143</td>
<td>0.400</td>
</tr>
<tr>
<td>256 ms</td>
<td>0.303</td>
<td>0.298</td>
<td>0.831</td>
</tr>
<tr>
<td>1.0 sec</td>
<td>0.573</td>
<td>0.564</td>
<td>1.664</td>
</tr>
<tr>
<td>4.1 sec</td>
<td>1.119</td>
<td>1.109</td>
<td>3.649</td>
</tr>
<tr>
<td>16.4 sec</td>
<td>2.043</td>
<td>2.031</td>
<td>7.851</td>
</tr>
</tbody>
</table>
Measuring execution price uncertainty: \(\sqrt{\text{Wavlet rough (cumulative) variances}}\) for AAPL Bid (units: $0.01/sh)

<table>
<thead>
<tr>
<th>(\tau_j)</th>
<th>Median</th>
<th>99(^{th}) percentile</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 ms</td>
<td>0.024</td>
<td>0.076</td>
</tr>
<tr>
<td>4 ms</td>
<td>0.053</td>
<td>0.172</td>
</tr>
<tr>
<td>16 ms</td>
<td>0.103</td>
<td>0.302</td>
</tr>
<tr>
<td>64 ms</td>
<td>0.205</td>
<td>0.571</td>
</tr>
<tr>
<td>256 ms</td>
<td>0.423</td>
<td>1.153</td>
</tr>
<tr>
<td>1.0 sec</td>
<td>0.828</td>
<td>2.337</td>
</tr>
<tr>
<td>4.1 sec</td>
<td>1.621</td>
<td>4.857</td>
</tr>
<tr>
<td>16.4 sec</td>
<td>3.167</td>
<td>10.331</td>
</tr>
</tbody>
</table>
**Cumulative wavelet variances**
for AAPL Bid (units: $0.01/sh)

The price uncertainty for a trader who can only time his marketable trades within a 4-second window has \( \sigma = \$0.016 \)

<table>
<thead>
<tr>
<th>Time</th>
<th>Variance</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 ms</td>
<td>0.052</td>
</tr>
<tr>
<td>4 ms</td>
<td>0.053</td>
</tr>
<tr>
<td>16 ms</td>
<td>0.103</td>
</tr>
<tr>
<td>64 ms</td>
<td>0.205</td>
</tr>
<tr>
<td>256 ms</td>
<td>0.423</td>
</tr>
<tr>
<td>1.0 sec</td>
<td>0.828</td>
</tr>
<tr>
<td>4.1 sec</td>
<td>1.621</td>
</tr>
<tr>
<td>16.4 sec</td>
<td>3.167</td>
</tr>
</tbody>
</table>

Compare: current access fees \( \approx \$0.003 \)
The wavelet correlation $\rho_{X,Y}(\tau_j)$

- For two series $X$ and $Y$, the wavelet variances are $\nu_X^2(\tau_j)$ and $\nu_Y^2(\tau_j)$
- The wavelet covariance is $\nu_{X,Y}(\tau_j)$
- The wavelet correlation is
  \[ \rho_{X,Y}(\tau_j) = \frac{\nu_{X,Y}(\tau_j)}{\sqrt{\nu_X^2(\tau_j)\nu_Y^2(\tau_j)}} \]
- Fundamental value changes should affect both the bid and the ask.
- The wavelet correlation at scale $\tau_j$ indicates the contribution of fundamental volatility.
- Next: wavelet correlation for AAPL bid and ask:
How closely do the wavelet variances for AAPL’s bid correspond to a random walk?

<table>
<thead>
<tr>
<th>Scale</th>
<th>Wavelet variance estimate</th>
<th>Random-walk variance factors</th>
<th>Implied random-walk variance</th>
</tr>
</thead>
<tbody>
<tr>
<td>𝜏_𝑗</td>
<td>( \hat{\nu}^2(\tau_j) )</td>
<td>( \phi(\tau_j) )</td>
<td>( \hat{\nu}^2(\tau_j) / \phi(\tau_j) )</td>
</tr>
<tr>
<td>1 ms</td>
<td>0.0009</td>
<td>0.1875</td>
<td>0.0050</td>
</tr>
<tr>
<td>4 ms</td>
<td>0.0024</td>
<td>0.4849</td>
<td>0.0049</td>
</tr>
<tr>
<td>16 ms</td>
<td>0.0075</td>
<td>1.8960</td>
<td>0.0039</td>
</tr>
<tr>
<td>64 ms</td>
<td>0.0309</td>
<td>7.5740</td>
<td>0.0041</td>
</tr>
<tr>
<td>256 ms</td>
<td>0.1360</td>
<td>30.2935</td>
<td>0.0045</td>
</tr>
<tr>
<td>1.0 sec</td>
<td>0.5140</td>
<td>121.1730</td>
<td>0.0042</td>
</tr>
<tr>
<td>4.1 sec</td>
<td>2.1434</td>
<td>484.6930</td>
<td>0.0044</td>
</tr>
<tr>
<td>16.4 sec</td>
<td>8.6867</td>
<td>1,938.7700</td>
<td>0.0045</td>
</tr>
</tbody>
</table>
If 0.005 (cents per share)^2 is the random-walk variance over one ms., the accumulated variance over a 6-hour mid-day period is:

0.005 \times 1,000 \times 3,600 \times 6 = 108,000

The implied 6-hour standard deviation is about 329 (cents per share).

AAPL’s average price in the sample is about $340

\frac{3.29}{340} \approx 1\%
Volatility Signature Plots

- Suggested by Andersen and Bollerslev.
- Plot realized volatility (per constant time unit) vs. length of interval used to compute the realized volatility.
- Basic idea works for wavelet variances.
- “How much is short-run quote volatility inflated, relative to what we’d expect from a random walk?”
Normalization of wavelet variances

- For a given stock, the implied random-walk variance at scale $\tau_j$ is $\hat{\nu}^2(\tau_j)/\phi(\tau_j)$.
- The longest time scale in the analysis is about 20 minutes.
- The ratio $\frac{\hat{\nu}^2(\tau_j)/\phi(\tau_j)}{\hat{\nu}^2(20 \text{ min})/\phi(20 \text{ min})}$ measures variance at scale $\tau_j$ relative to the wavelet variance at 20 minutes, under a random-walk benchmark.
- If the price is truly a random walk, this should be unity for all $\tau_j$. 
Market cap deciles collapsed into quintiles.

Within each quintile, I average
\[
\frac{\hat{\nu}^2(\tau_j) / \phi(\tau_j)}{\hat{\nu}^2(20 \text{ min}) / \phi(20 \text{ min})}
\]
across firms.

Results from millisecond- and second-resolution analyses are spliced.

Next: the (normalized) volatility signature plot.
The take-away

- For high-cap firms
  - Wavelet variances at short time scales have modest elevation relative to random-walk.

- Low-cap firms
  - Wavelet variances are strongly elevated at short time scales.
  - Significant price risk relative to TWAP.
Sample bid-ask wavelet correlations

- These are already normalized.
- Compute quintile averages across firms.
How closely do movements in the bid and ask track?

- Positive in all cases (!)
- For high-cap stocks, $\rho \approx 0.7$ (one second) and $\rho > 0.9$ (20 seconds)
- For bottom cap-quintile, $\rho < 0.2$ (one second) and $\rho < 0.5$ (20 minutes)
For each firm in mkt cap deciles ≤ 6, I examined the day with the highest wavelet variance at time scales of 1 second and under.

- HFQ is easiest to see against a backdrop of low activity.

Next slides ... some examples
Conclusions

- High frequency quoting is a real (but episodic) fact of the market.
- Time-scale decompositions are useful in measuring the overall effect.
  - ... and detecting the episodes
- Remaining questions ...

Why does HFQ occur?

- Why not? The costs are extremely low.
- Testing?
- Malfunction?
- Interaction of simple algos?
- Genuinely seeking liquidity (counterparty)?
- Deliberately introducing noise?
- Deliberately pushing the NBBO to obtain a favorable price in a dark trade?