Does Algorithmic Trading Improve Liquidity?

Terry Hendershott\textsuperscript{1}  Charles M. Jones\textsuperscript{2}  Albert J. Menkveld\textsuperscript{3}

\textsuperscript{1}Haas School of Business, UC Berkeley
\textsuperscript{2}Graduate School of Business, Columbia University
\textsuperscript{3}Tinbergen Institute, VU University Amsterdam

WFA, June 2008
Relative bid-ask spread Dow Jones stocks
(all stocks 1900-1928, DJIA stocks 1929-2000)
Time trend bid-ask spread (ctd)

NYSE value-weighted average effective spread

Figure 1. NYSE value-weighted proportional effective spreads
Institutional trading

How do institutions trade prior to algorithms? To buy 100,000 IBM shares, they

- hire a broker-dealer to take down or shop a block
- hire NYSE floor broker who uses judgement to slowly “work” the order

Broker-dealers now offer algos that minimize price concession through a dynamic trading strategy that optimizes over price, quantity, time, and venue

And, broker-dealers and hedge funds supply liquidity with algos (e.g. D.E. Shaw, Getco,...)
Related literature

IO of liquidity supply


Free trading option of limit orders (Copeland and Galai (1983))

▶ monitoring public information flow is costly (Foucault, Roëll, and Sandas (2003))

▶ AT may raise costs of non-AT limit orders (Rock (1990))

Optimal execution of large orders (Keim and Madhavan (1995), Bertsimas and Lo (1998), Almgren and Chriss (2000))

What do we do?

We measure algo trading through normalized (electronic) message traffic at the NYSE

- message traffic is electronic order submissions, cancels, and trade reports

Panel regressions associate time-series increases in algo trading with more liquid markets

- we exploit the exogenous, staggered introduction of autoquote at the NYSE as an instrument to establish causality
$messages_{it}$ (#electronic messages per minute i.e. proxy for algorithmic activity (/minute))

- **Q1**
- **Q2**
- **Q3**
- **Q4**
- **Q5** (95% conf. interval)
$algo_{trad}\_{it}$ (dollar volume per electronic message times $(-1)$ to proxy for algorithmic trading ($100$))
Autoquote

Decimals in 2001 shrink inside quote depth

In October 2002 NYSE proposes “liquidity quote”
  ▶ firm bid and offer for substantial size (> 15,000 shares)
“Autoquote” is proposed simultaneously to free up the specialist to concentrate on the liquidity quote
  ▶ specialists had been manually disseminating the inside quote
  ▶ software would now “autoquote” any change to book
Liquidity quote delayed, autoquote immediate

Autoquote is important for AT
  ▶ immediate feedback about terms of trade
    ▶ algo liquidity suppliers see abnormally wide inside quote
    ▶ algo liquidity demanders access quote more quickly
#stocks that trade in Autoquote

- Q1
- Q2
- Q3
- Q4
- Q5

↑ 5/27/03 last stocks to Autoquote
↓ 12/2/02 start of sample
↓ 1/29/03 first stocks to Autoquote
↑ 5/27/03 last stocks to Autoquote
↓ 1/29/03 first stocks to Autoquote
↓ 12/2/02 start of sample

7/31/03 end of sample ↓

0 20 40 60 80 100 120 140 160
25
50
75
100
125
150
175
200
↓ 12/2/02 start of sample
↓ 1/29/03 first stocks to Autoquote
↑ 5/27/03 last stocks to Autoquote
7/31/03 end of sample ↓
Autoquote dummy as instrument for \( algo\_{trad} \)  

Table 4: Overall, Between, and Within Correlations Autoquote Analysis

This table presents the overall, between, and within correlations for the variables used in the autoquote analysis. It is based on daily observations in the period when autoquote was phased in, i.e. December 2, 2003, through July 31, 2003. For variable definitions, we refer to Table 1. We exploit the exogenous autoquote dummy (0 before the autoquote introduction, 1 after) to instrument for \( algo\_{trad} \) in order to identify causality from \( algo\_{trad} \) to our liquidity measures. In the IV estimation, we exclude identification off of a time trend (by adding time dummies) and thus solely rely on the nonsynchronous introduction of autoquote (see Figure 5). Before we report the IV estimation results in subsequent tables, this table reports correlations between the instrument \( \text{autoquote} \) and the endogenous variable \( \text{algo}_{\text{trad}} \) after removing the time trend.

### Panel A: Overall, between, and within correlation after removing the time trend

<table>
<thead>
<tr>
<th></th>
<th>( \text{messagess}_{it} )</th>
<th>( \text{algo}_{trad} )</th>
<th>( \text{shareturnover}_{it} )</th>
<th>( \text{volatility}_{it} )</th>
<th>( 1/\text{price}_{it} )</th>
<th>( \ln \text{marketcap}_{it} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \text{autoquote}_{it} )</td>
<td>( \rho(\text{overall}) )</td>
<td>0.15*</td>
<td>-0.05*</td>
<td>0.02*</td>
<td>0.03*</td>
<td>0.02*</td>
</tr>
<tr>
<td></td>
<td>( \rho(\text{between}) )</td>
<td>0.23*</td>
<td>-0.16*</td>
<td>0.06</td>
<td>0.09*</td>
<td>0.04</td>
</tr>
<tr>
<td></td>
<td>( \rho(\text{within}) )</td>
<td>0.08*</td>
<td>0.03*</td>
<td>-0.01*</td>
<td>0.00</td>
<td>0.01*</td>
</tr>
</tbody>
</table>

### Panel B: Within correlation by quintile after removing the time trend

| \( \text{autoquote}_{it} \) Q1 | \( \rho(\text{within}) \) | 0.15* | 0.03* | 0.01* | -0.00 | 0.03* | -0.03* |
| \( \text{autoquote}_{it} \) Q2 | \( \rho(\text{within}) \) | 0.03* | 0.04* | -0.01* | 0.00  | -0.02* | 0.01* |
| \( \text{autoquote}_{it} \) Q3 | \( \rho(\text{within}) \) | 0.05* | 0.03* | 0.00  | -0.00 | 0.01  | -0.02* |
| \( \text{autoquote}_{it} \) Q4 | \( \rho(\text{within}) \) | 0.01* | 0.00  | -0.00 | -0.00 | -0.01 | 0.01  |
| \( \text{autoquote}_{it} \) Q5 | \( \rho(\text{within}) \) | -0.00 | 0.03* | -0.02* | 0.00  | 0.05* | -0.04* |

\( a: \) Based on the time means i.e. \( \bar{x}_i = \frac{1}{T} \sum_{t=1}^{T} x_{i,t}. \)  
\( b: \) Based on the deviations from time means i.e. \( x^*_i,t = x_{i,t} - \bar{x}_i. \)  
\( *: \) Significant at a 95% level.

\( F\)-tests reject null that instruments do not enter first-stage regression for all our IV regressions
IV regression including T/O, volatility, price, and size

\[ L_{it} = \alpha_i + \gamma_t + \beta A_{it} + \delta X_{it} + \varepsilon_{it} \]

Table 5: Effect of AT on Spread: Nonsynchronous Autoquote Introduction as Instrumental Variable

This table regresses various measures of the (half) spread on our algorithmic trading proxy. It is based on daily observations in the period when autoquote was phased in, i.e. December 2, 2003, through July 31, 2003. We use the exogenous nonsynchronous autoquote introduction to instrument for the endogenous algorithmic trading to identify causality from algorithmic trading to liquidity. We estimate

\[ L_{it} = \alpha_i + \gamma_t + \beta A_{it} + \delta X_{it} + \varepsilon_{it} \]

where \( L_{it} \) is a spread measure for stock \( i \) on day \( t \), \( A_{it} \) is the algorithmic trading measure \( \text{algo\_trad}_{it} \), and \( X_{it} \) is a vector of control variables, including share turnover, volatility, \( 1/\text{price} \), and \( \log \text{market\_cap} \). We always include fixed effects and time dummies. The set of instruments we use consists of all explanatory variables, except that we replace \( \text{algo\_trad}_{it} \) with \( \text{autoquote}_{it} \). We regress by quintile and report \( t \)-values based on standard errors that are robust to general cross-section and time-series heteroskedasticity and within-group autocorrelation (see Arellano and Bond (1991)).

<table>
<thead>
<tr>
<th>Coefficient on ( \text{algo_trad}_{it} )</th>
<th>Q1</th>
<th>Q2</th>
<th>Q3</th>
<th>Q4</th>
<th>Q5</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Panel A: quoted spread, quoted depth, and effective spread</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( q\text{spread}_{it} )</td>
<td>-0.52**</td>
<td>-0.42**</td>
<td>-0.43</td>
<td>-0.16</td>
<td>9.92</td>
</tr>
<tr>
<td></td>
<td>(-3.23)</td>
<td>(-2.21)</td>
<td>(-1.44)</td>
<td>(-0.05)</td>
<td>(1.22)</td>
</tr>
<tr>
<td>( q\text{depth}_{it} )</td>
<td>-3.47**</td>
<td>-1.43</td>
<td>-1.99</td>
<td>15.49</td>
<td>0.61</td>
</tr>
<tr>
<td></td>
<td>(-2.50)</td>
<td>(-1.16)</td>
<td>(-1.07)</td>
<td>(0.39)</td>
<td>(0.19)</td>
</tr>
<tr>
<td>( \text{espread}_{it} )</td>
<td>-0.18**</td>
<td>-0.32**</td>
<td>-0.35</td>
<td>-1.63</td>
<td>4.65</td>
</tr>
<tr>
<td></td>
<td>(-2.65)</td>
<td>(-2.23)</td>
<td>(-1.56)</td>
<td>(-0.42)</td>
<td>(1.16)</td>
</tr>
<tr>
<td><strong>Panel B: spread decompositions</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( r\text{spread}_{it} )</td>
<td>0.35**</td>
<td>0.76**</td>
<td>1.03**</td>
<td>14.26</td>
<td>15.88</td>
</tr>
<tr>
<td></td>
<td>(3.53)</td>
<td>(3.97)</td>
<td>(2.06)</td>
<td>(0.46)</td>
<td>(1.36)</td>
</tr>
<tr>
<td>( \text{adv_selection}_{it} )</td>
<td>-0.53**</td>
<td>-1.07**</td>
<td>-1.39**</td>
<td>-15.48</td>
<td>-11.21</td>
</tr>
<tr>
<td></td>
<td>(-3.57)</td>
<td>(-4.08)</td>
<td>(-2.06)</td>
<td>(-0.47)</td>
<td>(-1.33)</td>
</tr>
</tbody>
</table>

#observations: 1082*167 (stock*day)

*/**: Significant at a 95%/99% level.
$stdev_{\text{tradedcorr\_comp}}_{it}$ (stdev of trade–correlated component of eff. price innovations cf. Hasbrouck (1991a,1991b) (bps))

Q1
Q3
Q5
Q2
Q4
95% conf. interval
$\text{stdev\_nontradecorr\_comp}_{it}$ (stdev of non-trade-correlated component of eff. price innovations cf. Hasbrouck (1991a, 1991b) (bps))

Q1
Q3
Q5
Q2
Q4
95% conf. interval
IV regression for LSB and Hasbrouck decompositions

\[ M_{it} = \alpha_i + \gamma_t + \beta A_{it} + \delta X_{it} + \varepsilon_{it} \]

<table>
<thead>
<tr>
<th>Coefficient on algo_trad_{it}</th>
<th>Q1</th>
<th>Q2</th>
<th>Q3</th>
<th>Q4</th>
<th>Q5</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Panel A: Lin, Sanger, and Booth (1995)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( LSB_{95_fixed_{it}} )</td>
<td>0.26**</td>
<td>0.59**</td>
<td>0.69**</td>
<td>9.91</td>
<td>8.97</td>
</tr>
<tr>
<td></td>
<td>(3.63)</td>
<td>(4.16)</td>
<td>(2.26)</td>
<td>(0.46)</td>
<td>(1.36)</td>
</tr>
<tr>
<td>( LSB_{95_adv_sel_{it}} )</td>
<td>-0.26**</td>
<td>-0.61**</td>
<td>-0.84**</td>
<td>-12.19</td>
<td>-7.72</td>
</tr>
<tr>
<td></td>
<td>(-3.46)</td>
<td>(-3.80)</td>
<td>(-2.14)</td>
<td>(-0.46)</td>
<td>(-1.32)</td>
</tr>
<tr>
<td>( LSB_{95_order_persist_{it}} )</td>
<td>-0.18**</td>
<td>-0.30**</td>
<td>-0.21</td>
<td>0.66</td>
<td>3.30</td>
</tr>
<tr>
<td></td>
<td>(-3.06)</td>
<td>(-3.10)</td>
<td>(-1.60)</td>
<td>(0.28)</td>
<td>(1.21)</td>
</tr>
<tr>
<td><strong>Panel B: “Hasbrouck decomposition”</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( stdev_tradecorr_comp_{it} )</td>
<td>-0.22**</td>
<td>-0.26**</td>
<td>-0.30*</td>
<td>-3.39</td>
<td>-0.57**</td>
</tr>
<tr>
<td></td>
<td>(-2.62)</td>
<td>(-3.08)</td>
<td>(-1.69)</td>
<td>(-0.30)</td>
<td>(-2.73)</td>
</tr>
<tr>
<td>( stdev_nontradecorr_comp_{it} )</td>
<td>0.13**</td>
<td>0.13**</td>
<td>0.13</td>
<td>1.03</td>
<td>0.13</td>
</tr>
<tr>
<td></td>
<td>(2.48)</td>
<td>(2.36)</td>
<td>(1.47)</td>
<td>(0.28)</td>
<td>(1.12)</td>
</tr>
</tbody>
</table>

#observations: 1082*167 (stock*day)

 adultos: Significant at a 95%/99% level.
Interpretation: generalized Roll model

i.i.d. innovation in efficient price in each of two periods,
\[ m_t = m_{t-1} + w_t, \text{ with } w_t \in \{-\varepsilon, +\varepsilon\} \text{ equally likely} \]

1. At \( t = 0 \), risk-neutral humans submit a bid and ask quote and, given full competition, the first one arriving bids her reservation price.

2. At \( t = 1 \), humans can buy the information \( w_1 \) at cost \( c \). If bought, they can submit a new limit order.

3. At \( t = 2 \), two informed liquidity demanders arrive, one with a positive private value associated with a trade, \( +\theta \), the other with a negative private value, \( -\theta \).

Assume

1. \( 2c > \theta \) i.e. cost of “observing” for humans is sufficiently high (“quotes become stale”)

2. \( \varepsilon > \theta \) i.e. large innovations prevent simultaneous transaction by both liquidity demanders (unimportant)
Interpretation: generalized Roll model (ctd)

Humans only

The "game" has two periods, each with an i.i.d. innovation in the efficient price:

\[ m_t = m_{t-1} + w_t, \]

where \( w_t \in \{ \epsilon, -\epsilon \} \), each with probability 0.5. The game features three stages:

- At \( t = 0 \), risk-neutral humans can submit a bid and ask quote and, given full competition, the first one arriving bids her reservation price.
- At \( t = 1 \), humans can buy the information \( w_1 \) at cost \( c \). If they buy the information, they can submit a new limit order.
- At \( t = 2 \), two informed liquidity demanders arrive, one with a positive private value associated with a trade, +\( \theta \), the other with a negative private value, -\( \theta \).

We assume that \( 2c > \theta \), i.e., the cost of “observing” information for humans is sufficiently high that they do not update their quotes. The technical assumption \( 2\epsilon > \theta \) is also required so that only one of the two arriving liquidity demanders transacts at \( t = 2 \).

There are four equally likely paths through the binomial tree:

- uu
- ud
- du
- dd

where \( u \) represents a positive increment of \( \epsilon \) to the fundamental value and \( d \) is a negative increment. In equilibrium, humans do not buy the \( w_1 \) information and update the quote at \( t = 1 \), since they have to quote so far away from the efficient price to make up for \( c \) that neither liquidity demander will transact at that quote (as \( 2c > \theta \)). Given that they do not acquire the \( w_1 \) information, humans protect themselves by setting the bid price equal to

\[
\text{probability} \quad \text{state} \quad \text{efficient price} \quad \text{transaction price}
\]

\[
\begin{array}{cccc}
.25 & uu & m_{2uu} & m_{2uu} \\
.50 & ud \text{ and } du & m_{2ud} = m_{2du} & \text{no transaction} \\
.25 & dd & m_{2dd} & m_{2dd}
\end{array}
\]

- at \( t = 1 \) public information does not enter quotes
- “welfare loss” due to possible unrealized private value
Interpretation: generalized Roll model (ctd)

Introduce an algo that buys information at zero cost

\[
\begin{array}{cccc}
| \text{probability} | \text{state} | \text{efficient price} & \text{transaction price} \\
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>.25</td>
<td>uu</td>
<td>(m^u_2)</td>
<td>(m^u_2)</td>
</tr>
<tr>
<td>.25</td>
<td>ud</td>
<td>(m^u_d = m^d_u)</td>
<td>(m^d_u - \theta)</td>
</tr>
<tr>
<td>.50</td>
<td>du</td>
<td>(m^d_u = m^d_u)</td>
<td>(m^d_u + \theta)</td>
</tr>
<tr>
<td>.25</td>
<td>dd</td>
<td>(m^d_d)</td>
<td>(m^d_d)</td>
</tr>
</tbody>
</table>
\end{array}
\]

- at \(t = 1\) public information enters quotes, but midquote becomes “noisy” measure of true value
- no unrealized private value
Interpretation: generalized Roll model (ctd)

Efficient price is revealed without trades i.e. public information enters quotes without trades

Revenue to liquidity suppliers is positive

Also matches other findings: more frequent trades, narrower quotes

Note: model assumes that algo competition is less intense than human competition
Conclusion

1. Panel regressions time-series increases in algo trading correlate with liquidity improvement

2. Staggered introduction of structural change (autoquote) as an instrument confirms algo trading lowers trading cost and increases price informativeness

3. Surprisingly, revenues to liquidity suppliers increase with algo trading. Market power for some period after introduction?
Does Algorithmic Trading Improve Liquidity?

Terry Hendershott\(^1\) Charles M. Jones\(^2\) Albert J. Menkveld\(^3\)

\(^1\)Haas School of Business, UC Berkeley
\(^2\)Graduate School of Business, Columbia University
\(^3\)Tinbergen Institute, VU University Amsterdam

WFA, June 2008


The summary informativeness of stock trades: An econometric analysis.

Technology and liquidity provision: The blurring of traditional definitions.


Continuous auctions and insider trading.
Econometrica 53, 1315–1335.

Trade size and components of the bid-ask spread.

Econometric models of limit-order executions.

Optimal trading strategy and supply/demand dynamics.
Technical report, MIT.

The specialist’s order book and price anomalies.
Technical report, Harvard University.