Modeling Trade Direction

Dale W.R. Rosenthal

UIC Finance
Liautaud Graduate School of Business

7 March 2009
Motivation

- Financial markets trades result from two or more orders.
- Later arriving order: the *initiator* (aggressor).
- Was the initiator a buy or a sell? aka
  - What was the initiating trade’s *direction?* *sign?* *side?*
- Needed for some microstructure research, *e.g.* price impact.
  - Trades impart a small bias (impact) to the price process.
  - Price impact modeled as function of trade size and direction.
- Initiating side (buy/sell) is not available in real-time.
- Fitting impact models is hard, can save $ billions/year.
- Want to guess initiator as accurately as possible.
Lee and Ready (1991) first considered delays.
- Compare trade to midpoint with earlier timestamp.
- 1988 NYSE and AMEX data: 5 seconds; 1987 data: 2 seconds.
- Resolve ties with tick test (+ and 0+ ticks: buys)

The current debate: What method and delay to use?
- Bid/ask test, 0s: Ellis, Michaely, and O’Hara (2000); Peterson and Sirri (2003).
Problems with Previous Studies

Previous work on trade signing has some problems:

**Old Data**  Pre-electronic, pre-decimalization trades.  
1987 (Lee and Ready) to 1999 (Henker and Wang).

**Narrow Data**  Trades for only a few stocks.  
144 (TORQdb) to 401 (Henker and Wang).

**Biased Data**  Only large-cap stocks (*all* preceding studies).

**Time Skew**  No simultaneous analyses of NYSE, Nasdaq trades.

**Polluted?**  Some now-common problems affect many studies.

Why care? “This delay is decreasing to nearly 0 seconds.” 
Still a problem: delay decreased, but quote volume increased.
Better Quotes and a Modeled Approach

- Picking the correct prevailing quotes may be noisy.
  - Instead, try to get *close* to the prevailing quote.
- Average quotes across time via approximate delay distribution.
- Also use an approach that allows for richer models:
  - Include other information (*e.g.* tick test, bid/ask test);
  - Account for information strengths;
  - Allow for auto-correlated and cross-correlated buys/sells.
  - Acknowledge differences in markets (*e.g.* NYSE vs. Nasdaq).
  - Accommodate effects of market capitalization, liquidity, etc.
- Model can even estimate probability of correct prediction.
Model Notation

\( b_t, a_t, m_t \) = bid, ask, midpoint initiator saw at time \( t \).
\( p_t \) = price of trade at time \( t \);
\( p_{t-} \) = price of trade preceding time \( t \);
\( p'_{t-} \) = differing trade price preceding time \( t \);
\( B_t \) = side of trade at time \( t \) (1=buy, 0=sell);
\( g \) = normalized difference function, e.g. \( \log(p_t) - \log(\hat{m}_t) \);
\( J \) = signed indicator-like function (-1,+1 if \( p_t \approx \hat{b}_t, \hat{a}_t \); 0 else).

\( J \) needed: estimated quotes may not be decimalized.

\( J \) for 1% spread; \( \tau: -0.1% \ - - 1% \ - - 5% \)
Trade Direction Model

\[ P(B_{jt} = \text{Buy} | \mathcal{F}_t; \theta_o, c_k, d_{kl}) = \pi_{jt} = \logit(\eta_{jt}) \]

\[ \eta_{jt} = \beta_0 + \beta_{o1} g(p_{jt}, \hat{m}_{jt}) + \beta_{o2} g(p_{jt}, p'_{jt-}) + \beta_{o3} J(p_t, \hat{b}_t, \hat{a}_t) + \phi_o \eta_{jt-} + c_k + d_{kl} \]

(1)

\(j\) indexes stocks; \(k\) indexes ten-minute time “bins”; \(\ell\) indexes sectors; \(o\) indexes markets.

- Random effects: handle (+) correlations, pseudoreplication.
- Instead of \(\phi_o \eta_{jt-}\) AR term, used lagged metrics.
Dataset and Estimation

- Use ArcaTrade dataset from NYSE Archipelago ECN.
  - Includes initiating side for NYSE, Nasdaq, and AMEX stocks\(^1\).
  - Universe: 2,836 different stocks (2004 “Russell 3000”).
  - Dec 2004: 1, 2 for estimation; 3–31 for out-of-sample testing.
- In-sample estimation uses almost 2.2 MM observations.
- Out-of-sample testing uses 16.5 MM observations.
- Nonlinear parameters found by conjugate direction (CD).
  - CD uses loop: try parameters, estimate quotes, fit GLMM.
  - Penalized quasi-likelihood used to fit GLMM.

\(^1\)Volume share: 2.3%, 22.5%, 23.3% of NYSE, Nasdaq, AMEX.
## Estimated Model

<table>
<thead>
<tr>
<th>Fixed Effect</th>
<th>AMEX</th>
<th>Nasdaq</th>
<th>NYSE</th>
</tr>
</thead>
<tbody>
<tr>
<td>$J$ width $\tau$</td>
<td>Overall: $2.1 \times 10^{-4}$ (0.3)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Delay scale $\nu$</td>
<td>1.66 (0.58)</td>
<td>1.65 (0.65)</td>
<td>0.62 (0.47)</td>
</tr>
<tr>
<td>Delay rate $\lambda$</td>
<td>0.35 (3.7)</td>
<td>0.33 (0.40)</td>
<td>0.78 (0.35)</td>
</tr>
<tr>
<td>Intercept</td>
<td>Overall: 0.06 (0.02)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Midpoint</td>
<td>—</td>
<td>209 (11)</td>
<td>122 (13)</td>
</tr>
<tr>
<td>Tick</td>
<td>—</td>
<td>29.4 (8.4)</td>
<td>-20.5 (8.5)</td>
</tr>
<tr>
<td>Bid/Ask</td>
<td>1.20 (0.25)</td>
<td>1.41 (0.02)</td>
<td>2.04 (0.20)</td>
</tr>
<tr>
<td>Prev. Bid/Ask</td>
<td>0.33 (0.31)</td>
<td>-0.14 (0.01)</td>
<td>-0.17 (0.05)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Random Effect</th>
<th>Std. Dev.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Time Bin</td>
<td>0.08 (0.01)</td>
</tr>
<tr>
<td>Sector × Time Bin</td>
<td>0.27 (0.03)</td>
</tr>
</tbody>
</table>

Overdispersion Parameter: 1.0086
Estimation Summary

- Negative prior bid/ask coefficient: agrees with bid-ask bounce.
- Opposite tick coefficient signs: differing short-sale price tests?
- Random effects non-zero, imply buying/selling correlation of:
  - 0.2% across all stocks in 10-minute period.
  - 2% across same-sector stocks in 10-minute period.
- Delay parameter fitting preferred old quotes (30s–120s)
  - Indicates ultra-short-term persistence of quote changes.
- Overdispersion parameter not of practical concern.
### Out of Sample: Across Markets

<table>
<thead>
<tr>
<th>Market</th>
<th>N</th>
<th>Percent of Trades Correctly Classified</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Modeled</td>
</tr>
<tr>
<td>AMEX</td>
<td>19,435</td>
<td>69.8%</td>
</tr>
<tr>
<td>Nasdaq</td>
<td>15,220,579</td>
<td>74.3%</td>
</tr>
<tr>
<td>NYSE</td>
<td>1,264,866</td>
<td>80.7%</td>
</tr>
<tr>
<td>Overall</td>
<td>16,504,880</td>
<td>74.7%</td>
</tr>
</tbody>
</table>

**EMO** = Ellis, Michaely, and O’Hara bid/ask test.  
**LR** = Lee and Ready midpoint test.  
**Tick** = tick test.  

- **Shocker:** LR is the current “gold standard”.

---
Out of Sample: Across Sectors, Spread, Time

- Sectors: Best method across all sectors except one (small).
- Spread: Best method across spread with two exceptions:
  - 0.1% less accurate for 4.4MM trades at ask; and,
  - More abysmal than winner\(^2\) for 30,000 trades at midpoint.
- Dates: Best method for each out-of-sample date.

\(^2\)45.5\% vs. 48.8\%.
### Performance Attribution: Results

<table>
<thead>
<tr>
<th>Market</th>
<th>N</th>
<th>Baseline (All Tests)</th>
<th>Convert Tests to Metrics</th>
<th>Add Lag-1 Metrics</th>
<th>Ad-hoc Delay</th>
<th>Full Model</th>
</tr>
</thead>
<tbody>
<tr>
<td>AMEX</td>
<td>19,435</td>
<td>67.7%</td>
<td>+2.5%</td>
<td>+0.4%</td>
<td>-0.8%</td>
<td>+0.0%</td>
</tr>
<tr>
<td>Nasdaq</td>
<td>15,220,579</td>
<td>70.3%</td>
<td>+3.0%</td>
<td>-0.1%</td>
<td>+0.9%</td>
<td>+0.2%</td>
</tr>
<tr>
<td>NYSE</td>
<td>1,264,866</td>
<td>79.8%</td>
<td>+1.1%</td>
<td>-0.6%</td>
<td>+0.7%</td>
<td>-0.3%</td>
</tr>
<tr>
<td>Overall</td>
<td>16,504,880</td>
<td>71.1%</td>
<td>+2.7%</td>
<td>-0.1%</td>
<td>+0.9%</td>
<td>-0.1%</td>
</tr>
</tbody>
</table>

- To attribute performance, I fit a series of nested models.
- Information strength (±1 tests to metrics) gains 1%–3%.
- Adding lagged bid/ask metric gains 0.4% for AMEX trades.
- Basic delay model gains 0.8% for NYSE, Nasdaq trades.
Contributions and Further Work

- Beat next-best method by 1–2%\(^3\) across almost all groupings.
- Introduced delay theory to estimation of prevailing quotes.
- Opened doors to richer trade signing models:
  - Use multiple sources of information.
  - Consider strength of information.
  - Correct for microstructure peculiarities.
  - Allow for autocorrelations and cross-correlations.
  - Interaction between volume/volatility/spread and metrics?
- Developed Edgeworth expansions for average delays.
- Conduct experiments to infer BLUPs and make money?

\(^3\)cf. Most published results beat EMO or LR by 0.5% in one group.