 Modeling Limit Order Book Dynamics and Predicting Mid-Price Movement Using Hawkes Processes

Yuanda Chen
Major Advisor: Dr. Alec N. Kercheval

Department of Mathematics
Florida State University

October 22, 2015
Introduction

1. Electronic system containing limit orders.

2. Interested buyers and sellers submit orders at a fixed price.

3. A limit order sits in the order book until it is either executed against a market order or canceled.

4. Three market events: adding a new order, execution and cancellation.

5. Queue sizes change.


7. Mid-price moves when either best ask or best bid queue is depleted, when the spread is 1.
1. High frequency trading and data make modeling at transaction level possible.

2. Irregular occurrence suggests using of point processes.


An example

**Figure:** An Example of the Intensity of a Hawkes Process
($\nu = 1.2, \alpha = 0.6, \beta = 0.8$)
Univariate Hawkes process

Definition

A simple version of Hawkes process $\lambda(t)$ is defined by using a constant base intensity and an exponential kernel $\gamma(t) = \alpha e^{-\beta t}$:

(i) $N(0) = 0$.

(ii) $\lambda(t) = \nu + \alpha \int_0^t e^{-\beta(t-s)} dN_s = \nu + \alpha \sum_{t_i < t} e^{-\beta(t-t_i)}$.

(iii) $P\{N(t + h) - N(t) = 1|\mathcal{F}_t\} = \lambda(t)h + o(h)$.

(iv) $P\{N(t + h) - N(t) \geq 2|\mathcal{F}_t\} = o(h)$. 
Multi-variate Hawkes processes

- 4-dimensional Hawkes processes for 4 types of market events.
- Their intensity functions interact with each other.
- Likelihood function: Rubin (1972) ”Regular point processes and their detection”.
- QQ-plot: Ogata (1988) ”Statistical models for earthquake occurrences and residual analysis for point processes”.

Mid-price movement prediction

- 10000 independent simulations to estimate the probability that the next mid-price move is upward.
- Multiple predictions made before the next move.
- Prediction: up signal and down signal.
- Labeling rate and accuracy
## Prediction performance

<table>
<thead>
<tr>
<th>Training/Testing Period</th>
<th>Sample</th>
<th>Labelled</th>
<th>Correct</th>
<th>Sample</th>
<th>Labelled</th>
<th>Correct</th>
<th>Sample</th>
<th>Labelled</th>
<th>Correct</th>
</tr>
</thead>
<tbody>
<tr>
<td>10:00-11:00</td>
<td>20947</td>
<td>11973(57%)</td>
<td>9323(78%)</td>
<td>19343</td>
<td>6158(32%)</td>
<td>4742(77%)</td>
<td>22760</td>
<td>6679(29%)</td>
<td>5581(84%)</td>
</tr>
<tr>
<td>11:00-11:30</td>
<td>5015(25%)</td>
<td>7672(38%)</td>
<td>4138(83%)</td>
<td>17391</td>
<td>7989(46%)</td>
<td>5444(68%)</td>
<td>19930</td>
<td>15910</td>
<td>5638(73%)</td>
</tr>
<tr>
<td>12:00-12:30</td>
<td>9151(57%)</td>
<td>9994(68%)</td>
<td>7195(79%)</td>
<td>16039</td>
<td>2689(16%)</td>
<td>4468(73%)</td>
<td>14739</td>
<td>14693</td>
<td>6952(70%)</td>
</tr>
<tr>
<td>13:00-13:30</td>
<td>7176(61%)</td>
<td>9277(66%)</td>
<td>3856(54%)</td>
<td>11848</td>
<td>16460</td>
<td>6559(71%)</td>
<td>14149</td>
<td>12333</td>
<td>4468(73%)</td>
</tr>
<tr>
<td>14:00-14:30</td>
<td>11249(68%)</td>
<td>10176(83%)</td>
<td>6399(57%)</td>
<td>22632</td>
<td>7738(34%)</td>
<td>6147(79%)</td>
<td>22967</td>
<td>18889(82%)</td>
<td>13707(73%)</td>
</tr>
<tr>
<td>15:30-16:00</td>
<td>42.67%</td>
<td>54.18%</td>
<td>71.76%</td>
<td>70.74%</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Figure:** Prediction Performance with Trust Level of 0.25
Testing strategy

- Although not predicting profit directly, this should be better than knowing nothing.

- at long position:
  - up signal: buy 1 share
  - down signal: sell all shares

- at short position:
  - up signal: buy all shares back
  - down signal: short 1 share
Strategy performance with INTC

**Figure:** Strategy performance with INTC
Figure: Strategy performance with MSFT
Diffusion approximation

- Approximate the queue size processes by correlated Brownian motions.
- Enables fast calculation
- How good the approximation is?