Non-parametric prediction of mid-price dynamics in a limit order book

Deepan Palguna 1  Ilya Pollak 2

1Purdue University

2Purdue University and Pragma Securities
dpalguna@purdue.edu, ipollak@ecn.purdue.edu

October 24, 2013

5th Annual Modeling High Frequency Data in Finance Conference, Stevens Institute of Technology, NJ
Motivation and description of the approach

- Main motivation
  - Executing a large order within a fixed duration
Motivation and description of the approach

Main motivation
- Executing a large order within a fixed duration

Approach to mid-price prediction
- Describe order book configuration using a state
- Use historical data to estimate empirical means of future mid-price changes given the current state
- Use the empirical means and empirical probabilities (features) to predict future mid-price changes
State at time $t$ described using the order imbalance $S(t)$ which is:

$$S(t) = \left\lfloor \frac{\text{Number of shares at the highest bid price at time } t}{100} \right\rfloor - \left\lfloor \frac{\text{Number of shares at the lowest ask price at time } t}{100} \right\rfloor$$

State : $8 - 2 = 6$

Figure: Example order book state
Clustering states

- Problem: Features cannot be estimated reliably if states are plentiful.
- Cluster similar states to increase number of observations for feature estimation.
- Distance between states $S_1$ and $S_2$ measured using Euclidean distance if $\text{sign}(S_1) = \text{sign}(S_2)$.
- Distance set to $\infty$ otherwise.
- Cluster by iteratively merging states till smallest cluster probability exceeds $P_{\text{min}}$. 
\[ \delta = \text{a parameter, in units of time, to be selected} \]

\[ P(t) = \text{mid-price at time } t \]

\[ \Omega(t) = \text{cluster at time } t \]

Given \( \Omega(t) \), we define four conditional features:

- The conditional probability that \( P(t) \neq P(t + \delta) \)
- The conditional expectation of \( P(t + \delta) - P(t) \)
- The conditional expectation of \( \tau(t) \), the time until the first mid-price change after time \( t \)
- The conditional expectation of \( P(t + \tau(t)) - P(t) \)

- Estimate features as temporal averages
Algorithm to detect mid-price change between times $t$ and $t + \delta$

**Figure:** First stage: Change detection

1. **Rule 1**
   - Conditional probability of non-zero mid price change $> 0$?
   - If both are $> 0$, predict change.
   - If both are $< 0$, predict no change.
   - If the two disagree, pick the prediction of the rule with the lower empirical error rate.

2. **Rule 2**
   - Conditional time to first mid-price change $> 0$?
Up or down prediction given that change was detected

Figure: Second stage: Sign prediction
Order execution application

- **Aim**
  - Buy $X_0$ shares of a stock over time steps $t_0, \cdots, t_N$ that are $\delta$ seconds apart
  - In doing so, minimize average price paid per share
Order execution application

- **Aim**
  - Buy $X_0$ shares of a stock over time steps $t_0, \cdots, t_N$ that are $\delta$ seconds apart
  - In doing so, minimize average price paid per share

**Uniform Benchmark**

- Send market orders of size $X_0/(N + 1)$ at each time step
Our execution algorithm

- A modification of uniform execution that makes use of our predictors

- \( X_{t_i} \) = Number of shares still to be bought just before time \( t_i \)

- At time \( t_i \), for parameter \( \pi \in [0, 1] \), execute buy market order of size

\[
\begin{align*}
\min \left( X_{t_i}, \frac{X_0}{N + 1} \right) & \quad \text{if the price is predicted to stay the same,} \\
\min \left( X_{t_i}, \frac{X_0(1 + \pi)}{N + 1} \right) & \quad \text{if the price is predicted to go up,} \\
\min \left( X_{t_i}, \frac{X_0(1 - \pi)}{N + 1} \right) & \quad \text{if the price is predicted to go down.}
\end{align*}
\]
Simulations with order book data

- Order book data from NYSE openbook for June 30, 2010
- Complete order series to construct limit order books for NYSE securities

**Performance measured using**

\[
\frac{\text{Cost of uniform execution}}{\text{Cost of uniform execution}} - \frac{\text{Cost of our strategy}}{\text{Cost of uniform execution}}
\]

Our strategy outperforms uniform execution if the above quantity is consistently positive.
Evaluating the predictors: Simulation setup

- Difficulty in simulating permanent market impact. Simulated executions leave the order book unchanged.

- Execute $X_0$ shares between 3–4 pm. $X_0$ is 1% of average hourly volume over June 1–29, 2010.

- Prediction time interval $\delta$ used by the predictors $\rightarrow \{0.5, 1, 2\}$ second.

- Discretization time step $\Delta t \rightarrow 1$ ms.

- Trade acceleration-deceleration parameter $\pi \rightarrow 1$. 

Evaluating the predictors: Simulation results

Measure the significance of the results using standard errors estimated cross-sectionally over \( S = 100 \) liquid stocks

- \( S \) stocks produce relative improvements \( R_1, \ldots, R_S \)
- \( \mu \) and \( \sigma \) estimated as sample mean and sample standard deviation of \( R_1, \ldots, R_S \)
- Assume improvements are uncorrelated with same variance and estimate standard error as \( \hat{\eta} = \frac{\hat{\sigma}}{\sqrt{S}} \)
Simulation results

Table: Cross-sectional mean, sample standard deviation, and standard errors of relative execution cost improvements for $S = 100$ most liquid stocks. Clustering probability constraint, $P_{\text{min}} = 0.03$.

<table>
<thead>
<tr>
<th>$\delta$ (seconds)</th>
<th>0.5</th>
<th>1</th>
<th>2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean improvement ($\hat{\mu}$) $\times 10^{-4}$</td>
<td>3.041</td>
<td>2.847</td>
<td>3.136</td>
</tr>
<tr>
<td>Standard deviation ($\hat{\sigma}$) $\times 10^{-4}$</td>
<td>11.034</td>
<td>10.13</td>
<td>10.220</td>
</tr>
<tr>
<td>Standard error ($\hat{\eta} = \hat{\sigma}/\sqrt{S}$) $\times 10^{-5}$</td>
<td>11.034</td>
<td>10.13</td>
<td>10.220</td>
</tr>
<tr>
<td>$\hat{\mu}/\hat{\eta}$</td>
<td>2.756</td>
<td>2.810</td>
<td>3.068</td>
</tr>
</tbody>
</table>

- **Statistical significance:** All mean to standard error ratios $> 2$
- **Economic significance:** For a company trading $100$ million worth shares daily, 2 basis-point improvement in trading costs $\Rightarrow$ annual savings of $5$ million $= 2 \times 10^{-4} \times 100 \times 10^6 \times 250$
- **Directly predicting ask price movements with 3-D states gives higher improvements**
Thank you