NON-PARAMETRIC PREDICTION OF PRICE DYNAMICS IN LIMIT ORDER BOOKS

Deepan Palguna[†] and Ilya Pollak[∗]
[†]Purdue University and [∗]Pragma Securities.

Introduction

**Contribution:** Novel non-parametric method to predict mid-price dynamics.

**Applications:**
- A broker in charge of executing a large trade.
- An asset manager contemplating a portfolio rebalancing.
- A regulator or exchange in charge of maintaining orderly markets.

Limit order books

- Order book at time $t$ summarized using order imbalance $S(t)$.
- Estimate features conditioned on states using empirical averages.
- Cluster similar states together to enable accurate feature estimation.
- Use cluster features for predicting mid-price movements.

Outline of prediction approach

**Prediction of mid-price change**

- At time $t$, state $S(t)$ and corresponding cluster $Ω(t)$ are computed.
- Mid-price change $\delta$ seconds into the future is predicted in two stages, using features associated with $Ω(t)$.

Clustering by iterative merging

- Distance between two clusters $Ω_1$ and $Ω_2$ is the Euclidean distance between their centers if $\text{sign}(\text{center}(Ω_1)) = \text{sign}(\text{center}(Ω_2))$. Otherwise distance is $\infty$.

Performance Analysis

- **Aim:** Buy $X_0$ shares of a stock over time steps $t_0, \ldots, t_N$ that are $\delta$ seconds apart.
- **Uniform Benchmark:** Send market orders of size $X_0/(N+1)$ at each time step.
- **Our modification of uniform benchmark:**
  - $X_t =$ Number of shares to be bought just before time $t_t$.
  - At time $t_t$, for parameter $\pi \in [0, 1]$, execute buy market order of size.
  \[
  \min \left\{ \frac{X_t}{N+1}, \frac{X_t(1+\pi)}{N+1}, \frac{X_t(1-\pi)}{N+1} \right\}
  \]
  if the price is predicted to stay the same,

Simulation results

- **Performance measured using**
  \[
  \text{Cost of uniform execution} - \text{Cost of our strategy} = \frac{\text{Cost of uniform execution}}{\text{Cost of our strategy}}
  \]
- Execute 1% of average hourly volume from 3 pm to 4 pm using clusters computed from 9:30 am for:
  - $S =$ 100 most liquid stocks
  - Clustering probability ($P_{\text{mm}}$) = 0.03, $\pi = 1$

<table>
<thead>
<tr>
<th>$\delta$ (seconds)</th>
<th>0.5</th>
<th>1</th>
<th>2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean Improvement ($\mu$) $\times 10^{-4}$</td>
<td>3.041</td>
<td>2.847</td>
<td>3.136</td>
</tr>
<tr>
<td>Standard Deviation ($\sigma$) $\times 10^{-4}$</td>
<td>11.034</td>
<td>10.13</td>
<td>10.220</td>
</tr>
<tr>
<td>Standard Error ($\eta$) $\times 10^{-5}$</td>
<td>11.034</td>
<td>10.13</td>
<td>10.220</td>
</tr>
<tr>
<td>$\mu/\eta$</td>
<td>2.756</td>
<td>2.819</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 $= 2 \times 10^{-4} \times 100 \times 10^6 \times 250$. 