Time is Money:
Estimating the Cost of Latency in Trading

Sasha Stoikov

Cornell University

April 16, 2015
Background

- Automated or computerized trading
  - Accounts for 70% of equity trades taking place in the US
Background

- Automated or computerized trading
  - Accounts for 70% of equity trades taking place in the US
- High frequency trading
  - Short position-holding periods
  - Market-making (payment for order flow)
  - Latency arbitrage across trading venues
Background

- Automated or computerized trading
  - Accounts for 70% of equity trades taking place in the US
- High frequency trading
  - Short position-holding periods
  - Market-making (payment for order flow)
  - Latency arbitrage across trading venues
- Algorithmic trading
  - Brokers executing large client transactions
  - Optimally splitting client orders
  - Opportunistic trading algorithms (React, Bolt, Stealth, Ambush, Guerrilla, Sniper)
Background

- Automated or computerized trading
  - Accounts for 70% of equity trades taking place in the US
- High frequency trading
  - Short position-holding periods
  - Market-making (payment for order flow)
  - Latency arbitrage across trading venues
- Algorithmic trading
  - Brokers executing large client transactions
  - Optimally splitting client orders
  - Opportunistic trading algorithms (React, Bolt, Stealth, Ambush, Guerrilla, Sniper)
- Why low-latency trading?
Market twenty years ago: the pit
Market today: the order book

<table>
<thead>
<tr>
<th>Mkt Name</th>
<th>Price</th>
<th>Size</th>
<th>Curr Size</th>
<th>Avg Price</th>
</tr>
</thead>
<tbody>
<tr>
<td>NSDJ</td>
<td>47.96</td>
<td>68</td>
<td>68</td>
<td>47.96</td>
</tr>
<tr>
<td>NBK</td>
<td>47.96</td>
<td>2</td>
<td>70</td>
<td>47.96</td>
</tr>
<tr>
<td>BATS</td>
<td>47.96</td>
<td>12</td>
<td>62</td>
<td>47.96</td>
</tr>
<tr>
<td>DRCTEDGE</td>
<td>47.96</td>
<td>1</td>
<td>83</td>
<td>47.96</td>
</tr>
<tr>
<td>ARCA</td>
<td>47.96</td>
<td>128</td>
<td>211</td>
<td>47.96</td>
</tr>
<tr>
<td>NSDJQ</td>
<td>47.96</td>
<td>906</td>
<td>1.117</td>
<td>47.952</td>
</tr>
<tr>
<td>EDEBA</td>
<td>47.96</td>
<td>123</td>
<td>1.246</td>
<td>47.982</td>
</tr>
<tr>
<td>CHX</td>
<td>47.95</td>
<td>59</td>
<td>1.298</td>
<td>47.952</td>
</tr>
<tr>
<td>CBK</td>
<td>47.95</td>
<td>35</td>
<td>1.322</td>
<td>47.952</td>
</tr>
<tr>
<td>BEX</td>
<td>47.95</td>
<td>193</td>
<td>1.485</td>
<td>47.951</td>
</tr>
<tr>
<td>ARCA</td>
<td>47.95</td>
<td>893</td>
<td>2.343</td>
<td>47.951</td>
</tr>
<tr>
<td>NSDQ</td>
<td>47.94</td>
<td>1,626</td>
<td>3.969</td>
<td>47.948</td>
</tr>
<tr>
<td>ARCA</td>
<td>47.94</td>
<td>1,314</td>
<td>5.283</td>
<td>47.948</td>
</tr>
<tr>
<td>KSO</td>
<td>47.93</td>
<td>1,252</td>
<td>6.833</td>
<td>47.940</td>
</tr>
<tr>
<td>ARCA</td>
<td>47.93</td>
<td>1,313</td>
<td>8.146</td>
<td>47.940</td>
</tr>
<tr>
<td>TMQ</td>
<td>47.92</td>
<td>10</td>
<td>8.196</td>
<td>47.940</td>
</tr>
<tr>
<td>KSOQ</td>
<td>47.92</td>
<td>1,473</td>
<td>9.625</td>
<td>47.957</td>
</tr>
<tr>
<td>ARCA</td>
<td>47.92</td>
<td>1,201</td>
<td>10.830</td>
<td>47.935</td>
</tr>
<tr>
<td>UBS</td>
<td>47.91</td>
<td>1</td>
<td>10.631</td>
<td>47.935</td>
</tr>
<tr>
<td>HCR</td>
<td>47.91</td>
<td>1</td>
<td>10.632</td>
<td>47.935</td>
</tr>
<tr>
<td>KSO</td>
<td>47.91</td>
<td>1,504</td>
<td>12.556</td>
<td>47.952</td>
</tr>
<tr>
<td>KSO</td>
<td>47.90</td>
<td>1,362</td>
<td>13.698</td>
<td>47.929</td>
</tr>
<tr>
<td>KSO</td>
<td>47.89</td>
<td>1,334</td>
<td>15.062</td>
<td>47.925</td>
</tr>
<tr>
<td>ARCA</td>
<td>47.88</td>
<td>1,177</td>
<td>16.281</td>
<td>47.917</td>
</tr>
<tr>
<td>KSO</td>
<td>47.87</td>
<td>934</td>
<td>17.193</td>
<td>47.918</td>
</tr>
<tr>
<td>KSO</td>
<td>47.86</td>
<td>923</td>
<td>18.116</td>
<td>47.916</td>
</tr>
<tr>
<td>ARCA</td>
<td>47.85</td>
<td>1,173</td>
<td>19.926</td>
<td>47.904</td>
</tr>
<tr>
<td>KSO</td>
<td>47.85</td>
<td>982</td>
<td>19.066</td>
<td>47.910</td>
</tr>
<tr>
<td>KSO</td>
<td>47.84</td>
<td>940</td>
<td>19.948</td>
<td>47.909</td>
</tr>
<tr>
<td>KSOQ</td>
<td>47.83</td>
<td>900</td>
<td>20.746</td>
<td>47.908</td>
</tr>
<tr>
<td>ARCA</td>
<td>47.82</td>
<td>670</td>
<td>31.998</td>
<td>47.906</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Mkt Name</th>
<th>Price</th>
<th>Size</th>
<th>Curr Size</th>
<th>Avg Price</th>
</tr>
</thead>
<tbody>
<tr>
<td>NSDQ</td>
<td>47.97</td>
<td>1,261</td>
<td>1,261</td>
<td>47.970</td>
</tr>
<tr>
<td>EDEBA</td>
<td>47.97</td>
<td>243</td>
<td>1,554</td>
<td>47.970</td>
</tr>
<tr>
<td>ARCA</td>
<td>47.97</td>
<td>58</td>
<td>1,562</td>
<td>47.970</td>
</tr>
<tr>
<td>KSO</td>
<td>47.97</td>
<td>20</td>
<td>1,652</td>
<td>47.970</td>
</tr>
<tr>
<td>ARCA</td>
<td>47.97</td>
<td>112</td>
<td>1,714</td>
<td>47.970</td>
</tr>
<tr>
<td>KSO</td>
<td>47.97</td>
<td>239</td>
<td>2,073</td>
<td>47.970</td>
</tr>
<tr>
<td>ARCA</td>
<td>47.97</td>
<td>1,127</td>
<td>3,200</td>
<td>47.910</td>
</tr>
<tr>
<td>BATS</td>
<td>47.97</td>
<td>1,241</td>
<td>4,441</td>
<td>47.970</td>
</tr>
<tr>
<td>DRCTEDGE</td>
<td>47.97</td>
<td>424</td>
<td>4,865</td>
<td>47.970</td>
</tr>
<tr>
<td>ARCA</td>
<td>47.97</td>
<td>1,648</td>
<td>6,514</td>
<td>47.970</td>
</tr>
<tr>
<td>NSDQ</td>
<td>47.97</td>
<td>1,576</td>
<td>7,990</td>
<td>47.974</td>
</tr>
<tr>
<td>NSDQ</td>
<td>47.97</td>
<td>1,562</td>
<td>9,452</td>
<td>47.977</td>
</tr>
<tr>
<td>ARCA</td>
<td>47.97</td>
<td>1,346</td>
<td>10,000</td>
<td>47.978</td>
</tr>
<tr>
<td>KSO</td>
<td>47.96</td>
<td>1,648</td>
<td>12,246</td>
<td>41.981</td>
</tr>
<tr>
<td>ARCA</td>
<td>47.96</td>
<td>1,265</td>
<td>13,553</td>
<td>47.963</td>
</tr>
<tr>
<td>ARCA</td>
<td>47.96</td>
<td>1,494</td>
<td>15,027</td>
<td>47.965</td>
</tr>
<tr>
<td>ARCA</td>
<td>47.96</td>
<td>1,261</td>
<td>15,200</td>
<td>47.977</td>
</tr>
<tr>
<td>ARCA</td>
<td>47.96</td>
<td>1,323</td>
<td>17,591</td>
<td>47.960</td>
</tr>
<tr>
<td>UBS</td>
<td>47.96</td>
<td>1,322</td>
<td>18,913</td>
<td>47.952</td>
</tr>
<tr>
<td>ARCA</td>
<td>47.96</td>
<td>1,061</td>
<td>19,974</td>
<td>47.995</td>
</tr>
<tr>
<td>ARCA</td>
<td>47.96</td>
<td>5</td>
<td>19,998</td>
<td>47.995</td>
</tr>
<tr>
<td>ARCA</td>
<td>47.96</td>
<td>1,062</td>
<td>21,011</td>
<td>47.998</td>
</tr>
<tr>
<td>ARCA</td>
<td>47.96</td>
<td>1</td>
<td>21,012</td>
<td>47.998</td>
</tr>
<tr>
<td>ARCA</td>
<td>47.96</td>
<td>965</td>
<td>21,977</td>
<td>48.000</td>
</tr>
<tr>
<td>ARCA</td>
<td>47.96</td>
<td>1,043</td>
<td>23,000</td>
<td>48.054</td>
</tr>
<tr>
<td>ARCA</td>
<td>47.96</td>
<td>1</td>
<td>23,000</td>
<td>48.054</td>
</tr>
<tr>
<td>ARCA</td>
<td>47.96</td>
<td>991</td>
<td>23,955</td>
<td>48.067</td>
</tr>
<tr>
<td>ARCA</td>
<td>47.96</td>
<td>960</td>
<td>24,865</td>
<td>48.010</td>
</tr>
<tr>
<td>ARCA</td>
<td>47.96</td>
<td>940</td>
<td>24,874</td>
<td>48.010</td>
</tr>
<tr>
<td>ARCA</td>
<td>47.96</td>
<td>771</td>
<td>25,445</td>
<td>48.070</td>
</tr>
<tr>
<td>ARCA</td>
<td>47.96</td>
<td>927</td>
<td>26,957</td>
<td>48.114</td>
</tr>
</tbody>
</table>
The worst kept secret in HFT

SELL when imbalance $I = \frac{x}{x+y}$ is small, where $x =$ bid size and $y =$ ask size
Outline

• Why bid and ask sizes matter:
  • Forecasting Prices from Level-I Quotes in the Presence of Hidden Liquidity, with M. Avellaneda and J. Reed
  • Modeling bid and ask sizes
  • \( P(\text{up}) \): the probability that the price will move up
  • The imbalance: \( I = \frac{x}{x+y} \)
Outline

• Why bid and ask sizes matter:
  • Forecasting Prices from Level-I Quotes in the Presence of Hidden Liquidity, with M. Avellaneda and J. Reed
  • Modeling bid and ask sizes
  • \( P(\text{up}) \): the probability that the price will move up
  • The imbalance: \( I = \frac{x}{x+y} \)

• Why latency matters:
  • Optimal Asset Liquidation using Limit Order Book Information, with R. Waeber
  • Modeling latency
  • Optimal liquidation time
  • Trade regions
Modeling Level I quotes

Assume the bid-ask spread is 1 tick
One of the following must happen first:

1. The ask queue is depleted and the price “moves up”.
2. The bid queue is depleted and the price “moves down”.

![Diagram showing bid and ask prices with bid size and ask size arrows]
Continuous model

- Bid size: $x_t$
- Ask size: $y_t$
Continuous model

- Bid size: \( x_t \)
- Ask size: \( y_t \)
- The process \((x_t, y_t)\) can be approximated by the diffusion

\[
\begin{align*}
    dx_t &= \sigma dW_t \\
    dy_t &= \sigma dZ_t \\
    E(\,dWdZ) &= \rho dt,
\end{align*}
\]
Continuous model

- Bid size: $x_t$
- Ask size: $y_t$
- The process $(x_t, y_t)$ can be approximated by the diffusion

$$dx_t = \sigma dW_t$$
$$dy_t = \sigma dZ_t$$
$$E(dWdZ) = \rho dt,$$

- $\tau_x$ and $\tau_y$ are the times when the sizes hit zero
The diffusion

\[ X_t = \sigma W_t \]
\[ Y_t = \sigma Z_t \]
\[ E(dW_t dZ_t) = \rho dt \]
The partial differential equation

- Let \( u(x, y) = P(\tau_y < \tau_x | x_t = x, y_t = y) \) be the probability that the next price move is up, given the bid and ask sizes.
Let \( u(x, y) = P(\tau_y < \tau_x | x_t = x, y_t = y) \) be the probability that the next price move is up, given the bid and ask sizes. It solves the following PDE:

\[
\sigma^2 (u_{xx} + 2\rho u_{xy} + u_{yy}) = 0, \quad x > 0, \ y > 0,
\]
The partial differential equation

- Let \( u(x, y) = P(\tau_y < \tau_x | x_t = x, y_t = y) \) be the probability that the next price move is up, given the bid and ask sizes.
- It solves the following PDE:
  \[
  \sigma^2 (u_{xx} + 2\rho u_{xy} + u_{yy}) = 0, \quad x > 0, \; y > 0,
  \]
- Boundary conditions
  \[
  u(0, y) = 0, \quad \text{for} \quad y > 0,
  \]
  \[
  u(x, 0) = 1, \quad \text{for} \quad x > 0.
  \]
  The price moves as soon as \( x_t \) or \( y_t \) hit zero.
Solution

**Theorem**

The probability of an upward move in the mid price is given by

\[
    u(x, y) = \frac{1}{2} \left( 1 - \frac{\text{Arctan} \left( \sqrt{\frac{1+\rho}{1-\rho}} \frac{y-x}{y+x} \right)}{\text{Arctan} \left( \sqrt{\frac{1+\rho}{1-\rho}} \right)} \right).
\]  

(1)
Uncorrelated queues ($\rho = 0$)

- Problem

\[ u_{xx} + u_{yy} = 0, \quad x > 0, \ y > 0, \]

and

\[ u(0, y) = 0, \quad \text{for} \quad y > 0, \]
\[ u(x, 0) = 1, \quad \text{for} \quad x > 0. \]
Uncorrelated queues ($\rho = 0$)

- **Problem**
  \[ u_{xx} + u_{yy} = 0, \quad x > 0, \quad y > 0, \]
  and
  \[ u(0, y) = 0, \quad \text{for} \quad y > 0, \]
  \[ u(x, 0) = 1, \quad \text{for} \quad x > 0. \]

- **Solution**
  \[ u(x, y) = \frac{2}{\pi} \arctan \left( \frac{x}{y} \right). \]
Perfectly negatively correlated queues ($\rho = -1$)

- Problem

\[ u_{xx} - 2u_{xy} + u_{yy} = 0, \quad x > 0, \quad y > 0, \]

and

\[ u(0, y) = 0, \quad \text{for} \quad y > 0, \]
\[ u(x, 0) = 1, \quad \text{for} \quad x > 0. \]
Perfectly negatively correlated queues \((\rho = -1)\)

\begin{itemize}
  \item Problem

\[
  u_{xx} - 2u_{xy} + u_{yy} = 0, \quad x > 0, \quad y > 0,
\]

and

\[
  u(0, y) = 0, \quad \text{for} \quad y > 0,
\]

\[
  u(x, 0) = 1, \quad \text{for} \quad x > 0.
\]

  \item Solution

\[
  u(x, y) = \frac{x}{x + y}.
\]
\end{itemize}
The data

- Best bid and ask quotes for tickers QQQQ, XLF, JPM, over the first five trading days in 2010
The data

- Best bid and ask quotes for tickers QQQQ, XLF, JPM, over the first five trading days in 2010
- All tickers are traded on various exchanges (NASDAQ, NYSE and BATS)
The data

- Best bid and ask quotes for tickers QQQQ, XLF, JPM, over the first five trading days in 2010
- All tickers are traded on various exchanges (NASDAQ, NYSE and BATS)
- Consider the perfectly negatively correlated queues model, i.e.

\[ u(x, y) = \frac{x}{x + y} \]
**Data sample**

Obtained from the consolidated quotes of the NYSE-TAQ database, provided by WRDS

<table>
<thead>
<tr>
<th>symbol</th>
<th>date</th>
<th>time</th>
<th>bid</th>
<th>ask</th>
<th>bsize</th>
<th>asize</th>
<th>exchange</th>
</tr>
</thead>
<tbody>
<tr>
<td>QQQQ</td>
<td>2010-01-04</td>
<td>09:30:23</td>
<td>46.32</td>
<td>46.33</td>
<td>258</td>
<td>242</td>
<td>T</td>
</tr>
<tr>
<td>QQQQ</td>
<td>2010-01-04</td>
<td>09:30:23</td>
<td>46.32</td>
<td>46.33</td>
<td>260</td>
<td>242</td>
<td>T</td>
</tr>
<tr>
<td>QQQQ</td>
<td>2010-01-04</td>
<td>09:30:23</td>
<td>46.32</td>
<td>46.33</td>
<td>264</td>
<td>242</td>
<td>T</td>
</tr>
<tr>
<td>QQQQ</td>
<td>2010-01-04</td>
<td>09:30:24</td>
<td>46.32</td>
<td>46.33</td>
<td>210</td>
<td>271</td>
<td>P</td>
</tr>
<tr>
<td>QQQQ</td>
<td>2010-01-04</td>
<td>09:30:24</td>
<td>46.32</td>
<td>46.33</td>
<td>210</td>
<td>271</td>
<td>P</td>
</tr>
<tr>
<td>QQQQ</td>
<td>2010-01-04</td>
<td>09:30:24</td>
<td>46.32</td>
<td>46.33</td>
<td>161</td>
<td>271</td>
<td>P</td>
</tr>
</tbody>
</table>
### Summary statistics

<table>
<thead>
<tr>
<th>Ticker</th>
<th>Exchange</th>
<th>num qt</th>
<th>qt/sec</th>
<th>spread</th>
<th>bsize+asize</th>
<th>price</th>
</tr>
</thead>
<tbody>
<tr>
<td>XLF</td>
<td>NASDAQ</td>
<td>0.7M</td>
<td>7</td>
<td>0.010</td>
<td>8797</td>
<td>15.02</td>
</tr>
<tr>
<td>XLF</td>
<td>NYSE</td>
<td>0.4M</td>
<td>4</td>
<td>0.010</td>
<td>10463</td>
<td>15.01</td>
</tr>
<tr>
<td>XLF</td>
<td>BATS</td>
<td>0.4M</td>
<td>4</td>
<td>0.011</td>
<td>7505</td>
<td>14.99</td>
</tr>
<tr>
<td>QQQQ</td>
<td>NASDAQ</td>
<td>2.7M</td>
<td>25</td>
<td>0.010</td>
<td>1455</td>
<td>46.30</td>
</tr>
<tr>
<td>QQQQ</td>
<td>NYSE</td>
<td>4.0M</td>
<td>36</td>
<td>0.011</td>
<td>1152</td>
<td>46.27</td>
</tr>
<tr>
<td>QQQQ</td>
<td>BATS</td>
<td>1.6M</td>
<td>15</td>
<td>0.011</td>
<td>1055</td>
<td>46.28</td>
</tr>
<tr>
<td>JPM</td>
<td>NASDAQ</td>
<td>1.2M</td>
<td>11</td>
<td>0.011</td>
<td>87</td>
<td>43.81</td>
</tr>
<tr>
<td>JPM</td>
<td>NYSE</td>
<td>0.7M</td>
<td>6</td>
<td>0.012</td>
<td>47</td>
<td>43.77</td>
</tr>
<tr>
<td>JPM</td>
<td>BATS</td>
<td>0.6M</td>
<td>5</td>
<td>0.014</td>
<td>39</td>
<td>43.82</td>
</tr>
</tbody>
</table>

**Table**: Summary statistics
Estimation procedure

1. We filter the data set by exchange and ticker
Estimation procedure

1. We filter the data set by exchange and ticker
2. We “bucket” the imbalance

\[ I = \frac{x}{x + y} \]

into intervals: \( 0 < I \leq 0.05 \), \( 0.05 < I \leq 0.1 \), etc..
Estimation procedure

1. We filter the data set by exchange and ticker
2. We “bucket” the imbalance
   \[ I = \frac{x}{x+y} \]
   into intervals: 0 < I ≤ 0.05, 0.05 < I ≤ 0.1, etc..
3. For each bucket, we compute the empirical probability that the price goes up at the next price move, \( u(I) \).
Estimation procedure

1. We filter the data set by exchange and ticker
2. We “bucket” the imbalance
   \[ I = \frac{x}{x + y} \]
   into intervals: 0 < \( I \leq 0.05 \), 0.05 < \( I \leq 0.1 \), etc..
3. For each bucket, we compute the empirical probability that the price goes up at the next price move, \( u(I) \).
4. We plot the probability that the next price move is up, conditional on the imbalance.
Probability of an upward move

as a function of the imbalance
Empirically, the probability of the price going up when the ask size is small BUT does not tend to zero.
Hidden liquidity

- Empirically, the probability of the price going up when the ask size is small BUT does not tend to zero.
- Orders on other exchanges prevent the price from moving up (REG NMS)
Hidden liquidity

- Empirically, the probability of the price going up when the ask size is small BUT does not tend to zero.
- Orders on other exchanges prevent the price from moving up (REG NMS)
- Hidden orders, iceberg orders, dark pools
Boundary condition

- We model a fixed hidden liquidity \( H \)
Boundary condition

- We model a fixed hidden liquidity $H$
- This translates in

$$\sigma^2 (p_{xx} + 2 \rho p_{xy} + p_{yy}) = 0, \quad x > -H, \ y > -H,$$

with the boundary condition

$$p(-H, y) = 0, \quad \text{for} \quad y > -H,$$

$$p(x, -H) = 1, \quad \text{for} \quad x > -H.$$
We model a fixed hidden liquidity $H$
This translates in
$$\sigma^2 (p_{xx} + 2\rho p_{xy} + p_{yy}) = 0, \quad x > -H, \ y > -H,$$
with the boundary condition
$$p(-H, y) = 0, \quad \text{for} \quad y > -H,$$
$$p(x, -H) = 1, \quad \text{for} \quad x > -H.$$
Boundary condition

- We model a fixed hidden liquidity $H$
- This translates in
  \[ \sigma^2 (p_{xx} + 2\rho p_{xy} + p_{yy}) = 0, \quad x > -H, \, y > -H, \]
  with the boundary condition
  \[
  p(-H, y) = 0, \quad \text{for} \quad y > -H, \\
  p(x, -H) = 1, \quad \text{for} \quad x > -H.
  \]
- In other words we can solve the problem with boundary conditions at zero and use the relation
  \[ p(x, y; H) = u(x + H, y + H) \]
Perfectly negatively correlated queues \((\rho = -1)\)

Solution

\[
p(x, y; H) = \frac{x + H}{x + y + 2H}.
\]
Estimation procedure

1. We filter the data set by exchange and ticker
Estimation procedure

1. We filter the data set by exchange and ticker
2. We “bucket” the imbalance in the intervals [0, 0.05), [0.05, 0.1), etc...
## Estimation procedure

1. We filter the data set by exchange and ticker
2. We “bucket” the imbalance in the intervals \([0, 0.05), [0.05, 0.1), \text{ etc...}\)
3. For each bucket, we compute the empirical probability that the price goes up \(\hat{p}(I)\).
Estimation procedure

1. We filter the data set by exchange and ticker
2. We “bucket” the imbalance in the intervals \([0, 0.05), [0.05, 0.1), \text{etc...}\]
3. For each bucket, we compute the empirical probability that the price goes up \(\hat{p}(I)\).
4. The correlation \(-1\) model predicts

\[
p = \frac{x + H}{x + y + 2H} = \frac{I + \frac{H}{x+y}}{1 + 2\frac{H}{x+y}} = \frac{I + h}{1 + 2h} = \frac{1}{1 + 2h}(I-0.5)+0.5
\]

where \(h\) is the normalized hidden size
Estimation procedure

1. We filter the data set by exchange and ticker
2. We “bucket” the imbalance in the intervals $[0, 0.05)$, $[0.05, 0.1)$, etc...
3. For each bucket, we compute the empirical probability that the price goes up $\hat{p}(I)$.
4. The correlation $-1$ model predicts
   \[
   p = \frac{x + H}{x + y + 2H} = \frac{l + \frac{H}{x+y}}{1 + 2\frac{H}{x+y}} = \frac{l + h}{1 + 2h} = \frac{1}{1 + 2h}(l - 0.5) + 0.5
   \]
   where $h$ is the normalized hidden size
5. We regress
   \[
   \hat{p}(I) - 0.5 = \beta(I - 0.5) + \epsilon
   \]
   and obtain an implied hidden liquidity $h = 0.5(1/\beta - 1)$ for each exchange.
Probability of an upward move

hidden size = 0.09
**Results**

The hidden size tells us of how informative the level 1 quotes are.

<table>
<thead>
<tr>
<th>Ticker</th>
<th>NYSE</th>
<th>BATS</th>
<th>NASDAQ</th>
</tr>
</thead>
<tbody>
<tr>
<td>XLF</td>
<td>0.21</td>
<td>0.09</td>
<td>0.09</td>
</tr>
<tr>
<td>QQQQ</td>
<td>0.42</td>
<td>0.39</td>
<td>0.28</td>
</tr>
<tr>
<td>JPM</td>
<td>0.32</td>
<td>0.23</td>
<td>0.18</td>
</tr>
</tbody>
</table>

*Table*: Implied hidden liquidity across tickers and exchanges

But this information decays with latency.
Latency

- Latency arises in every trade execution:
  1. Time of datafeed to travel from exchange to execution machine;
  2. The algorithm making a decision;
  3. The order being sent back to the market.

- We assume there is a fixed latency $L$

- What you see is not what you get
The cost of latency

There is empirical evidence that selling on small imbalances can be profitable:

- On each quote $i$, record the imbalance $I_i$ and the bid price $S^b_i$
- At a later quote in the future $j$, $L$ milliseconds later, record the bid price $S^b_j$
- Take averages of $(S^b_j - S^b_i)$ for $I_i$ in different buckets
Cost as a fraction of the spread

x axis is time, y axis is cost. Each graph is an imbalance decile.
The cost of latency

Transaction cost for $L=1\text{ms}$, $10\text{ms}$, $100\text{ms}$, $1000\text{ms}$
The Optimal Liquidation Problem

The imbalance process $I_t$ is a Markov process.

- **Goal:** Identify an optimal time $\tau$ in $[0, T - L]$ to sell the share at the bid price, i.e.,

$$V(T, L, x) = \sup_{0 \leq \tau \leq T - L} E[P_{\tau + L}^b - P_0^b | I_0 = x],$$

for $x \in [0, 1]$. 
Modeling the imbalance

- $I(n)$ for $0 \leq n \leq N$ is a finite state Markov process.
- 20 transient states, $(0, 0.05]$, $(0.05, 0.1]$, etc...
- Assume the bid ask spread equals the tick size
- Assume that in a $dt = 10$ second time step, the price can only go up or down by one tick
- We estimate the transition probabilities $p_{ij}^{up}$, $p_{ij}^{down}$ and $p_{ij}^{stay}$ empirically
Dynamic Program

- We estimate the payoff function $G^L(i) = E[P^b_l - P^b_0|l(0) = i]$
- Bellman’s recursion:
  $$V^L(n, i) = \max \left\{ G^L(i), E[V^L(n + 1, l(n + 1))|l(n) = i] \right\},$$
- Conditional probability:
  $$E[V^L(n + 1, l(n + 1))|l(n) = i] = \sum_{k=1}^{20} p^{stay}_{ik} V^L(n + 1, k)$$
  $$+ \sum_{k=1}^{20} p^{up}_{ik} (V^L(n + 1, k) + 1) + \sum_{k=1}^{20} p^{down}_{ik} (V^L(n + 1, k) - 1)$$
Trade/no Trade Regions

Define

\[ D = \left\{ (t, x) \in [0, T] \times [0, 1) : V(t, x) = G^L(x) \right\}, \]

\[ C = \left\{ (t, x) \in [0, T] \times [0, 1) : V(t, x) < G^L(x) \right\}. \]

**Proposition**

Fix \( t \in [0, T], x \in [0, 1], \) then \( V^L(t, x) \) is decreasing in \( L \) for \( L \in [0, T]. \)
Trade Regions on 4 different days
The value function for $T = 5$ min
Backtesting the Trade Regions

1. The value function

\[ V(T, L) = \frac{1}{20} \sum_{i=1}^{20} V(T, L, i) \]

2. Out of sample performance

\[ \hat{V}(T, L) = IMB(T, L) - TWAP(T) = \frac{1}{n_T} \sum_{i=1}^{n_T} \frac{P^b(\tau_i + L) - P^b(T_i)}{\Delta}, \]
Backtesting the Trade Regions for $T = 5$ min
Conclusion

1. We can estimate the probability of the next price move:
   - Conditional on the bid and ask sizes
   - Conditional on imbalance if the sizes are negatively correlated
   - Conditional on hidden liquidity for a ticker/exchange pair

2. We can estimate the cost of latency:
   - By solving an optimal stopping problem
   - Backtesting trade/no trade regions on level I data