Forecasting prices from level-I quotes in the presence of hidden liquidity

S. Stoikov, M. Avellaneda and J. Reed

December 5, 2011

Background

- Automated or computerized trading
  - Accounts for 70% of equity trades taking place in the US
  - U.S. Securities and Exchange Commission (SEC) authorized electronic exchanges in 1998
  - Archipelago-Arca-NYSE, Island-Instinet-Inet-NASDAQ, BATS, CME, Tokyo stock exchange, Eurex, London stock exchange

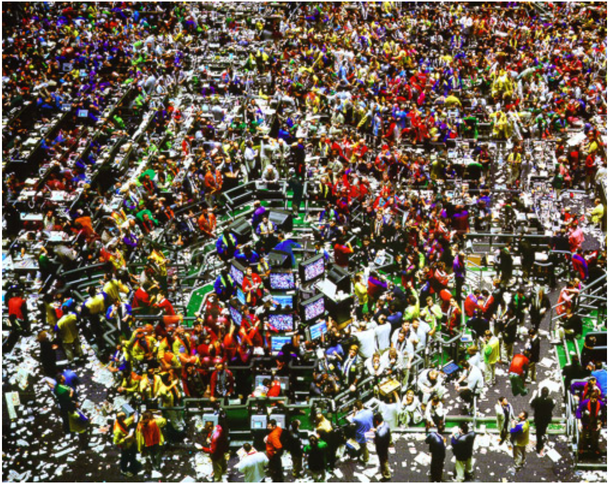
Background

- Automated or computerized trading
  - Accounts for 70% of equity trades taking place in the US
  - U.S. Securities and Exchange Commission (SEC) authorized electronic exchanges in 1998
  - Archipelago-Arca-NYSE, Island-Instinet-Inet-NASDAQ, BATS, CME, Tokyo stock exchange, Eurex, London stock exchange
- Algorithmic trading
  - Brokers executing client transactions
  - "Optimally" splitting of client orders

Background

- Automated or computerized trading
  - Accounts for 70% of equity trades taking place in the US
  - U.S. Securities and Exchange Commission (SEC) authorized electronic exchanges in 1998
  - Archipelago-Arca-NYSE, Island-Instinet-Inet-NASDAQ, BATS, CME, Tokyo stock exchange, Eurex, London stock exchange
- Algorithmic trading
  - Brokers executing client transactions
  - "Optimally" splitting of client orders
- High frequency trading
  - Computerized trading strategies characterized by extremely short position-holding periods
  - Market-making
  - Flash crash!

Market in the 90s



Market today

Bid					Ask				
MM Name	Price	Size	Cum Size	Avg Price	MM Name	Price	Size	Cum Size	Avg Price
NSDQ	47.96	68	68	47.960	NSDQ	47.97	1,281	1,281	47.970
NSX	47.96	2	70	47.960	EDGEA	47.97	243	1,524	47.970
BATS	47.96	12	82	47.960	CHX	47.97	58	1,582	47.970
DRCTEDGE	47.96	1	83	47.960	CBSX	47.97	20	1,602	47.970
ARCA	47.96	128	211	47.960	NSX	47.97	112	1,714	47.970
NSDQ	47.95	906	1,117	47.952	BEX	47.97	359	2,073	47.970
EDGEA	47.95	123	1,240	47.952	ARCA	47.97	1,127	3,200	47.970
CHX	47.95	58	1,298	47.952	BATS	47.97	1,241	4,441	47.970
CBSX	47.95	35	1,333	47.952	DRCTEDGE	47.97	424	4,865	47.970
BEX	47.95	152	1,485	47.951	NSDQ	47.98	1,649	6,514	47.973
ARCA	47.95	858	2,343	47.951	ARCA	47.98	1,376	7,890	47.974
NSDQ	47.94	1,626	3,969	47.946	NSDQ	47.99	1,562	9,452	47.977
ARCA	47.94	1,314	5,283	47.945	ARCA	47.99	1,348	10,800	47.978
NSDQ	47.93	1,550	6,833	47.941	NSDQ	48.00	1,448	12,248	47.981
ARCA	47.93	1,313	8,146	47.940	ARCA	48.00	1,285	13,533	47.983
TMBR	47.92	10	8,156	47.940	NSDQ	48.01	1,494	15,027	47.985
NSDQ	47.92	1,473	9,629	47.937	ARCA	48.01	1,241	16,268	47.987
ARCA	47.92	1,201	10,830	47.935	NSDQ	48.02	1,323	17,591	47.990
UBSS	47.91	1	10,831	47.935	NSDQ	48.03	1,322	18,913	47.992
HDSN	47.91	1	10,832	47.935	NSDQ	48.04	1,061	19,974	47.995
NSDQ	47.91	1,504	12,336	47.932	TMBR	48.05	10	19,984	47.995
NSDQ	47.90	1,362	13,698	47.929	UBSS	48.05	5	19,989	47.995
NSDQ	47.89	1,384	15,082	47.925	NSDQ	48.05	1,022	21,011	47.998
NSDQ	47.88	1,177	16,259	47.922	HDSN	48.05	1	21,012	47.998
NSDQ	47.87	934	17,193	47.919	NSDQ	48.06	965	21,977	48.000
NSDQ	47.86	923	18,116	47.916	NSDQ	48.07	1,043	23,020	48.004
UBSS	47.85	10	18,126	47.916	UBSS	48.08	4	23,024	48.004
NSDQ	47.85	882	19,008	47.913	NSDQ	48.08	901	23,925	48.007
NSDQ	47.84	940	19,948	47.909	NSDQ	48.09	940	24,865	48.010
NSDQ	47.83	800	20,748	47.906	UBSS	48.10	9	24,874	48.010
UBSS	47.82	40	20,788	47.906	NSDQ	48.10	571	25,445	48.012
NSDQ	47.82	520	21,308	47.904	NSDQ	48.11	482	25,927	48.014

This is often referred to as “the order book”

A simplified view of the trading world

Agent	Type of decision	Data
Mutual/hedge fund	Investment	Daily close prices
Banks, brokers	Order splitting	5 min prices
Algorithms, HFT	Market vs. limit, order routing	Level I trades and quotes
Electronic market	Order matching, messenging	Level II trades and quotes

Literature

- Price impact and optimal execution
  - Almgren and Chriss (2000)
  - Schied and Schoneborn (2007)
  - Bouchaud (2009)

Literature

- Price impact and optimal execution
  - Almgren and Chriss (2000)
  - Schied and Schoneborn (2007)
  - Bouchaud (2009)
- Market microstructure and the information content of the order book
  - Hasbrouck (1993)
  - Parlour and Seppi (2008)
  - Hellstroem and Simonsen (2009)
  - Cao, Hansch and Wang (2009)

Literature

- Price impact and optimal execution
  - Almgren and Chriss (2000)
  - Schied and Schoneborn (2007)
  - Bouchaud (2009)
- Market microstructure and the information content of the order book
  - Hasbrouck (1993)
  - Parlour and Seppi (2008)
  - Hellstroem and Simonsen (2009)
  - Cao, Hansch and Wang (2009)
- Limit order book models, zero-intelligence
  - Smith, Farmer, Gillemot, and Krishnamurthy (2003)
  - Cont, Stoikov and Talreja (2010)
  - Cont, De Larrard (2011)

Model objectives

- Making short term price predictions
  1. Given the best bid/ask quotes
  2. Given statistics on the arrival rates of orders
  3. Given a single *hidden liquidity* parameter

Model objectives

- Making short term price predictions
  1. Given the best bid/ask quotes
  2. Given statistics on the arrival rates of orders
  3. Given a single *hidden liquidity* parameter
- Improving the micro-price or "fair" price

$$p_{micro} = p_{bid}\left(\frac{q_{ask}}{q_{ask}+q_{bid}}\right) + p_{ask}\left(\frac{q_{bid}}{q_{ask}+q_{bid}}\right)$$

Model objectives

- Making short term price predictions
  1. Given the best bid/ask quotes
  2. Given statistics on the arrival rates of orders
  3. Given a single *hidden liquidity* parameter
- Improving the micro-price or "fair" price

$$p_{micro} = p_{bid}\left(\frac{q_{ask}}{q_{ask}+q_{bid}}\right) + p_{ask}\left(\frac{q_{bid}}{q_{ask}+q_{bid}}\right)$$

- Comparing the quality of various exchanges

Model objectives

- Making short term price predictions
  1. Given the best bid/ask quotes
  2. Given statistics on the arrival rates of orders
  3. Given a single *hidden liquidity* parameter
- Improving the micro-price or "fair" price

$$p_{micro} = p_{bid}\left(\frac{q_{ask}}{q_{ask}+q_{bid}}\right) + p_{ask}\left(\frac{q_{bid}}{q_{ask}+q_{bid}}\right)$$

- Comparing the quality of various exchanges
- Estimating hidden liquidity

Outline

1. The discrete model
   - A queuing model for level 1 quotes
   - The probability of an upward move in price

Outline

1. The discrete model
   - A queuing model for level 1 quotes
   - The probability of an upward move in price
2. The diffusion limit
   - Diffusion approximation
   - Hidden liquidity and boundary conditions
   - Closed form solution

Outline

1. The discrete model
    - A queuing model for level 1 quotes
    - The probability of an upward move in price
2. The diffusion limit
    - Diffusion approximation
    - Hidden liquidity and boundary conditions
    - Closed form solution
3. Data analysis
    - Trades and quotes (TAQ) data
    - Estimating hidden liquidity

Outline

1. The discrete model
   - A queuing model for level 1 quotes
   - The probability of an upward move in price
2. The diffusion limit
   - Diffusion approximation
   - Hidden liquidity and boundary conditions
   - Closed form solution
3. Data analysis
   - Trades and quotes (TAQ) data
   - Estimating hidden liquidity
4. Conclusion

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”.

A continuous-time Markov chain

Let (X_t, Y_t) be the bid and ask sizes.
Changes in the bid and ask sizes occur at exponential times with rates:

$$\begin{aligned}
\lambda &= \text{arrival rate of orders at the ask (bid)} \\
\mu &= \text{departure rate of orders at the ask (bid)} \\
\eta &= \text{rate of simultaneous arrival at the bid (ask)} \\
&\quad\ \text{and departure at the ask (bid)} \\
h &= \text{minimum order size}
\end{aligned}$$

Infinitesimal means and variances

$$
\begin{aligned}
E\left[X_{t+\Delta t} - X_t | X_t, Y_t\right] &= h\left(\lambda - \mu\right)\Delta t + o(\Delta t) \\
E\left[Y_{t+\Delta t} - Y_t | X_t, Y_t\right] &= h\left(\lambda - \mu\right)\Delta t + o(\Delta t) \\
E\left[(X_{t+\Delta t} - X_t)^2 | X_t, Y_t\right] &= h^2\left(\lambda + \mu + 2\eta\right)\Delta t + o(\Delta t) \\
E\left[(Y_{t+\Delta t} - Y_t)^2 | X_t, Y_t\right] &= h^2\left(\lambda + \mu + 2\eta\right)\Delta t + o(\Delta t) \\
E\left[(X_{t+\Delta t} - X_t)(Y_{t+\Delta t} - Y_t) | X_t, Y_t\right] &= h^2\left(2\eta\right)\Delta t + o(\Delta t).
\end{aligned}
$$

If $\lambda = \mu$, drifts and the variances of the queue sizes are given by

$$
\begin{aligned}
m_X &= m_Y &= 0 \\
\sigma_X^2 &= \sigma_Y^2 &= 2h^2\left(\lambda + \eta\right) \\
\rho & &= \frac{-\eta}{\lambda + \eta}
\end{aligned}
$$

The probability of an upward move in price

- τ_X is the first time the bid size hits zero
- τ_Y is the first time the ask size hits zero
- The probability that the price moves up before it moves down

$$Prob.\{\Delta P > 0 \,|\, X_t, Y_t\} = Prob.\{\tau_Y < \tau_X \,|\, X_t, Y_t\} = p(X_t, Y_t)$$

- This probability may be computed using Laplace transform methods (see Cont. et al. (2010))
- Here we will look at the diffusion limit.

Continuous limit

- Assume that the average queue sizes are much larger than the minimum size $< X >=< Y >\gg h$
- Assume that the frequency of orders per unit time is high, $\lambda, \eta \gg 1$.
- Define the coarse-grained variables

$$x = X/ < X >, \quad y = Y/ < Y >,$$

$$\sigma^2 = \frac{2h^2 (\lambda + \eta)}{< X >^2},$$

Continuous limit

- Assume that the average queue sizes are much larger than the minimum size $< X >=< Y >\gg h$
- Assume that the frequency of orders per unit time is high, $\lambda, \eta \gg 1$.
- Define the coarse-grained variables

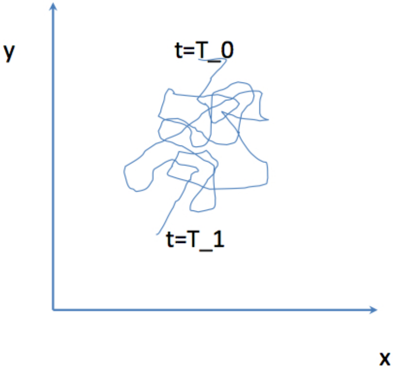
$$x = X / < X >, \quad y = Y / < Y >,$$

$$\sigma^2 = \frac{2h^2 (\lambda + \eta)}{< X >^2},$$

- The process (x_t, y_t) can be approximated by the diffusion

$$\begin{aligned} dx_t &= \sigma dW_t \\ dy_t &= \sigma dZ_t \\ E(dWdZ) &= \rho dt, \end{aligned}$$

The diffusion limit



X= bid size
Y = ask size

$$X_t = \sigma W_t$$
$$Y_t = \sigma Z_t$$
$$E(dW_t dZ_t) = \rho dt$$

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.

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:

$$\sigma^2 \left(u_{xx} + 2\rho u_{xy} + u_{yy}\right) = 0, \quad x > 0, \; y > 0,$$

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:

$$\sigma^2 \left(u_{xx} + 2\rho u_{xy} + u_{yy} \right) = 0, \quad x > 0, \; y > 0,$$

- Boundary conditions

$$\begin{aligned} u(0,y) &= 0, \quad \text{for} \quad y > 0, \\ u(x,0) &= 1, \quad \text{for} \quad x > 0. \end{aligned}$$

The price moves as soon as x_t or y_t hit zero

Hidden liquidity

- Empirically, the probability of the price going up when the ask size is small does not tend to zero.

Hidden liquidity

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

Boundary condition

- We model a fixed hidden liquidity H

Boundary condition

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

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

with the boundary condition

$$\begin{aligned} p(-H, y) &= 0, \quad \text{for} \quad y > -H, \\ p(x, -H) &= 1, \quad \text{for} \quad x > -H. \end{aligned}$$

Boundary condition

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

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

with the boundary condition

$$\begin{aligned} p(-H, y) &= 0, \quad \text{for} \quad y > -H, \\ p(x, -H) &= 1, \quad \text{for} \quad x > -H. \end{aligned}$$

Boundary condition

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

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

with the boundary condition

$$\begin{aligned} p(-H, y) &= 0, \quad \text{for} \quad y > -H, \\ p(x, -H) &= 1, \quad \text{for} \quad x > -H. \end{aligned}$$

- In other words we can solve the problem with boundary conditions at zero and use the relation

$$p(x, y; H) = u(x + H, y + H)$$

Solution

Theorem

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

$$p(x, y; H) = u(x + H, y + H), \tag{1}$$

where

$$u(x, y) = \frac{1}{2}\left(1 - \frac{\mathit{Arctan}\left(\sqrt{\frac{1+\rho}{1-\rho}}\frac{y-x}{y+x}\right)}{\mathit{Arctan}\left(\sqrt{\frac{1+\rho}{1-\rho}}\right)}\right). \tag{2}$$

Uncorrelated queues ($\rho = 0$)

- Problem

$$p_{xx} + p_{yy} = 0, \quad x > -H, \ y > -H,$$

and

$$\begin{aligned} p(-H, y) &= 0, \quad \text{for} \quad y > -H, \\ p(x, -H) &= 1, \quad \text{for} \quad x > -H. \end{aligned}$$

Uncorrelated queues ($\rho = 0$)

- Problem

$$p_{xx} + p_{yy} = 0, \quad x > -H, \; y > -H,$$

and

$$\begin{aligned} p(-H, y) &= 0, \quad \text{for} \quad y > -H, \\ p(x, -H) &= 1, \quad \text{for} \quad x > -H. \end{aligned}$$

- Solution

$$p(x, y; H) = \frac{2}{\pi} Arctan \left(\frac{x + H}{y + H} \right).$$

Perfectly negatively correlated queues ($\rho = -1$)

- Problem

$$p_{xx} - 2p_{xy} + p_{yy} = 0, \quad x > -H, \; y > -H,$$

and

$$\begin{aligned} p(-H, y) &= 0, \quad \text{for} \quad y > -H, \\ p(x, -H) &= 1, \quad \text{for} \quad x > -H. \end{aligned}$$

Perfectly negatively correlated queues ($\rho = -1$)

- Problem

$$p_{xx} - 2p_{xy} + p_{yy} = 0, \quad x > -H, \ y > -H,$$

and

$$\begin{aligned} p(-H, y) &= 0, \quad \text{for} \quad y > -H, \\ p(x, -H) &= 1, \quad \text{for} \quad x > -H. \end{aligned}$$

- Solution

$$p(x, y; H) = \frac{x + H}{x + y + 2H}.$$

The data

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

The data

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

The data

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

$$p(x, y; H) = \frac{x + H}{x + y + 2H}$$

we obtain the "implied hidden size" for each ticker and exchange.

Data sample

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

symbol	date	time	bid	ask	bsize	asize	exchange
QQQQ	2010-01-04	09:30:23	46.32	46.33	258	242	T
QQQQ	2010-01-04	09:30:23	46.32	46.33	260	242	T
QQQQ	2010-01-04	09:30:23	46.32	46.33	264	242	T
QQQQ	2010-01-04	09:30:24	46.32	46.33	210	271	P
QQQQ	2010-01-04	09:30:24	46.32	46.33	210	271	P
QQQQ	2010-01-04	09:30:24	46.32	46.33	161	271	P

Summary statistics

Ticker	Exchange	num qt	qt/sec	spread	bsize+asize	price
XLF	NASDAQ	0.7M	7	0.010	8797	15.02
XLF	NYSE	0.4M	4	0.010	10463	15.01
XLF	BATS	0.4M	4	0.011	7505	14.99
QQQQ	NASDAQ	2.7M	25	0.010	1455	46.30
QQQQ	NYSE	4.0M	36	0.011	1152	46.27
QQQQ	BATS	1.6M	15	0.011	1055	46.28
JPM	NASDAQ	1.2M	11	0.011	87	43.81
JPM	NYSE	0.7M	6	0.012	47	43.77
JPM	BATS	0.6M	5	0.014	39	43.82
AAPL	NASDAQ	1.3M	13	0.034	9.1	212.50
AAPL	NYSE	0.4M	4	0.046	5.7	212.66
AAPL	BATS	0.6M	6	0.054	4.5	212.43

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 bid and ask sizes in deciles

Estimation procedure

1. We filter the data set by exchange and ticker
2. We "bucket" the bid and ask sizes in deciles
3. For each bucket (i, j), we compute the empirical probability that the price goes up u_{ij}.

Estimation procedure

1. We filter the data set by exchange and ticker
2. We "bucket" the bid and ask sizes in deciles
3. For each bucket (i, j), we compute the empirical probability that the price goes up u_{ij}.
4. We count the number of occurrences of the (i, j) bucket, and denote this distribution d_{ij}.

Estimation procedure

1. We filter the data set by exchange and ticker
2. We "bucket" the bid and ask sizes in deciles
3. For each bucket (i, j), we compute the empirical probability that the price goes up u_{ij}.
4. We count the number of occurrences of the (i, j) bucket, and denote this distribution d_{ij}.
5. We minimize least squares for the negatively correlated queues model, i.e.

$$\min_H \sum_{i,j=1}^{10} \left[\left(u_{ij} - \frac{i+H}{i+j+2H} \right)^2 d_{ij} \right]$$

and obtain an implied hidden liquidity H for each exchange.

Empirical probability (XLF on NASDAQ)

decile	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
0.1	0.50	0.38	0.25	0.25	0.32	0.26	0.23	0.23	0.15
0.2	0.61	0.50	0.47	0.41	0.36	0.40	0.38	0.27	0.20
0.3	0.75	0.53	0.50	0.43	0.39	0.37	0.43	0.39	0.28
0.4	0.74	0.58	0.57	0.50	0.42	0.42	0.47	0.46	0.37
0.5	0.68	0.64	0.61	0.58	0.50	0.51	0.48	0.49	0.41
0.6	0.74	0.60	0.63	0.58	0.49	0.50	0.50	0.50	0.49
0.7	0.78	0.62	0.57	0.53	0.52	0.50	0.50	0.60	0.53
0.8	0.77	0.73	0.61	0.54	0.51	0.50	0.40	0.50	0.42
0.9	0.85	0.79	0.72	0.63	0.60	0.51	0.47	0.57	0.50

Model probabilities (XLF on NASDAQ)

decile	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
0.1	0.50	0.42	0.36	0.31	0.28	0.25	0.23	0.21	0.19
0.2	0.58	0.50	0.44	0.39	0.35	0.32	0.29	0.27	0.25
0.3	0.64	0.56	0.50	0.45	0.41	0.37	0.35	0.32	0.30
0.4	0.69	0.61	0.55	0.50	0.46	0.42	0.39	0.37	0.34
0.5	0.72	0.65	0.59	0.54	0.50	0.46	0.43	0.41	0.38
0.6	0.75	0.68	0.63	0.58	0.54	0.50	0.47	0.44	0.42
0.7	0.77	0.71	0.65	0.61	0.57	0.53	0.50	0.47	0.45
0.8	0.79	0.73	0.68	0.63	0.59	0.56	0.53	0.50	0.47
0.9	0.81	0.75	0.70	0.66	0.62	0.58	0.55	0.53	0.50

Results

Ticker	NASDAQ	NYSE	BATS
XLF	0.15	0.17	0.17
QQQQ	0.21	0.04	0.18
JPM	0.17	0.17	0.11
AAPL $s = 1$	0.16	0.90	0.65
AAPL $s = 2$	0.31	0.60	0.64
AAPL $s = 3$	0.31	0.69	0.63

Table: Implied hidden liquidity across tickers and exchanges

Future research

- Level 2 data, predictions on longer time scales
- Bid ask spreads greater than 1
- High frequency volatility estimation
- Optimal execution with limit and market orders
- More general dynamics for the bid and ask processes