Time is Money:

Estimating the Cost of Latency in Trading

Sasha Stoikov

Cornell University

December 10, 2014

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

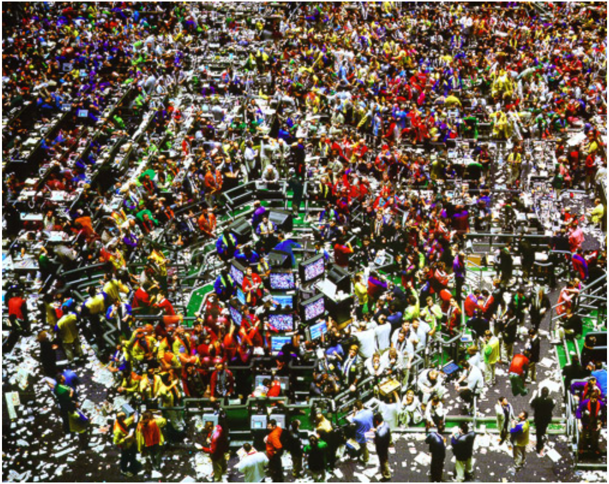
Background

- 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

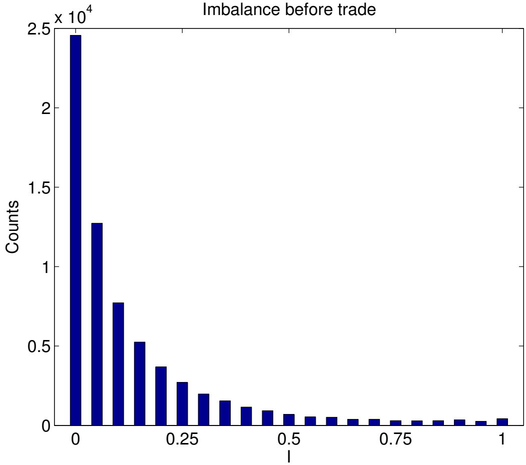


Market today: the order book

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

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(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(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".

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

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

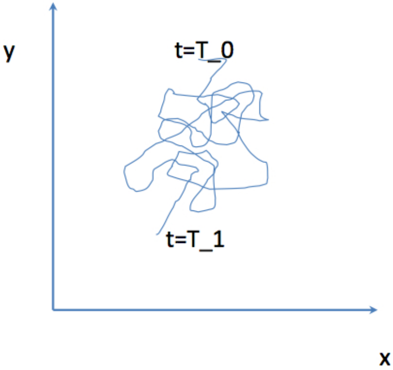
Continuous model

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

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

- τ_x and τ_y are the times when the sizes hit zero

The diffusion



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

Solution

Theorem

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

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

Uncorrelated queues ($\rho = 0$)

- Problem

$$u_{xx} + u_{yy} = 0, \quad x > 0, \; y > 0,$$

and

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

Uncorrelated queues ($\rho = 0$)

- Problem

$$u_{xx} + u_{yy} = 0, \quad x > 0, \; y > 0,$$

and

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

- Solution

$$u(x, y) = \frac{2}{\pi} Arctan \left(\frac{x}{y} \right).$$

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

- Problem

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

and

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

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

- Problem

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

and

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

- Solution

$$u(x, y) = \frac{x}{x + y}.$$

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

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

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: $0 < I \leq 0.05$, $0.05 < I \leq 0.1$, etc..
3. For each bucket, we compute the empirical probability that the price goes up at the next price move, $u(I)$.

Estimation procedure

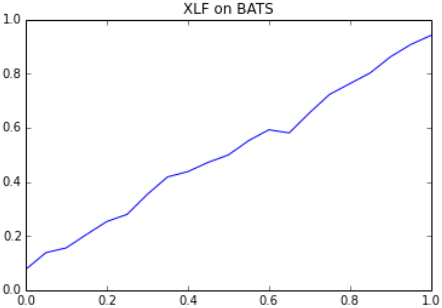
1. We filter the data set by exchange and ticker
2. We "bucket" the imbalance

$$I = \frac{x}{x+y}$$

   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



Hidden liquidity

- 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

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

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

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

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 $[0, 0.05)$, $[0.05, 0.1)$, 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 $[0, 0.05)$, $[0.05, 0.1)$, 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

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

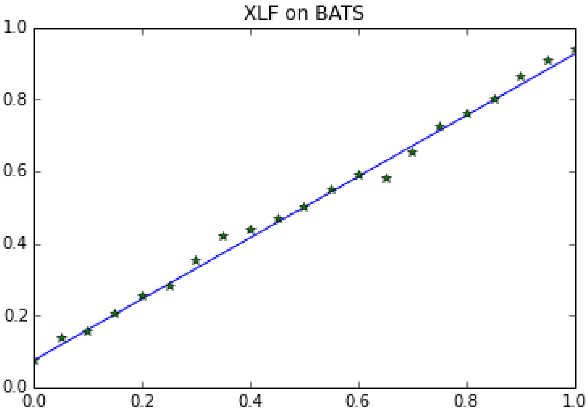
   where h is the normalized hidden size
5. We regress

$$\hat{p}(I) - 0.5 = \beta(I - 0.5) + \epsilon$$

   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 I quotes are.

Ticker	NYSE	BATS	NASDAQ
XLF	0.21	0.09	0.09
QQQQ	0.42	0.39	0.28
JPM	0.32	0.23	0.18

Table: Implied hidden liquidity across tickers and exchanges

But this information decays with latency

Latency

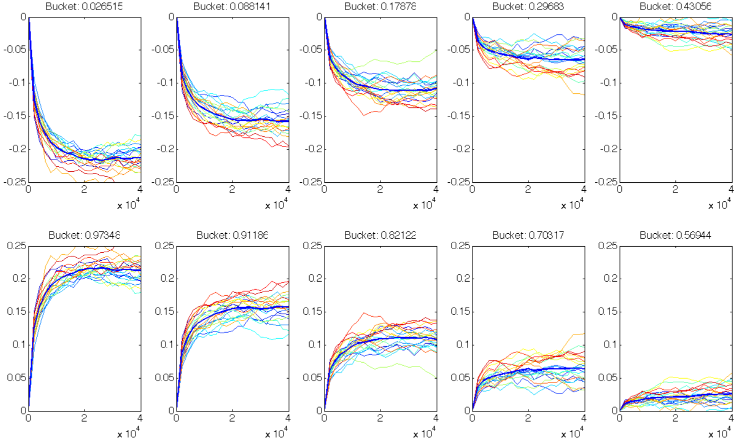
- Latency arises in every trade execution:
  1. Time of datafeed to travel from exchange to execution machine;
  2. The algorithm making a decision;
  3. The order being sent back to the market.
- We assume there is a fixed latency L
- What you see is **not** what you get

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_i^b
- At a later quote in the future j, L milliseconds later, record the bid price S_j^b
- Take averages of $(S_j^b - S_i^b)$ for I_i in different buckets

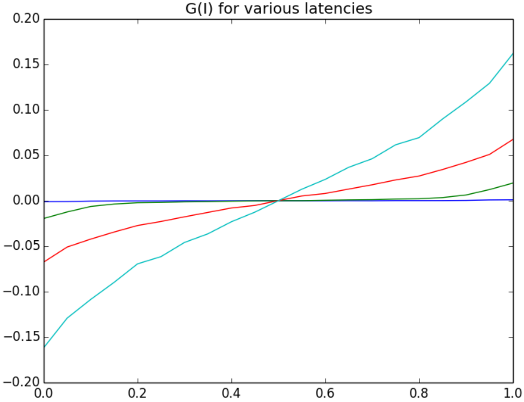
Cost as a fraction of the spread



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

The cost of latency

Transaction cost for L=1ms, 10ms, 100ms, 1000ms



The Optimal Liquidation Problem

The imbalance process I_t is a Markov process.

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

$$V(t,x) = \sup_{0 \le \tau \le T-L} E[P^b_{\tau+L} - P^b_0 | I_t = x],$$

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...
- We estimate the payoff function $G^L(x) = E[P_L^b - P_0^b | I_0 = x]$ for a given latency L and imbalance x
- We estimate the transition probabilities p_{ij}^{up}, p_{ij}^{down} and p_{ij}^{stay} empirically

Dynamic Program

- Bellman's recursion:

$$V^L(n,i) = \max\left\{G^L(i), E[V^L(n+1, I(n+1))|I(n)=i]\right\},$$

- Conditional probability:

$$E[V^L(n+1, I(n+1))|I(n)=i] = \sum_{k=1}^{20} p_{ik}^{stay} V^L(n+1,k)$$

$$+\sum_{k=1}^{20} p_{ik}^{up}(V^L(n+1,k)+1) + \sum_{k=1}^{20} p_{ik}^{down}(V^L(n+1,k)-1)$$

Trade/no Trade Regions

Define

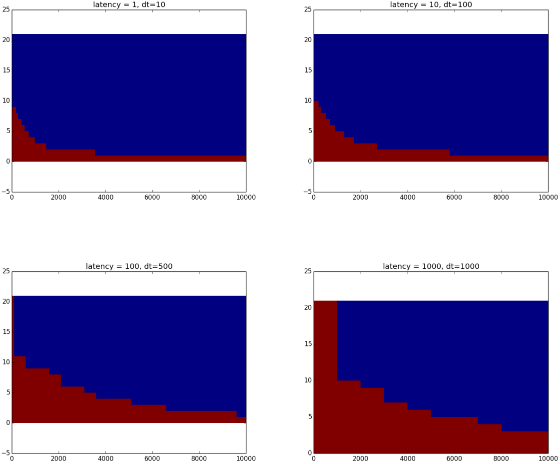
$$D = \left\{(t,x) \in [0,T] \times [0,1) : V(t,x) = G^L(x)\right\},$$

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

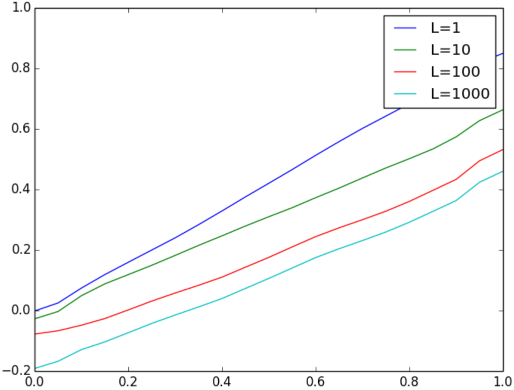
Proposition

Fix $t \in [0,T], x \in [0,1]$, then $V^L(t,x)$ is decreasing in L for $L \in [0,T]$.

Trade/no Trade Regions



The value function



Backtesting the Trade Regions

1. Calibrate the trade region based on 1 day of 5 year US treasuries data
2. Backtest the trade region on 10 out of sample days
3. Results:
   - $P^b_{\tau+L} - P^b_0 = 42\%$ of the bid ask spread, for L=1ms
   - $P^b_{\tau+L} - P^b_0 = 31\%$ of the bid ask spread, for L=10ms
   - $P^b_{\tau+L} - P^b_0 = 20\%$ of the bid ask spread, for L=100ms
   - $P^b_{\tau+L} - P^b_0 = 9\%$ of the bid ask spread, for L=1000ms

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