Volume Imbalance and Algorithmic Trading

Álvaro Cartea

a.cartea@ucl.ac.uk

University College London
joint work with

Ryan Donnelly, EPFL
Sebastian Jaimungal, University of Toronto

December, 2014

Outline

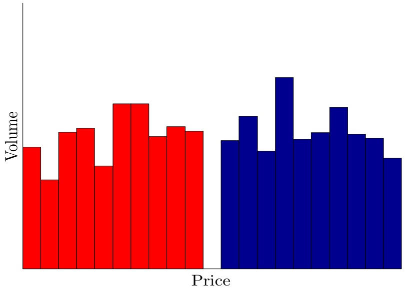
- The limit order book.
- Volume order imbalance as an indicator of market behaviour.
- Imbalance model and market model.
- Optimal trading problem.
- The value of knowing imbalance.

The Limit Order Book

- The limit order book is a record of collective interest to buy or sell certain quantities of an asset at a certain price.

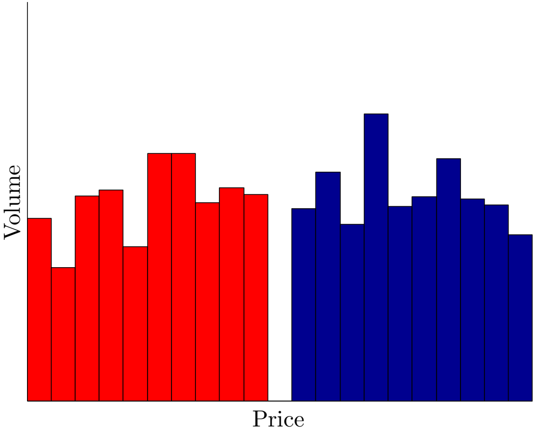
Buy Orders		Sell Orders	
Price	Volume	Price	Volume
60.00	80	60.10	75
59.90	100	60.20	75
59.80	90	60.30	50

- Graphical representation of the limit order book:



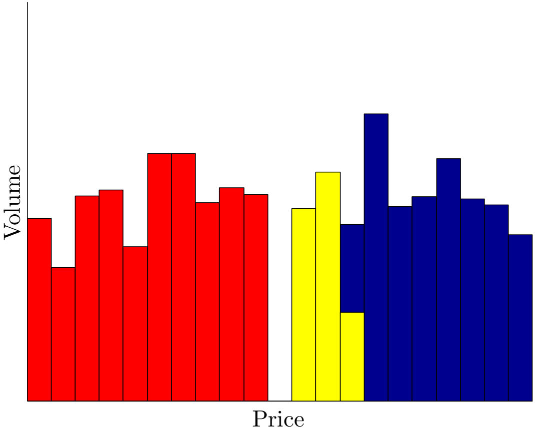
Market Orders

- An incoming market order lifts limit orders from the book.



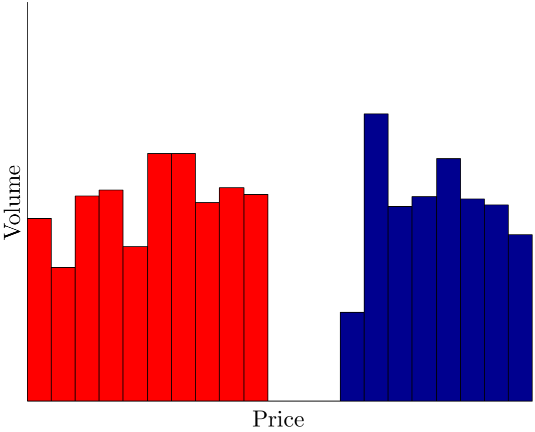
Market Orders

- An incoming market order lifts limit orders from the book.



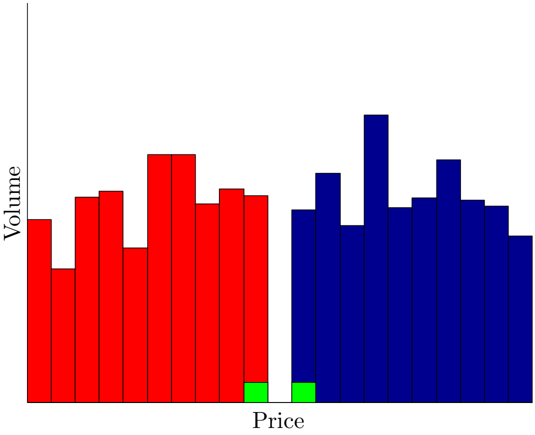
Market Orders

- An incoming market order lifts limit orders from the book.



Agent's Goal

- Optimally place limit orders in the limit order book (LOB)



Agent's Goal

Optimally placing limit orders in the limit order book requires the agent to specify dynamics of the market, namely:

- Dynamics of the midprice.
- Dynamics of the spread.
- Dynamics of incoming market buy and sell orders.
- Interaction between the agent's limit orders and incoming market orders.

Models from previous literature

- Avellaneda and Stoikov (2008): midprice is BM, trades arrive according to Poisson process, exponential fill rate.
- Cartea and Jaimungal (2012): midprice jumps due to market orders, introduce risk control via inventory penalisation.
- Fodra and Labadie (2012): midprice follows a diffusion process with general local drift and volatility terms, Poisson arrivals, exponential fill rate.
- Guilbaud and Pham (2013): discrete spread modelled as Markov chain, independent Levy process midprice, inventory penalisation.
- Guéant, Lehalle, and Fernandez-Tapia (2013): midprice is BM, trades arrive according to Poisson process, exponential fill rate.
- Cartea, Jaimungal, and Ricci (2014): multi-factor mutually-exciting process jointly models arrivals, fill probabilities, and midprice drift.

Volume Order Imbalance

Volume Order Imbalance

- Volume order imbalance is the proportion of best interest on the bid side.
- Defined as:

$$I_t = \frac{V_t^b}{V_t^b + V_t^a}\,.$$

- V_t^b is the volume at the best bid at time t.
- V_t^a is the volume at the best ask at time t.
- $I_t \in [0, 1]$.

Predictive Power of Volume Imbalance - MO type

- Consider the types of market orders that are placed depending on the level of imbalance.
- More market **buys** when imbalance is **high**, more market **sells** when imbalance is **low**.

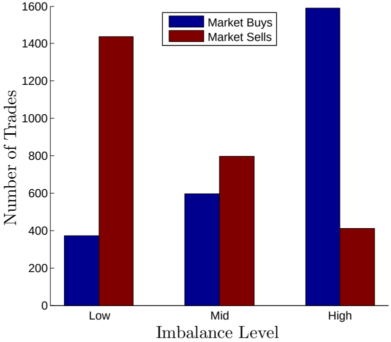


Figure : **BBBY**: one day of NASDAQ trades. Imbalance ranges are $[0, 0.35)$, $[0.35, 0.65]$, and $(0.65, 1]$.

Predictive Power of Volume Imbalance - MO type

- Consider the types of market orders that are placed depending on the level of imbalance.
- More market **buys** when imbalance is **high**, more market **sells** when imbalance is **low**.

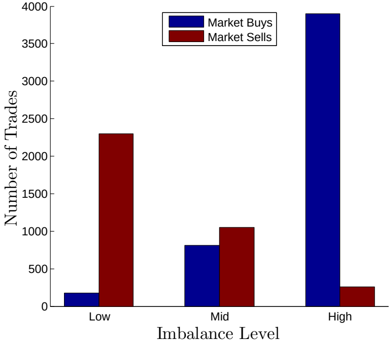


Figure : **MSFT**: one day of NASDAQ trades. Imbalance ranges are $[0, 0.35)$, $[0.35, 0.65]$, and $(0.65, 1]$.

Predictive Power of Volume Imbalance - MO type

- Consider the types of market orders that are placed depending on the level of imbalance.
- More market **buys** when imbalance is **high**, more market **sells** when imbalance is **low**.

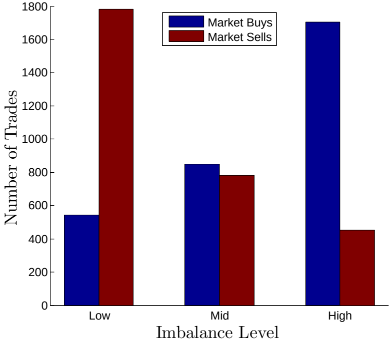


Figure : **TEVA**: one day of NASDAQ trades. Imbalance ranges are $[0, 0.35)$, $[0.35, 0.65]$, and $(0.65, 1]$.

Volume Imbalance and Midprice Change

Predictive Power of Volume Imbalance - Midprice Change

- Distribution of midprice change 20*ms* after a market order.

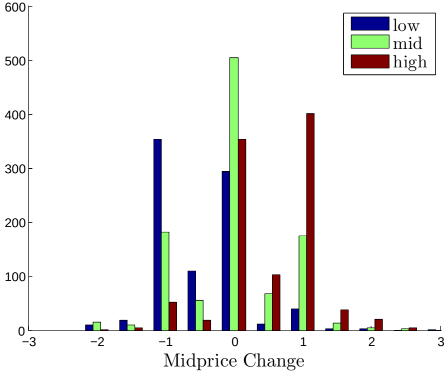


Figure : **BBBY**: one day of NASDAQ trades. Imbalance ranges are $[0, 0.35)$, $[0.35, 0.65]$, and $(0.65, 1]$.

Predictive Power of Volume Imbalance - Midprice Change

- Distribution of midprice change 20*ms* after a market order.

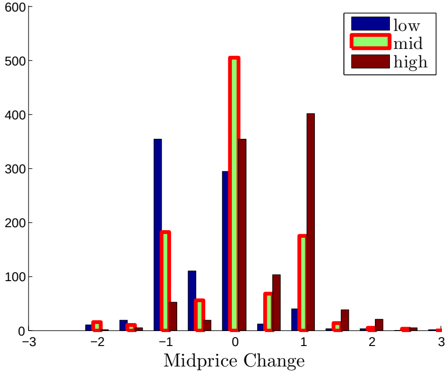


Figure : **BBBY**: one day of NASDAQ trades. Imbalance ranges are $[0, 0.35)$, $[0.35, 0.65]$, and $(0.65, 1]$.

Predictive Power of Volume Imbalance - Midprice Change

- Distribution of midprice change 20*ms* after a market order.

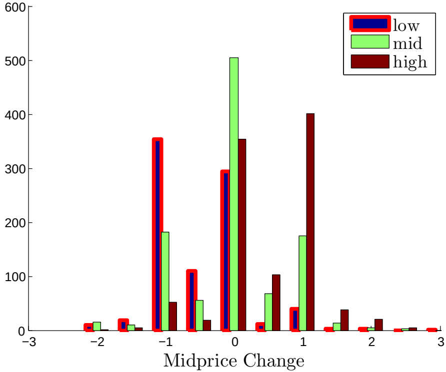


Figure : **BBBY**: one day of NASDAQ trades. Imbalance ranges are $[0, 0.35)$, $[0.35, 0.65]$, and $(0.65, 1]$.

Predictive Power of Volume Imbalance - Midprice Change

- Distribution of midprice change 20*ms* after a market order.

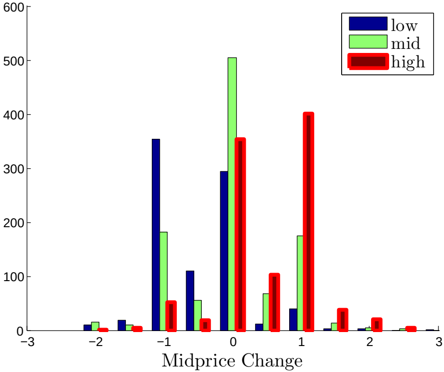


Figure : **BBBY**: one day of NASDAQ trades. Imbalance ranges are $[0, 0.35)$, $[0.35, 0.65]$, and $(0.65, 1]$.

Predictive Power of Volume Imbalance - Midprice Change

- Distribution of midprice change 20*ms* after a market order.

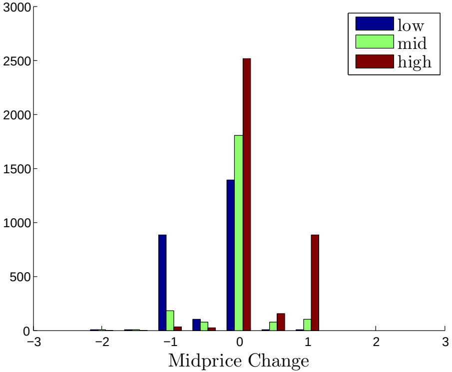


Figure : **MSFT**: one day of NASDAQ trades. Imbalance ranges are $[0, 0.35)$, $[0.35, 0.65]$, and $(0.65, 1]$.

Predictive Power of Volume Imbalance - Midprice Change

- Distribution of midprice change 20*ms* after a market order.

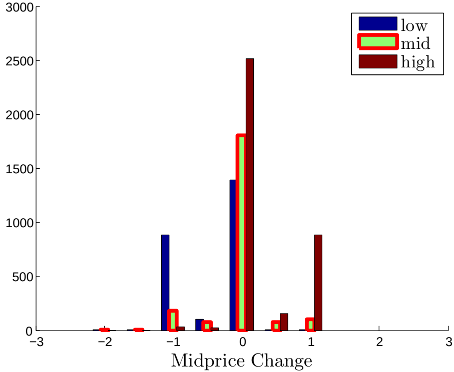


Figure : **MSFT**: one day of NASDAQ trades. Imbalance ranges are $[0, 0.35)$, $[0.35, 0.65]$, and $(0.65, 1]$.

Predictive Power of Volume Imbalance - Midprice Change

- Distribution of midprice change 20*ms* after a market order.

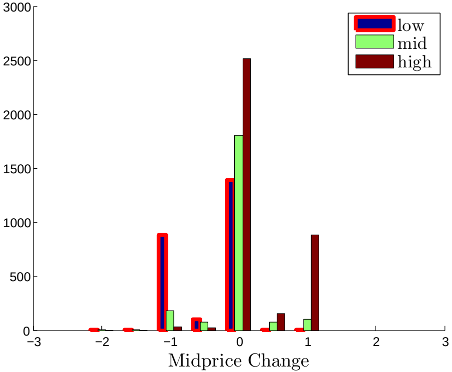


Figure : **MSFT**: one day of NASDAQ trades. Imbalance ranges are $[0, 0.35)$, $[0.35, 0.65]$, and $(0.65, 1]$.

Predictive Power of Volume Imbalance - Midprice Change

- Distribution of midprice change 20*ms* after a market order.

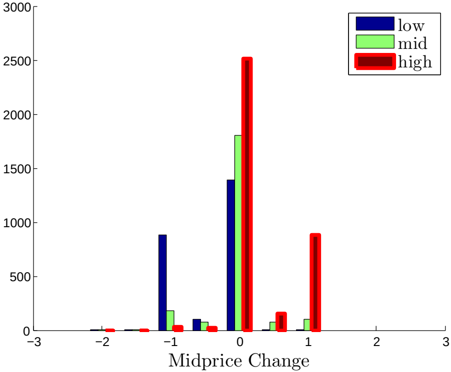


Figure : **MSFT**: one day of NASDAQ trades. Imbalance ranges are $[0, 0.35)$, $[0.35, 0.65]$, and $(0.65, 1]$.

Predictive Power of Volume Imbalance - Midprice Change

- Distribution of midprice change 20*ms* after a market order.

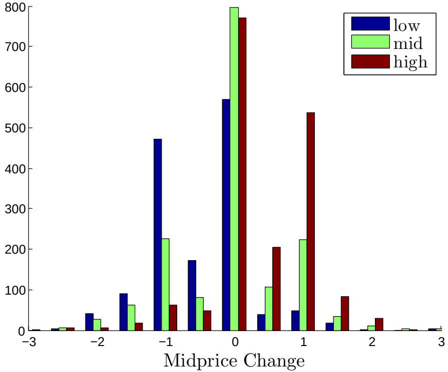


Figure : **TEVA**: one day of NASDAQ trades. Imbalance ranges are $[0, 0.35)$, $[0.35, 0.65]$, and $(0.65, 1]$.

Predictive Power of Volume Imbalance - Midprice Change

- Distribution of midprice change 20*ms* after a market order.

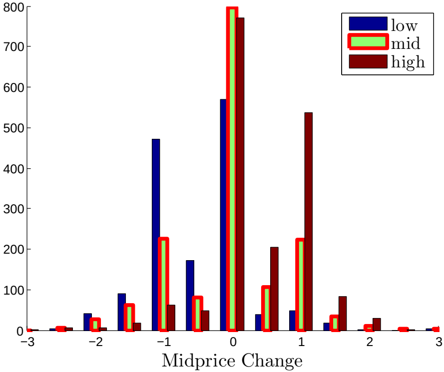


Figure : **TEVA**: one day of NASDAQ trades. Imbalance ranges are $[0, 0.35)$, $[0.35, 0.65]$, and $(0.65, 1]$.

Predictive Power of Volume Imbalance - Midprice Change

- Distribution of midprice change 20*ms* after a market order.

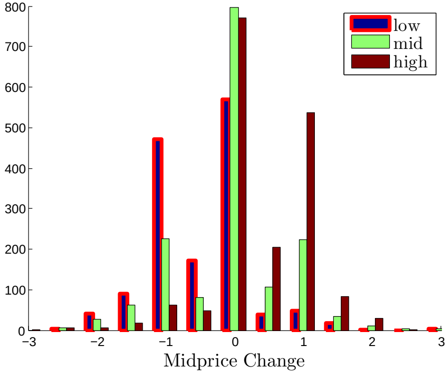


Figure : **TEVA**: one day of NASDAQ trades. Imbalance ranges are $[0, 0.35)$, $[0.35, 0.65]$, and $(0.65, 1]$.

Predictive Power of Volume Imbalance - Midprice Change

- Distribution of midprice change 20*ms* after a market order.

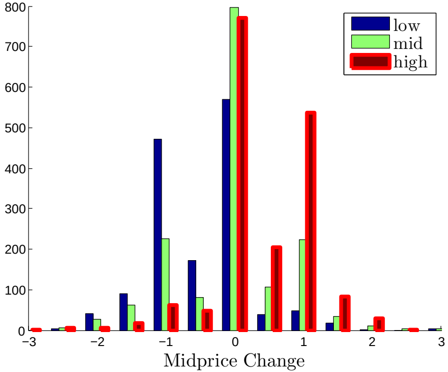


Figure : **TEVA**: one day of NASDAQ trades. Imbalance ranges are $[0, 0.35)$, $[0.35, 0.65]$, and $(0.65, 1]$.

Where to post in the LOB?

Tick Activity

- Number of market orders that take place at ticks from midprice.

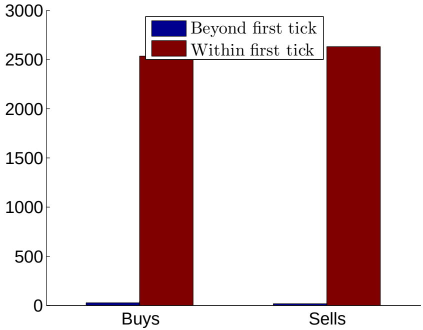


Figure : **BBBY**: one day of NASDAQ trades.

Tick Activity

- Number of market orders that take place at ticks from midprice.

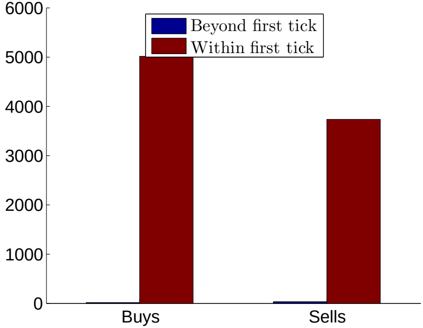


Figure : **MSFT**: one day of NASDAQ trades.

Tick Activity

- Number of market orders that take place at ticks from midprice.

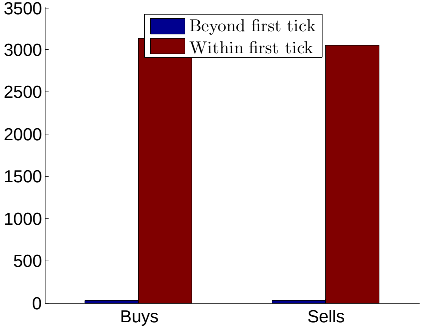


Figure : **TEVA**: one day of NASDAQ trades.

Market Model

Market Model

- Rather than model imbalance directly, a finite state imbalance regime process is considered, $Z_t \in \{1, \ldots, n_Z\}$.
  - Z_t will act as an approximation to the true value of imbalance.
  - The interval $[0, 1]$ is subdivided in to n_Z subintervals. $Z_t = k$ corresponds to I_t lying within the k^{th} subinterval.
- The spread Δ_t also takes values in a finite state space, $\Delta_t \in \{1, \ldots, n_\Delta\}$.

Market Model

- Let μ, μ^+, and μ^- be three doubly stochastic Poisson random measures.
- M_t^+ and M_t^-, the number of market buy and sell orders up to time t, are given by:

$$M_t^{\pm} = \int_0^t \int_{\bar{y}\in\mathbb{R}^3} \mu^{\pm}(d\bar{y}, du)$$

- The midprice, S_t, together with Z_t and Δ_t are modelled as:

$$\begin{aligned}
S_t &= S_0 + \int_0^t \int_{\bar{y}\in\mathbb{R}^3} y_1(\mu + \mu^+ - \mu^-)(d\bar{y}, du) \\
Z_t &= Z_0 + \int_0^t \int_{\bar{y}\in\mathbb{R}^3} (y_2 - Z_{u^-})(\mu + \mu^+ + \mu^-)(d\bar{y}, du) \\
\Delta_t &= \Delta_0 + \int_0^t \int_{\bar{y}\in\mathbb{R}^3} (y_3 - \Delta_{u^-})(\mu + \mu^+ + \mu^-)(d\bar{y}, du)
\end{aligned}$$

Main features of this model

- All three μ^i are conditionally independent given (Z_t, Δ_t) and have compensators of the form:

$$\nu^i(d\bar{y}, dt) = \lambda^i(Z_t, \Delta_t) F^i_{Z_t,\Delta_t}(d\bar{y}) dt$$

- This makes the joint process (Z_t, Δ_t) a continuous-time Markov chain.
- $\lambda^{\pm}(Z, \Delta)$ and $F^{\pm}_{Z,\Delta}(d\bar{y})$ are chosen to reflect realistic dependence between market order arrivals, volume imbalance, spread changes, and midprice changes.
- $F_{Z,\Delta}$ is chosen to have support only on $y_1 = \pm\frac{y_3 - \Delta}{2}$. Limit order activity must change the midprice and spread simultaneously.

Agent's Wealth and Inventory

- The agent may post bid and ask orders at the touch.
- Wealth has dynamics:

$$dX_t = \gamma_t^+ \left(S_{t^-} + \frac{\Delta_{t^-}}{2} \right) dM_t^+ - \gamma_t^- \left(S_{t^-} - \frac{\Delta_{t^-}}{2} \right) dM_t^-$$

  where $\gamma_t^\pm \in \{0, 1\}$ are the agent's control processes.
- Inventory has dynamics:

$$dq_t = -\gamma_t^+ dM_t^+ + \gamma_t^- dM_t^-$$

- Controls $\gamma_t^\pm$ are chosen such that inventory is constrained, $\underline{Q} \leq q_t \leq \overline{Q}$.

Optimal Trading

The Optimal Trading Problem

- The agent attempts to maximize expected terminal wealth:

$$H(t,x,q,S,Z,\Delta) = \sup_{(\gamma_t^{\pm})\in\mathcal{A}} \mathbb{E}\left[X_T + q_T\Big(S_T - \ell(q_T)\Big)\Big|\mathcal{F}_t\right]$$

- This value function has associated equation:

$$\begin{aligned}
&\partial_t H + \lambda(Z,\Delta)\mathbb{E}[\mathcal{D}H|Z,\Delta] \\
&+ \sup_{\gamma^+\in\{0,1\}} \lambda^+(Z,\Delta)\mathbb{E}[\mathcal{D}^+H|Z,\Delta] \\
&+ \sup_{\gamma^-\in\{0,1\}} \lambda^-(Z,\Delta)\mathbb{E}[\mathcal{D}^-H|Z,\Delta] = 0\,, \\
&H(T,x,q,S,Z) = x + q(S-\ell(q))\,.
\end{aligned}$$

Value Function Ansatz

- Making the ansatz $H(t, x, q, S, Z, \Delta) = x + qS + h(t, q, Z, \Delta)$ allows for a corresponding equation for h to be written:

$$
\begin{aligned}
\partial_t h + \lambda(Z,\Delta)(q\epsilon(Z,\Delta) + \Sigma(t,q,Z,\Delta)) & \\
+ \sup_{\gamma^+\in\{0,1\}} \lambda^+(Z,\Delta)\left(\gamma^+\frac{\Delta}{2} + (q-\gamma^+)\epsilon^+(Z,\Delta) + \Sigma^+_{\gamma^+}(t,q,Z,\Delta)\right) & \\
+ \sup_{\gamma^-\in\{0,1\}} \lambda^-(Z,\Delta)\left(\gamma^-\frac{\Delta}{2} - (q+\gamma^-)\epsilon^-(Z,\Delta) + \Sigma^-_{\gamma^-}(t,q,Z,\Delta)\right) & = 0 \\
h(T,q,Z,\Delta) & = -q\ell(q)
\end{aligned}
$$

- This is a system of ODE's of dimension $n_Z n_\Delta(\overline{Q} - \underline{Q} + 1)$.

Feedback Controls

- Feedback controls can be written as:

$$\gamma^{\pm}(t,q,Z,\Delta) \quad = \quad \begin{cases} 1, & \frac{\Delta}{2} - \epsilon^{\pm}(Z,\Delta) + \Sigma_1^{\pm}(t,q,Z,\Delta) > \Sigma_0^{\pm}(t,q,Z,\Delta) \\ 0, & \frac{\Delta}{2} - \epsilon^{\pm}(Z,\Delta) + \Sigma_1^{\pm}(t,q,Z,\Delta) \leq \Sigma_0^{\pm}(t,q,Z,\Delta) \end{cases}$$

where

$$\begin{aligned} \epsilon^{\pm}(Z,\Delta) &= \sum_{y_1,y_2,y_3} y_1 F^{\pm}_{Z,\Delta}(y_1,y_2,y_3) \\ \Sigma^{\pm}_{\gamma^{\pm}}(t,q,Z,\Delta) &= \sum_{y_1,y_2,y_3} \left(h(t,q\mp\gamma^{\pm},y_2,y_3) - h(t,q,Z,\Delta)\right) F^{\pm}_{Z,\Delta}(y_1,y_2,y_3) \end{aligned}$$

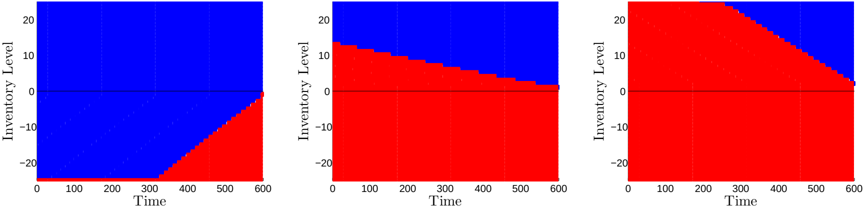
Optimal Trading Strategy – Parameters

- Allow three possible states of imbalance: $Z_t \in \{1, 2, 3\}$
- Two possible spreads: $\Delta_t \in \{1, 2\}$
- MO arrival rates and price impact account for imbalance. In matrices rows are spread ($n_\Delta = 2$) and columns are imbalance states ($n_Z = 3$)

$$\overline{\lambda}^{+} = \begin{pmatrix} 0.050 & 0.091 & 0.242 \\ 0.057 & 0.051 & 0.095 \end{pmatrix} \quad \overline{\varepsilon}^{+} = \begin{pmatrix} 0.247 & 0.556 & 0.710 \\ 0.539 & 0.959 & 1.036 \end{pmatrix}$$
$$\overline{\lambda}^{-} = \begin{pmatrix} 0.242 & 0.091 & 0.050 \\ 0.095 & 0.051 & 0.057 \end{pmatrix} \quad \overline{\varepsilon}^{-} = \begin{pmatrix} 0.710 & 0.556 & 0.247 \\ 1.036 & 0.959 & 0.539 \end{pmatrix}$$

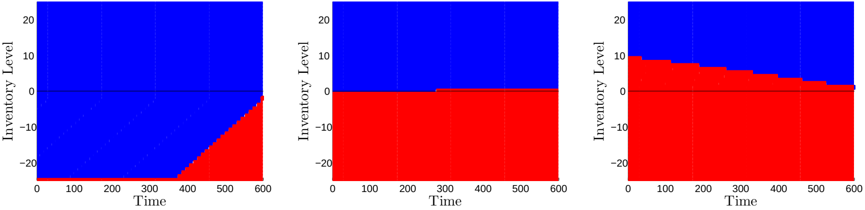
- Terminal penalty function chosen to be $\ell(q, \Delta) = 0.005q$.

Optimal Trading Strategy - Limit Sell Orders



Low Imbalance | Middle Imbalance | High Imbalance

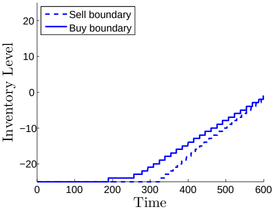
$\Delta = 1$



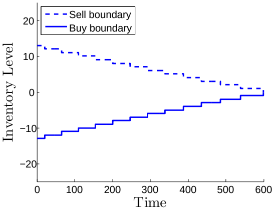
Low Imbalance | Middle Imbalance | High Imbalance

$\Delta = 2$

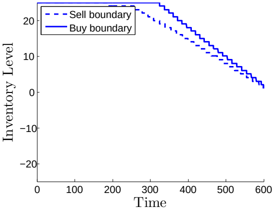
Optimal Trading Strategy - Buy and Sell Boundaries



Low Imbalance

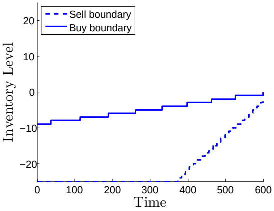


Middle Imbalance

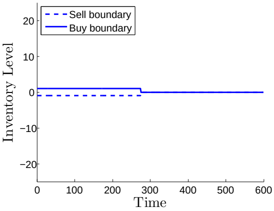


High Imbalance

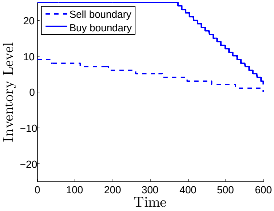
$\Delta = 1$



Low Imbalance



Middle Imbalance



High Imbalance

$\Delta = 2$

The Value of Knowing Imbalance

The Value of Knowing Imbalance

- The number of imbalance regimes is an important modelling choice.
- A large number of regimes can begin to cause observation and parameter estimation problems.
- A small number of regimes will not benefit as much from the predictive information.
- How does the performance of an agent depend on the number imbalance regimes in the model?

Simulation Procedure

- One day of data is simulated according to the model with $n_Z = 8$.
- These data are used to estimate parameters of the model when $n_Z = 1, 2, 4$, and 8 by collapsing observable imbalance states together.
- The "optimal" strategy is computed for each of these four choices of n_Z.
- Ten minutes of data are simulated according to the original model ($n_Z = 8$), and each trading strategy's performance is tested against it (plus two additional "naive" strategies).
- The previous step is repeated 50,000 times to get a distribution of performance results.

Simulation Results

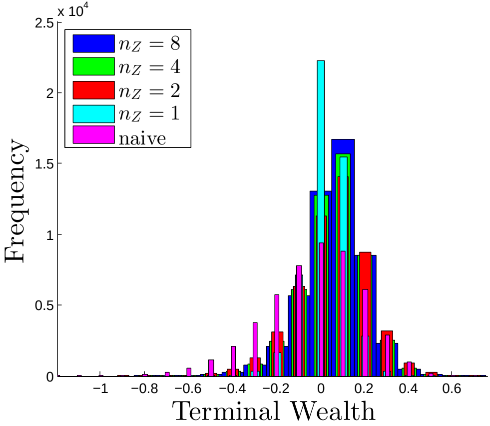


Figure : Distribution of terminal wealth for varying observable levels of imbalance and $\overline{Q} = -\underline{Q} = 5$. Data generating parameters estimated from BBBY.

Simulation Results

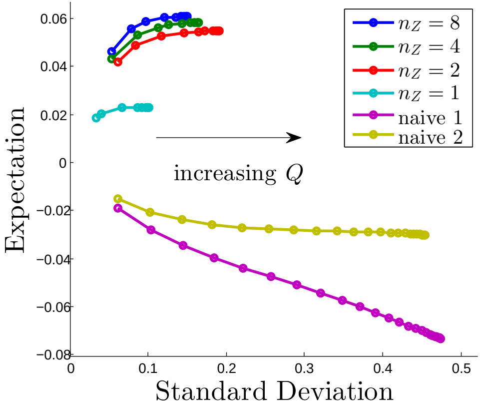


Figure : Expectation vs. Standard Deviation for varying observable levels of imbalance. Maximum inventory $Q = \overline{Q} = -\underline{Q}$ ranges from 1 to 25. Data generating parameters estimated from BBBY.

Simulation Results

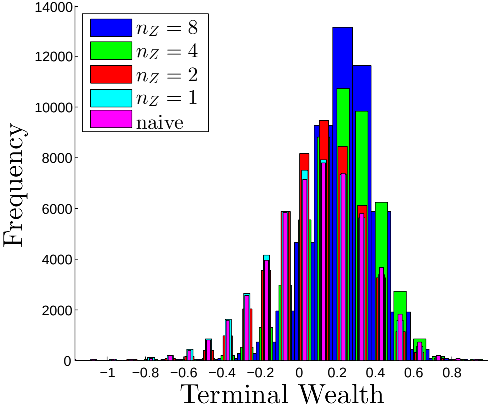


Figure : Distribution of terminal wealth for varying observable levels of imbalance and $\overline{Q} = -\underline{Q} = 5$. Data generating parameters estimated from MSFT.

Simulation Results

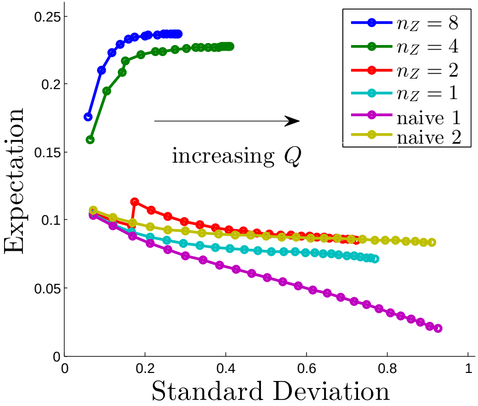


Figure : Expectation vs. Standard Deviation for varying observable levels of imbalance. Maximum inventory $Q = \overline{Q} = -\underline{Q}$ ranges from 1 to 25. Data generating parameters estimated from MSFT.

Simulation Results

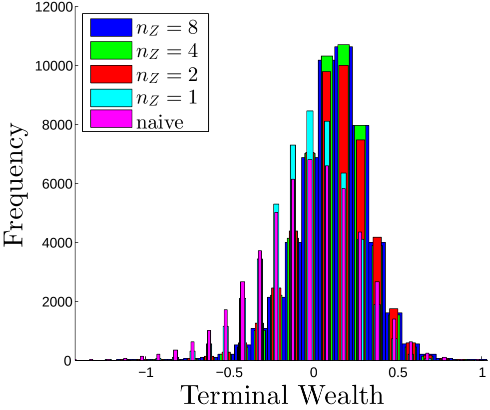


Figure : Distribution of terminal wealth for varying observable levels of imbalance and $\overline{Q} = -\underline{Q} = 5$. Data generating parameters estimated from TEVA.

Simulation Results

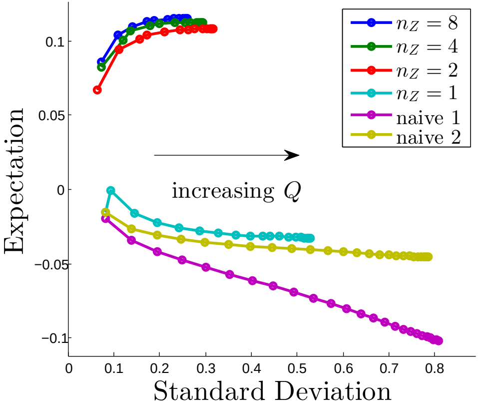


Figure : Expectation vs. Standard Deviation for varying observable levels of imbalance. Maximum inventory $Q = \overline{Q} = -\underline{Q}$ ranges from 1 to 25. Data generating parameters estimated from TEVA.

Conclusions

- The willingness of an agent to post limit orders is strongly dependent on the value of imbalance.
- Agent's should post buy orders more aggressively and sell orders more conservatively when imbalance is high. This reflects taking advantage of short term speculation and protecting against adverse selection.
- Corresponding opposite behaviour applies when imbalance is low.
- The additional value of being able to more accurately observe imbalance appears to have diminishing returns, but initially the additional value is very steep.

Future Endeavours

- Backtest strategies on real data.
- Investigate the effects of latency with respect to observing imbalance and spread.
- Expand the agent's controls to allow multiple limit order postings at different prices and of different volumes.
- Incorporate more realistic interactions between market orders and the agent's limit orders (i.e. queueing priority and partial fills).

Thanks for your attention!

Álvaro Cartea

a.cartea@ucl.ac.uk

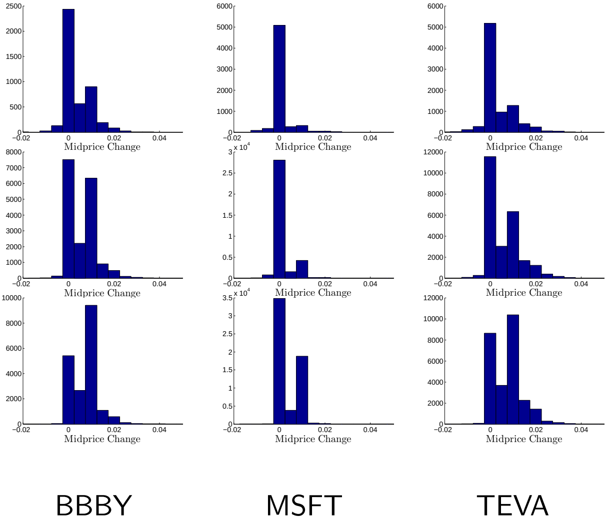


Figure : Distribution of midprice changes 20*ms* after a market buy order. Data is taken from a full month of trading (January 2011).

Simulation Parameters

$\lambda_\Delta = (0.119, 0.472)$	$\lambda_\Delta^\pm = (0.069, 0.030)$
$\lambda_Z = (0.97, 0.83, 0.89, 1.18, 1.18, 0.89, 0.83, 0.97)$	
$\lambda_Z^+ = (0.025, 0.025, 0.035, 0.041, 0.055, 0.073, 0.109, 0.186)$	
$\lambda_Z^- = (0.186, 0.109, 0.073, 0.055, 0.041, 0.035, 0.025, 0.025)$	
$\epsilon_\Delta = (0, 0)$	$\epsilon_\Delta^\pm = (0.556, 0.749)$
$\epsilon_Z = (-0.254, -0.084, -0.041, -0.008, 0.008, 0.041, 0.084, 0.254)$	
$\epsilon_Z^+ = (0.349, 0.333, 0.405, 0.432, 0.512, 0.550, 0.665, 0.854)$	
$\epsilon_Z^- = (0.854, 0.665, 0.550, 0.512, 0.432, 0.405, 0.333, 0.349)$	

Table : Estimated parameters for BBBY.

Simulation Parameters

$\lambda_\Delta = (0.050, 4.649)$	$\lambda_\Delta^\pm = (0.172, 0.521)$
$\lambda_Z = (1.43, 0.96, 0.71, 0.53, 0.53, 0.71, 0.96, 1.43)$	
$\lambda_Z^+ = (0.027, 0.033, 0.044, 0.056, 0.094, 0.210, 0.535, 1.565)$	
$\lambda_Z^- = (1.565, 0.535, 0.210, 0.094, 0.056, 0.044, 0.033, 0.027)$	
$\epsilon_\Delta = (0, 0)$	$\epsilon_\Delta^\pm = (0.277, 0.229)$
$\epsilon_Z = (-0.234, -0.012, -0.008, -0.004, 0.004, 0.008, 0.012, 0.234)$	
$\epsilon_Z^+ = (0.074, 0.148, 0.119, 0.152, 0.123, 0.227, 0.248, 0.436)$	
$\epsilon_Z^- = (0.436, 0.248, 0.227, 0.123, 0.152, 0.119, 0.148, 0.074)$	

Table : Estimated parameters for MSFT.

Simulation Parameters

$\lambda_\Delta = (0.225, 0.846)$	$\lambda_\Delta^\pm = (0.117, 0.049)$
$\lambda_Z = (1.62, 1.67, 1.81, 2.20, 2.20, 1.81, 1.67, 1.62)$	
$\lambda_Z^+ = (0.044, 0.050, 0.060, 0.066, 0.086, 0.119, 0.171, 0.331)$	
$\lambda_Z^- = (0.331, 0.171, 0.119, 0.086, 0.066, 0.060, 0.050, 0.044)$	
$\epsilon_\Delta = (0, 0)$	$\epsilon_\Delta^\pm = (0.534, 0.716)$
$\epsilon_Z = (-0.252, -0.084, -0.035, -0.006, 0.006, 0.035, 0.084, 0.251)$	
$\epsilon_Z^+ = (0.252, 0.329, 0.445, 0.479, 0.471, 0.541, 0.603, 0.752)$	
$\epsilon_Z^- = (0.752, 0.603, 0.541, 0.471, 0.479, 0.445, 0.329, 0.252)$	

Table : Estimated parameters for TEVA.