# The Micro-Price

Sasha Stoikov

Cornell University

Jim Gatheral @ NYU

# High frequency traders (HFT)

- HFTs are good:
    - Optimal order splitting
    - Pairs trading / statistical arbitrage
    - Market making / liquidity provision
    - Latency arbitrage
    - Sentiment analysis of news
- HFTs are evil:
    - The flash crash
    - Front running
    - Market manipulation and spoofing

# HFTs care about the imbalance

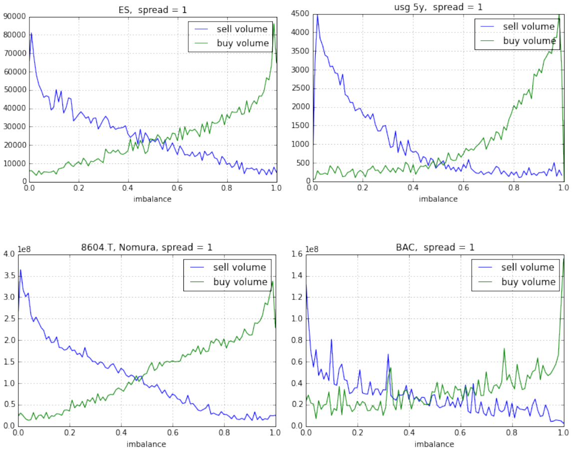


Figure: Buy and sell volume conditional on (pre-trade) Imbalance

# The mid-price

- The mid-price $M = \frac{P^b + P^a}{2}$
- $P^b$ is the best bid price
- $P^a$ is the best ask price
- Not a martingale (Bid-ask bounce)
- Low frequency signal
- Doesn't use volume at the best bid and ask prices.

# The weighted mid-price

- The weighted mid-price $M^w = IP^a + (1 - I)P^b$
- The imbalance $I = \frac{Q^b}{Q^b + Q^a}$
- $Q^b$ is the bid size and $Q^a$ is the ask size.
- Gatheral and Oomen (2009)
- Not a martingale
- Noisy
- Counter-intuitive examples

# The weighted mid-price example

- Assume $P^b = \$32.17$, $Q^b = 9$, $P^a = \$31.18$, $Q^a = 1$
- Assume the second best ask is \$31.19 and the second best ask size is 27
- $M^w = \$32.179 = 0.1 \cdot 32.17 + 0.9 \cdot 32.18$
- Order of size 1 at $P^a = \$31.18$ cancels
- New $M^w = \$32.1725 = 0.25 \cdot 32.17 + 0.75 \cdot 32.19$
- The ‘fair’ price just moved down after an ask order canceled?

# Features of the Micro-Price

- $P_t^{micro} = F(M_t, I_t, S_t)$
- Markov
- Martingale
- Computationally fast
- Better short term price predictions

# Outline

- General definition
- Toy models
  1. micro-price = mid price
  2. micro-price = weighted mid price
- A discrete Markov model
- Data analysis
- Conclusion

# Micro-price definition

Define

$$P_t^{micro} = \lim_{i \to \infty} P_t^i$$

where the approximating sequence of martingale prices is given by

$$P_t^i = \mathbb{E}\left[M_{\tau_i} | \mathcal{F}_t\right]$$

- $\mathcal{F}_t$ is the information contained in the order book at time $t$.
- $\tau_1, ..., \tau_n$ are (random) times when the mid-price $M_t$ changes

# Assumptions

## Assumption

*The information in the order book is given by the 3 dimensional Markov process* $\mathcal{F}_t = (M_t, I_t, S_t)$ *where* $M_t = \frac{1}{2}(P_t^b + P_t^a)$ *is the mid-price* $S_t = \frac{1}{2}(P_t^a - P_t^b)$ *is the bid-ask spread* $I_t = \frac{Q_t^b}{Q_t^b + Q_t^a}$ *is the imbalance at the top of the order book.*

## Assumption

*The dynamics of* $(M_t, I_t, S_t)$ *is independent of the level* $M_t$*, i.e.*

$$\mathbb{E}\left[M_{\tau_1} - M_t | M_t, I_t, S_t\right] \triangleq g^1(I_t, S_t)$$

# Main result

## Theorem

*Given Assumptions 1 and Assumption 2, the i-th approximation to the micro-price can be written as*

$$P_t^i = M_t + \sum_{k=1}^{i} g^k(I_t, S_t)$$

*where*

$$g^1(I_t, S_t) = \mathbb{E}\left[M_{\tau_1} - M_t | I_t, S_t\right]$$

*and*

$$g^{i+1}(I_t, S_t) = \mathbb{E}\left[g^i(I_{\tau_1}, S_{\tau_1}) | I_t, S_t\right], \forall j \geq 0$$

*can be computed recursively.*

# 3 examples

1. Mid-price independent of imbalance
2. Brownian imbalance
3. Discrete-time, finite state space

Interesting questions:

- Does the micro-price converge?
- What does it converge to?
- Is the micro-price between the bid and the ask?
- Is it sensible for large tick and small tick stocks?

# First example

If

- $M_s - M_t$ is independent of $I_t$ for all $s > t$
- $M_t$ is a continuous time random walk. The jumps are binomial and symmetric, i.e. $M_{\tau_{i+1}} - M_{\tau_i}$ takes values in $(-1, 1)$, have up and down probabilities of 0.5.
- The spread $S_t = 1$

then

$$P_t^{micro} = M_t$$

## Second example

If

- The process $I_t$ is a Brownian motion on the interval $[0,1]$.
- Let $\tau_{down} = \inf\{s > t : I_s = 0\}$ and $\tau_{up} = \inf\{s > t : I_s = 1\}$ and $\tau_1 = \min(\tau_{up}, \tau_{down})$
- When $I_t$ is absorbed to 1, the mid-price jumps up with probability 0.5 or bounces back with probability 0.5.
- When $I_t$ is absorbed to 0, the mid-price jumps down with probability 0.5 or bounces back with probability 0.5.
- The spread $S_t = 1$

then

$$P_t^{micro} = M_t + I_t - \frac{1}{2}$$

# Assumptions

- The time step is now discrete with $t \in \mathbb{Z}^+$,
- The imbalance $I_t$ takes discrete values $1 \le i_I \le n$,
- The spread $S_t$ takes discrete values $1 \le i_S \le m$
- The mid-price changes $M_{t+1} - M_t$ takes integer values in $K = \{k \mid 0 < |k| \le 2m\}$.
- Define the state $X_t = (I_t, S_t)$ with discrete values $1 \le i \le nm$

# Computing $g^1$

The first step approximation to the micro-price

$$
\begin{aligned}
g^1(i) =& \mathbb{E}\left[M_{\tau_1} - M_t | X_t = i\right] \\
=& \sum_{k \in K} k \cdot \mathbb{P}(M_{\tau_1} - M_t = k | X_t = i) \\
=& \sum_{k \in K} \sum_{s} k \cdot \mathbb{P}(M_{\tau_1} - M_t = k \wedge \tau_1 - t = s | X_t = i)
\end{aligned}
$$

# The transition probability matrix $T_1$

Then we define an *absorbing* Markov chain completely identified by the transition probability matrix $T^1$ in canonical form:

$$T^1 = \begin{pmatrix} Q & R^1 \\ 0 & \mathbb{I} \end{pmatrix}$$

- $Q$ is $nm \times nm$ matrix
- $R^1$ is $nm \times 4m$ matrix
- $\mathbb{I}$ is the $4m \times 4m$ matrix

# Computing $g^1$

Absorbing states

$$R^1_{ik} := \mathbb{P}(M_{t+1} - M_t = k | X_t = i)$$

Transient states

$$Q_{ij} := \mathbb{P}(M_{t+1} - M_t = 0 \wedge X_{t+1} = j | X_t = i)$$

Note that $R^1$ is an $nm \times 4m$ matrix and $Q$ is an $nm \times nm$ matrix.

$$g^1(i) = \Big(\sum_s Q^{s-1} R^1\Big)\underline{k} = \big(1 - Q\big)^{-1} R^1 \underline{k}$$

where $\underline{k} = \big[-2m, -2m+1, \ldots, -1, 1, \ldots 2m-1, 2m\big]^T$

## Computing $g^{i+1}$

Define a new matrix of absorbing states

$$R^2_{ik} := \mathbb{P}(M_{t+1} - M_t \neq 0 \wedge I_{t+1} = k | I_t = i)$$

Once again applying standard techniques for discrete time Markov processes with absorbing states

$$g^{i+1}(i) = \big(\sum_s Q^{s-1} R^2\big) g^i = \big(1 - Q\big)^{-1} R^2 g^i$$

## Checking that the micro-price converges

Define $B := \left(1 - Q\right)^{-1} R^2$.

**Theorem**

*If $B$ has strictly positive entries and $\lim_{k\to\infty} B^k = W$ where $W$ is the unique stationary distribution and $W\underline{g}^1 = 0$, then the limit*

$$\lim_{i\to\infty} p_t^i = p_t^{micro}$$

*converges.*

# A spectral decomposition for the micro-price

Perron-Frobenius decomposition

$$p_t^{micro} = \lim_{i \to \infty} p_t^i = M_t + \sum_{i=2}^{nm} \exp(\lambda_i) B_i g^1$$

where $\lambda_i$ are the eigenvalues of $B$ and $B_i$ are matrices formed from normalized left and right eigenvectors of $B$.

# The data

Bid and ask quotes for Bank of America (BAC) and Chevron (CVX), for the month of March 2011.

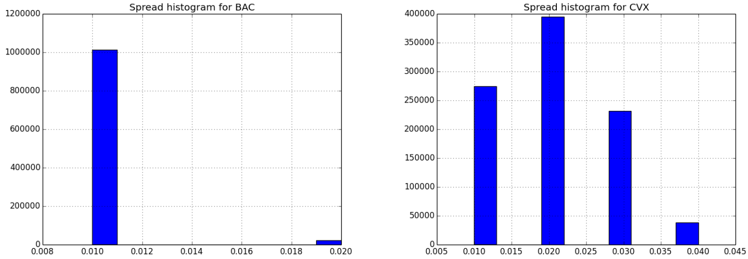


Figure: Spread histograms for BAC and CVX. BAC is a typical large tick stock and CVX is a typical small tick stock.

## The in-sample estimation

- Estimate transition probabilities $Q$, $R^1$ and $R^2$
- Compute $g^1 = \left(1 - Q\right)^{-1} R^1 \underline{k}$. This function is symmetrized to ensure that $g^1(i_I, i_S) = 1 - g^1(n - i_I, i_S)$.
- Compute $B = \left(1 - Q\right)^{-1} R^2$. This function is symmetrized to ensure that $B_{(i_I, i_S),(j_I, j_S)} = B_{(n-i_I, i_S),(n-j_I, j_S)}$. Note that the symmetrizing procedure ensures that $B\dot{g}^1 = 0$ and that the micro-price converges as guaranteed by Theorem 2.
- Perform a spectral decomposition of $B$ in terms of eigenvalues $\lambda_i$ and matrices $B_i$
- Compute the micro-price adjustment:

$$G^* = p^{micro} - M = \sum_{i=2}^{nm} \exp(\lambda_i) B_i g^1$$

# In-sample results

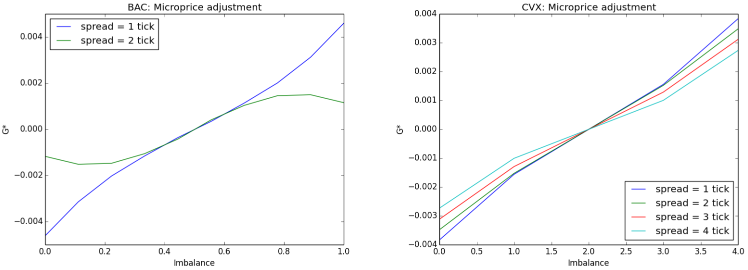


Figure: $G^* = p_t^{micro} - M_t$ as a function of $I$ and $S$

# Out of sample validation part 1

- Compute averages of $M_{t+60} - M_t$ grouped by $I_t$ and $S_t$ for 3 out of sample days
- Compare to $G^*$ obtained from the first day or March.

# Out of sample results part 1

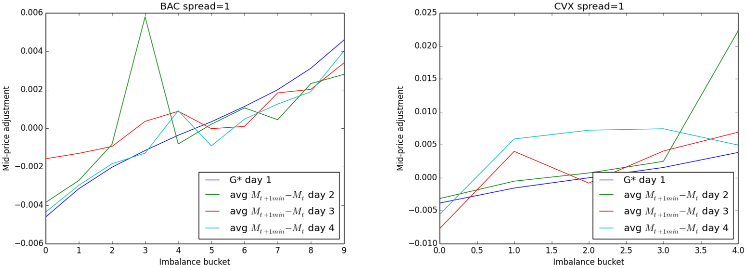


Figure: $G^*$ vs 1 min price predictions on three consecutive days

# Out of sample validation part 2

- Compute averages of $M_{t+60} - M_t$, $M_{t+300} - M_t$ and $M_{t+600} - M_t$ grouped by $I_t$ and $S_t$ for the entire month of March.
- Compare to $G^*$ obtained from the first day or March.

# Out of sample results part 2

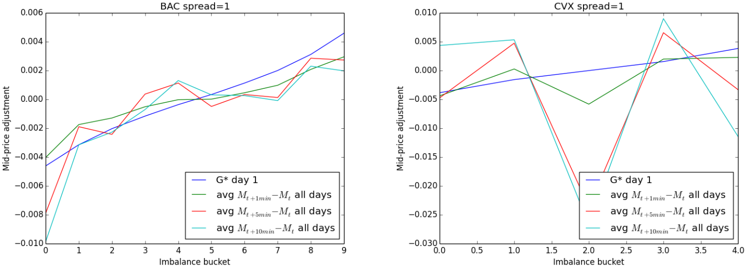


Figure: $G^*$ vs 1min, 5min and 10min price predictions for March 2011

## Summary

1. Have defined the micro-price as the expected mid-price in the distant future
2. When fitting a Markov model, we have conditions that ensures this micro-price converges
3. Micro-price is a good predictor of future mid prices
4. Micro-price can fit very different microstructures
5. Micro-price needs less data to converge than averaging mid price changes over fixed horizons
6. Micro-price is horizon independent
7. Micro-price seems to live between the bid and the ask

# Future work

1. Including other factors than imbalance and spread
2. Continuous models for the micro-price
3. Connections to quantities such as volatility, volume and tick size
4. High frequency volatility and correlation estimation
5. Applications to HFT strategies

Download the paper HERE