Multiscale problems in high frequency trading

Richard Sowers

Department of Industrial and Enterprise Systems Engineering
Department of Mathematics
University of Illinois at Urbana-Champaign

Large Deviations and Asymptotic Methods in Finance
Imperial College
April 2013
Modern algorithmic trading works at very high speed; clock cycles of computers.

Should be various problem with separations of scales.

Problems:

1. High frequency trading vs low frequency trading in the order book. Joint with A. Kirilenko (CFTC/MIT) and Carmen Meng

Reality is extremely complex. We want to understand some simple and stylized models. The goal is to think about how to begin to model things with provable theorems.
First problem: High frequency trading vs low frequency trading in the order book

Goal
Want to try to write down (and analyze) a model for high-frequency trading. Related work by

- Josh Reed
- Mike Ludkovski
- Cont and De Larrard
- Cont, Stoikov, and Taireja
The Limit Order Book

Limit Order Book

Order book is a collection of queues. Queues for buy orders are below current price, and queues for ask orders are above current price. Midpoint price is halfway between best bid and best ask. Most of high-frequency trading is at best bid or best ask (the level one queues). HFT'ers put and place orders at best bid and ask for a variety of reasons (some nefarious).
Can we model the level one queues with multiscale queues? Tools at hand are *fluid limits* and *diffusive limits*. $N$ a standard Poisson process. For $\kappa \ll 1$,

$$\kappa N \left( \frac{t}{\kappa} \right) \approx t + \sqrt{\kappa} B_t$$

where $B$ is a Brownian motion. First term is fluid limit, second is diffusion approximation.
Order queue dynamics

- Queues increase due to traders adding orders.
- Queues decrease due either to traders removing orders or orders being executed (crossing the book). In the latter case, we have a “pricing event”, but the midpoint price will not change (unless the queue empties). We focus on the midpoint price and do not distinguish between these two reasons the order queue decreases (should be corrected somehow).
Outline

Single queue (either bid or ask)
- Model for low-frequency (fundamental) traders
- Model for high frequency traders
- Heuristic idea of result

Volatility
- Model for bid and ask queues
- Volatility correction
Model for low-frequency (fundamental) traders

Let $\delta \ll 1$ be the tick size. Also need exponents $\nu_0$, $\gamma_0$, and $\iota$.

- Let’s assume that low-frequency traders add and remove orders of size $\delta^{\nu_0}$ and that they do this with rates (respectively) of $\lambda_L^+ / \delta^{\gamma_0}$ (rate of increase) and $\lambda_L^- / \delta^{\gamma_0}$ (rate of decrease).

- Let’s assume that the initial size of the queue is $q \delta^{\iota}$.

Queue size due to fundamental trading is:

$$Q^L_t = q \delta^{\iota} + \delta^{\nu_0} N^{L,+} \left( \frac{\lambda_L^+}{\delta^{\gamma_0}} t \right) - \delta^{\nu_0} N^{L,-} \left( \frac{\lambda_L^-}{\delta^{\gamma_0}} t \right)$$

$$\approx q \delta^{\iota} - \delta^{\nu_0-\gamma_0} \ell t + \delta^{\nu_0-\gamma_0}/2 \sigma_L B^L_t$$

where $\ell \overset{\text{def}}{=} \lambda_L^- - \lambda_L^+$ and $\sigma^2_L = (\lambda_L^+)^2 + (\lambda_L^-)^2$. Set $\tau \overset{\text{def}}{=} \inf \{ t \geq 0 : Q^L_t \leq 0 \}$, we have that

$$\tau \approx \frac{q}{\ell} \delta^{\iota-\nu_0+\gamma_0} = \frac{q}{\ell} \delta^2$$

if $\iota = 2 + \nu_0 - \gamma_0$; Diffusive scaling; $S$ moves by tick $\delta$ in time $\delta^2$. 
Assumption

High frequency traders are a higher-order perturbation of market; they don’t change *dominant* (diffusive) behavior of market. Model HFT as Poisson additions and removals from queue, but at a level *between* fluid and diffusive approximations for fundamental traders. Also, want to *balance* arrivals and removals so as to not change dominant diffusive behavior of market.
Fix $\varepsilon > 0$ (somehow smaller than $\delta$). Also need exponents $\nu$ and $\gamma$. Let's assume that low-frequency traders add and remove orders of size $\varepsilon^\nu$ and that they do this with common rate $\bar{\lambda}_L \varepsilon^\gamma$. Queue size due to both fundamental and high frequency trading is:

$$Q_t = q_{\delta^\iota} + \delta^{\nu_0} N^{L,+} \left( \frac{\lambda_L^+}{\delta^\gamma_0} t \right) - \delta^{\nu_0} N^{L,-} \left( \frac{\lambda_L^-}{\delta^\gamma_0} t \right)$$

$$+ \varepsilon^\nu N^{H,+} \left( \frac{\bar{\lambda}}{\varepsilon^\gamma} t \right) - \varepsilon^\nu N^{H,-} \left( \frac{\bar{\lambda}}{\varepsilon^\gamma} t \right)$$

$$\approx q_{\delta^\iota} - \delta^{\nu_0 - \gamma_0} \ell t + \delta^{\nu_0 - \gamma_0}/2 \sigma_L B_t^L$$

$$+ \varepsilon^{\nu - \gamma} \bar{\lambda} t - \varepsilon^{\nu - \gamma} \bar{\lambda} t + \varepsilon^{\nu - \gamma}/2 \sigma_H B_t^H$$

where $\sigma^2_H \overset{\text{def}}{=} 2\bar{\lambda}$. Want

$$\delta^{\nu_0 - \gamma_0}/2 \ll \varepsilon^{\nu - \gamma} \ll \delta^{\nu_0 - \gamma_0}$$

(also need another technical condition)
Core question

\[ Q_t \approx q \delta^t - \delta^{-\gamma} \nu \ell t + \varepsilon^{\nu-\gamma/2} \sigma_B H^t \]

\[ = \bar{q} - \bar{\mu} t + \bar{\kappa} B_t^H \]

\[ \tau_{\delta, \varepsilon} \overset{\text{def}}{=} \inf \{ t \geq 0 : Q_t \leq 0 \} \]

\[ = \inf \left\{ t \geq 0 : \frac{\bar{q}}{\bar{\kappa}} t - B_t^H = \frac{\bar{q}}{\bar{\kappa}} \right\} \]

What are asymptotics of \( \tau_{\delta, \varepsilon} \) as \( \kappa \downarrow 0 \)? Karatzas and Shreve, Exercise 3.5.9:

\[ \mathbb{E} \left[ \exp \left[ -\alpha \tau_{\delta, \varepsilon} \right] \right] = \exp \left[ \frac{\bar{\mu} \bar{q}}{\bar{\kappa}^2} - \frac{\bar{q}}{\bar{\kappa}} \sqrt{ \left( \frac{\bar{\mu}}{\bar{\kappa}} \right)^2 + 2\alpha} \right] \]

\[ = \exp \left[ \frac{\bar{\mu} \bar{q}}{\bar{\kappa}^2} \left\{ 1 - \sqrt{1 + \frac{2\alpha \bar{\kappa}^2}{\bar{\mu}^2}} \right\} \right] \]
Simulation

Richard Sowers
Multiscale problems in high frequency trading
\[ \tau_{\delta,\varepsilon} = \inf \left\{ t \geq 0 : \frac{\bar{q}}{\bar{\kappa}} t - B^H_t = \frac{\bar{q}}{\bar{\kappa}} \right\} \]

We expect that \( \mathbb{E}[\tau] \approx \frac{\bar{q}}{\bar{\mu}} \). Also, \( \sqrt{1 + 2x} \approx 1 + x - \frac{1}{2} x^2 \) for \( x \ll 1 \).

\[
\mathbb{E} \left[ \exp \left[ -\alpha \left( \tau_{\delta,\varepsilon} - \frac{\bar{q}}{\bar{\mu}} \right) \right] \right] = \exp \left[ \frac{\bar{\mu} \bar{q}}{\bar{\kappa}^2} \left\{ 1 - \sqrt{1 + \frac{2\alpha \bar{\kappa}^2}{\bar{\mu}^2}} \right\} - \frac{\alpha \bar{q}}{\bar{\mu}} \right]
\]
\[
\approx \exp \left[ \frac{\alpha^2 \bar{\kappa}^2 \bar{q}}{2 \bar{\mu}^3} \right]
\]

Setting \( \xi \overset{\text{def}}{=} \frac{\tau - (\bar{q}/\bar{\mu})}{\sqrt{\bar{q} \bar{\kappa}}/\bar{\mu}^{3/2}} \),
we get
\[
\mathbb{E} \left[ \exp \left[ -\alpha \xi \right] \right] \approx \exp \left[ \frac{1}{2} \alpha^2 \right]
\]
in other words (via a rigorous form of these equations), \( \xi \approx \mathcal{N}(0, 1) \), or rather
\[
\tau \approx \frac{\bar{q}}{\bar{\mu}} + \frac{\sqrt{\bar{q} \bar{\kappa}}}{\mu^{3/2}} \eta \quad (\eta = \mathcal{N}(0, 1))
\]

Richard Sowers  Multiscale problems in high frequency trading
Real proof is complicated (40+ pages). Consider process 

\[ Z_t = \exp [A_\theta Q_t + \nu \theta (t - \bar{q}/\bar{\mu})] . \]

Arrange \( A_\theta \) so that \( Z \) is a martingale. Note that \( Q_\tau \approx 0 \). Need to control various errors. \( A_\theta \) implicitly.
Unwrapping,

\[ \tau \approx \delta^2 + C\delta^{1-\nu_o+\gamma_o} \varepsilon^{\nu-\gamma/2} \eta \]

**Figure**: Poisson arrival rate for HFT; parameters are \( \bar{\lambda} = 40, \lambda^+ = 9, \lambda^- = 1, \delta = .3, \varepsilon = 0.01, \nu_o = 0.001, \gamma_o = 5, \nu = 0.001, \) and \( \gamma = 1.5. \)
Back to macroscopic problem of volatility

Assume bid and ask queues are independent and of same form;

\[ \tau_B \approx \delta^2 + C_B \delta^{1-\nu_1+\ve_1} \nu \gamma / 2 \eta_B \]

\[ \tau_A \approx \delta^2 + C_A \delta^{1-\nu_1+\ve_1} \nu \gamma / 2 \eta_A \]

Time that price moves is \( \tau \overset{\text{def}}{=} \tau_B \wedge \tau_A \), and

\[ \mathbb{E}[\tau] = \delta^2 + \delta^{1-\nu_1+\ve_1} \nu \gamma / 2 \mathbb{E}[C_B \eta_B \wedge C_A \eta_A]. \]

Let’s arrange things so that asymptotically, price change due to bid and ask are same (so \( \mathbb{P}\{\text{up}\} = \mathbb{P}\{\text{down}\} = \frac{1}{2} \)). Corrected volatility is

\[ \sigma^2 = \frac{\delta^2}{\delta^2 - \delta^{1-\nu_1+\ve_1} \nu \gamma / 2 C} \approx \frac{1}{1 - \delta^{-2} \delta^{1-\nu_1+\ve_1} \nu \gamma / 2 C} \approx \delta^2 + C \delta^{3-\nu_1-\ve_1} \nu \gamma / 2 \]

\[ = 1 + C \left( \frac{\ve \gamma / 2 2 \sqrt{\lambda}}{(q \delta_1)^{3/2}} \right) \left( \delta_1 \nu_1 \ell \right)^{1/2} \delta^2 \]
Excess volatility is proportional to

\[ \text{volatility of HFT arrival rate} \times \sqrt{\text{rate at which LT’ers empty queue}} \times (\text{tick size})^2 / (\text{initial size of queue})^{3/2}. \]
Recall that rate of HFT was rescaled $\bar{\lambda}$? Actually, we can (and did) consider \textit{stochastic arrival rates}; replace $\bar{\lambda}$ by CIR process

$$
 d\lambda_{t}^{\pm,\varepsilon} = -\frac{\alpha}{\varepsilon}(\lambda_{t}^{\pm,\varepsilon} - \bar{\lambda})dt + \frac{\sigma^{\pm}}{\sqrt{\varepsilon}}\sqrt{\lambda_{t}^{\pm,\varepsilon}} dW_{t}^{\pm}
$$

Adds another possible asymptotic regime. An attempt to model a \textit{statistical collection} of HFT traders with different strategies. In this other regime, dominant asymptotics are due to fluctuations of rate (Flash Crash).
Calculation is

\[
\varepsilon^\nu \tilde{N} \left( \varepsilon^{-\gamma} \int_{s=0}^{t} \lambda_s^\varepsilon ds \right)
\]

\[
\approx \varepsilon^{\nu-\gamma} \int_{s=0}^{t} \lambda_s^\varepsilon ds + \varepsilon^{\nu-\gamma}/2 \tilde{V}_t^a \int_{s=0}^{t} \lambda_s^\varepsilon ds
\]

\[
\approx \varepsilon^{\nu-\gamma} \bar{\lambda} t + \varepsilon^{\nu-\gamma+1/2} \frac{\sigma}{\alpha} \int_{s=0}^{t} \sqrt{\lambda_s^\varepsilon} dW_s + \varepsilon^{\nu-\gamma}/2 \tilde{V}_t^a \int_{s=0}^{t} \lambda_s^\varepsilon ds
\]

\[
\approx \varepsilon^{\nu-\gamma} \bar{\lambda} t + \varepsilon^{\nu-\gamma+1/2} \frac{\sigma}{\alpha} \tilde{V}_t^b \int_{s=0}^{t} \lambda_s^\varepsilon ds + \varepsilon^{\nu-\gamma}/2 \tilde{V}_t^a \int_{s=0}^{t} \lambda_s ds
\]

\[
\approx \varepsilon^{\nu-\gamma} \bar{\lambda} t + \varepsilon^{\nu-\gamma+1/2} \frac{\sigma}{\alpha} \tilde{V}_t^b + \varepsilon^{\nu-\gamma}/2 \tilde{V}_t^a
\]

where \(\tilde{V}\)'s are Brownian motions and
\[ \bar{\lambda} = 40 \text{ and } \lambda_L^- = 9 \]

\[ \bar{\lambda} = 100 \text{ and } \lambda_L^- = 2 \]

**Figure:** Common parameters are \( \delta = .3, \varepsilon = 0.01, \lambda_L^+ = 1, \nu_\circ = 0.001, \gamma_\circ = 5, \alpha_1 = 0.065, \alpha_2 = 0.061, \lambda_\circ = \bar{\lambda} + 1, \sigma_1 = 0.057, \sigma_2 = 0.0576, \nu = 0.001, \) and \( \gamma = 1.5. \)

Seems to introduce asymmetries (connected with time for CIR to reach equilibrium?)
Thanks