Limit Order Book Dynamics

My goal is to use the dynamics of the Limit Order Book (LOB) as an indicator for high-frequency stock price movement, thus enabling statistical arbitrage. Formally, I will the study limit order book imbalance process, $I(t)$, and the stock price process, $S(t)$, and attempt to establish a stochastic relationship $\dot{S} = f(S, I, t)$. I will then solve the stochastic control problem to derive an optimal trading strategy based on the observed relationship.

For Our Readers in the Middle East...

Oh well wouldn't you look at that, another one right in your fucking inbox, init? Bam, hundred bucks please!

Look last time I had the leisure of writing to you at length about life and the cosmos, but we're a little tight on time today so I might have to cut the goose chase a little short and go straight for the cow. Well why're we tight on time? As mentioned, I slept right the fuck in. Actually I was awakened at 5am due to my phone buzzing with your messages - forgot to silence that shit. Again at 730 with alarm number one, again at 800 with alarm number two, and at 1040 I finally darted awake, clearly a moment of clarity and omniscience as I immediately and accurately summed up the situation with an insightful outcry of "fuck!". Our lab lunch is in an hour, I really want to hit the gym because my relentless alcohol consumption isn't exactly doing favors for this weekend's gymnastics competition, and tonight I'm partying it up in Casa Loma. I'm excited as fuck about that. Look it's not the Casa Loma party I've always dreamed of ... in which we're wearing black smoking jackets, got fat fucking cigars in our mouths, open bar without queues, smokeshow babes in heels, we're bidding on auctions for some charity for camp for retards but it's like million dollar cars or butter sculptures, and by the end of the night we're fucking *tanked* but can still form sentences, no puking, straight line test probably okay, but tanked, and we're just told we won the chopper auction so we're like okay Jeeves, fuck this shit, fly us home. Basically I don't think it's gonna be that tonight, but maybe something close to it.

The Academic Week in Review

This week has been tits. I've gotten so much done it's not even funny, which included meeting with Sebastian once and with my dad twice. I mean, one might argue that *they* got a lot done and I was able to later reproduce it by referencing their notes, but who am I to downplay my own contribution/significance, right. Anyway, today I'm pleased to report that **Step 1** of the course project has been achieved. As a refresher:

Step 1 Re-derive and thereby check the coupled system of PDEs that solves the optimization problem.

An agent wants to liquidate a position by time T using limit orders (LOs) posted at an arbitrary depth δ_t at time t, which is the process that we control. The following processes are at play:

Order Imbalance	$Z = (Z_t)_{0 \le t \le T} \in \{-1, 0, +1\}$	a Markov chain with generator G
Arrival of MOs	$M^{\pm} = (M_t^{\pm})_{0 \le t \le T}$	Poisson with rate $\lambda^{\pm}(Z_t)$
Arrival of LOB Shuffling	$J^{\pm} = (J_t^{\pm})_{0 \le t \le T}$	Poisson with rate $\gamma^{\pm}(Z_t)$
ΔPrice: MO arrives	$\{\epsilon_{0,k}^{\pm}, \epsilon_{1,k}^{\pm}, \dots\} \sim F_k^{\pm}$	i.i.d. with $k \in \{-1, 0, 1\}$
ΔPrice: LOB shuffled	$\{\eta_{0,k}^{\pm}, \eta_{1,k}^{\pm}, \dots\} \sim L_k^{\pm}$	i.i.d. with $k \in \{-1, 0, 1\}$
Midprice	$S = (S_t)_{0 \le t \le T}$ $\mathrm{d}S_t = \epsilon^{+}_{M^{+}_{t^-}, Z_{t^-}} \,\mathrm{d}M_t^{+} - \epsilon^{-}_{M^{-}_{t^-}, Z_{t^-}} \,\mathrm{d}M_t^{-} + \eta^{+}_{J^{+}_{t^-}, Z_{t^-}} \,\mathrm{d}J_t^{+} - \eta^{-}_{J^{-}_{t^-}, Z_{t^-}} \,\mathrm{d}J_t^{-}$	
LO posted depth	$\delta = (\delta_t)_{0 \le t \le T}$	
LO fill count	$N^{\delta} = (N_t^{\delta})_{0 \le t \le T}$	
Cash	$X^{\delta} = (X_t^{\delta})_{0 \le t \le T}$ $\mathrm{d}X_t^{\delta} = (S_t + \delta_t)\,\mathrm{d}N_t^{\delta}$	
Inventory	$Q_t = Q_0 - N_t^{\delta}$	

Note that N_t^{δ} is NOT a Poisson process - it is a jump process satisfying the relationship that if at time t we have a sell limit order posted at a depth δ_t, then our fill probability is $e^{-\kappa \delta_t}$ conditional on a buy market order arriving; namely:

$$\mathcal{P}[\,\mathrm{d}N_t = 1 \mid \mathrm{d}M_t^{+} = 1] = e^{-\kappa \delta_t}$$

For our liquidation optimization problem, we stop trading either when we liquidate all our shares, or the trading day ends. Therefore, our stopping time τ can be expressed as

$$\tau = T \wedge \min\{t : Q_t^{\delta} = 0\}$$

We measure the performance of our LO posting strategy δ according to the *performance criteria* *function* H^δ:

$$\begin{aligned} H^\delta(t,x,s,z,q) &= \mathbb{E}[X_\tau^\delta + Q_\tau^\delta(S_\tau - \alpha Q_\tau^\delta) - \varphi \int_t^\tau (Q_u^\delta)^2 \, \mathrm{d}u \,|\, X_{t^-}^\delta = x, S_{t^-}^\delta = s, Z_{t^-} = z, Q_{t^-}^\delta = q] \\ &= \mathbb{E}_{t,x,s,z,q}[X_\tau^\delta + Q_\tau^\delta(S_\tau - \alpha Q_\tau^\delta) - \varphi \int_t^\tau (Q_u^\delta)^2 \, \mathrm{d}u] \end{aligned}$$

We seek to maximize this performance criteria, and thereby attain the *value function* H:

$$H(t,x,s,z,q) = \sup_{\delta \in \mathcal{A}} H^\delta(t,x,s,z,q)$$

where $\mathcal{A}$ is the set of $\mathcal{F}$-predictable, bounded from below processes – note I actually still have no idea what this means.

I've shown (canonically, nothing innovative) that we can apply the dynamic programming principle (DPP) to H. This means that instead of solving for the optimal strategy δ for the entire trading day, we instead only have to solve for it over a small time interval h, and after that time interval we 'behave optimally'. In other words, we can express H as:

$$H(t,x,s,z,q) = \sup_{\delta \in \mathcal{A}} \mathbb{E}_{t,x,s,z,q}\left[-\varphi \int_t^{t+h} (Q_u^\delta)^2 \, \mathrm{d}u + H(t+h, X_{t+h}^\delta, S_{t+h}^\delta, Z_{t+h}, Q_{t+h}^\delta)\right]$$

Now again, rather canonically, we can take this DPP and write it in infinitesimal form, which is called the dynamic programming equation (DPE). Writing it so, we end up with the following:

$$\begin{cases} \partial_t H(t,x,s,z,q) + \sup_{\delta \in \mathcal{A}} \left\{ \mathcal{L}_t^\delta H(t,x,s,z,q) + F(t,x,s,z,q,\delta) \right\} = 0 \\ \qquad\qquad\qquad\qquad\qquad\qquad H(\tau,x,s,z,q) = G(x,s,z,q) \end{cases} \tag{1}$$

where $F(t,x,s,z,q,\delta) = -\varphi q^2$, and $H(\tau,x,s,z,q) = x - q(s - \alpha q)$, both of which are read directly off of our performance criteria function H^δ, and $\mathcal{L}_t^\delta$ is the so-called 'infinitesimal generator' for our vector of random processes (x,s,z,q).

To solve for $\mathcal{L}_t^\delta$ we will determine $\mathrm{d}H$. In order to do so, we need to present some results on Ito's formula for Poisson processes and for Markov chains.

Look Teddy, if I had more time for this update, I would put them in, but as it stands I really wanna get that workout in, and it's already 3pm. So I'm going to omit the Ito formula derivations, suffice to say they took me a FUCKING LONG TIME.

Consider our Markov chain Z with generator G. Define $K_l(t)$ to be the number of jumps with $Z_s - Z_{s^-} = l$ up to time t, and $\beta_l(x) = G_{x,x+l}$. Then $\tilde{K}_l(t) = K_l(t) - \int_0^t \beta_l(Z_s) \, \mathrm{d}s$ is a martingale.

Now we solve for $\mathrm{d}H$, using the shorthand $H(\cdot) = H(t,x,s,z,q)$:

$$
\begin{aligned}
\mathrm{d}H(t,x,s,z,q) &= \partial_t H\,\mathrm{d}t + \partial_x H\,\mathrm{d}x + \partial_s H\,\mathrm{d}s + \partial_q H\,\mathrm{d}q + \partial_z H\,\mathrm{d}z \\
&= \partial_t H\,\mathrm{d}t + \partial_{M^+} H\,\mathrm{d}M^+ + \partial_{M^-} H\,\mathrm{d}M^- + \partial_{J^+} H\,\mathrm{d}J^+ + \partial_{J^-} H\,\mathrm{d}J^- \\
&\quad + \sum_l \left[H(t,x,s,z+l,q) - H(\cdot)\right]\mathrm{d}K_l \\
&= \partial_t H\,\mathrm{d}t + \Big\{e^{-\kappa\delta}\left(\mathbb{E}\left[H(t,x+(s+\delta),s+\epsilon^+_{0,z},z,q-1)\right] - H(\cdot)\right) \\
&\qquad\qquad + (1-e^{-\kappa\delta})\left(\mathbb{E}\left[H(t,x,s+\epsilon^+_{0,z},z,q)\right] - H(\cdot)\right)\Big\}\,\mathrm{d}M^+ \\
&\quad + \left\{\mathbb{E}\left[H(t,x,s-\epsilon^-_{0,z},z,q)\right] - H(\cdot)\right\}\mathrm{d}M^- \\
&\quad + \left\{\mathbb{E}\left[H(t,x,s+\eta^+_{0,z},z,q)\right] - H(\cdot)\right\}\mathrm{d}J^+ \\
&\quad + \left\{\mathbb{E}\left[H(t,x,s-\eta^-_{0,z},z,q)\right] - H(\cdot)\right\}\mathrm{d}J^- \\
&\quad + \sum_l \left[H(t,x,s,z+l,q) - H(\cdot)\right]\mathrm{d}K_l
\end{aligned}
$$

Substitute in the compensated identifies

$$
\begin{aligned}
\mathrm{d}M^\pm &= \mathrm{d}\tilde{M}^\pm + \lambda^\pm(z)\,\mathrm{d}t \\
\mathrm{d}M^\pm &= \mathrm{d}\tilde{M}^\pm + \gamma^\pm(z)\,\mathrm{d}t \\
\mathrm{d}K_l &= \mathrm{d}\tilde{K}_l + \beta_l(z)\,\mathrm{d}t
\end{aligned}
$$

$$
\begin{aligned}
&= \Bigg\{\partial_t H + \lambda^+(z)\Big\{e^{-\kappa\delta}\left(\mathbb{E}\left[H(t,x+(s+\delta),s+\epsilon^+_{0,z},z,q-1)\right] - H(\cdot)\right) \\
&\qquad\qquad + (1-e^{-\kappa\delta})\left(\mathbb{E}\left[H(t,x,s+\epsilon^+_{0,z},z,q)\right] - H(\cdot)\right)\Big\} \\
&\quad + \lambda^-(z)\left\{\mathbb{E}\left[H(t,x,s-\epsilon^-_{0,z},z,q)\right] - H(\cdot)\right\} \\
&\quad + \gamma^+(z)\left\{\mathbb{E}\left[H(t,x,s+\eta^+_{0,z},z,q)\right] - H(\cdot)\right\} \\
&\quad + \gamma^-(z)\left\{\mathbb{E}\left[H(t,x,s-\eta^-_{0,z},z,q)\right] - H(\cdot)\right\} \\
&\quad + \sum_l \beta_l(z)\left[H(t,x,s,z+l,q) - H(\cdot)\right]\Bigg\}\,\mathrm{d}t \\
&\quad + \Big\{e^{-\kappa\delta}\left(\mathbb{E}\left[H(t,x+(s+\delta),s+\epsilon^+_{0,z},z,q-1)\right] - H(\cdot)\right) \\
&\qquad + (1-e^{-\kappa\delta})\left(\mathbb{E}\left[H(t,x,s+\epsilon^+_{0,z},z,q)\right] - H(\cdot)\right)\Big\}\,\mathrm{d}\tilde{M}^+ \\
&\quad + \left\{\mathbb{E}\left[H(t,x,s-\epsilon^-_{0,z},z,q)\right] - H(\cdot)\right\}\mathrm{d}\tilde{M}^- \\
&\quad + \left\{\mathbb{E}\left[H(t,x,s+\eta^+_{0,z},z,q)\right] - H(\cdot)\right\}\mathrm{d}\tilde{J}^+ \\
&\quad + \left\{\mathbb{E}\left[H(t,x,s-\eta^-_{0,z},z,q)\right] - H(\cdot)\right\}\mathrm{d}\tilde{J}^- \\
&\quad + \sum_l \left[H(t,x,s,z+l,q) - H(\cdot)\right]\mathrm{d}\tilde{K}_l
\end{aligned}
$$

From which we can see that the infinitesimal generator is given by

$$\begin{aligned}
\mathcal{L}_t^\delta = \lambda^+(z) \Big\{ & e^{-\kappa\delta} \left(\mathbb{E}\left[H(t, x + (s+\delta), s + \epsilon_{0,z}^+, z, q-1) \right] - H(\cdot) \right) \\
& + (1 - e^{-\kappa\delta}) \left(\mathbb{E}\left[H(t, x, s + \epsilon_{0,z}^+, z, q) \right] - H(\cdot) \right) \Big\} \\
& + \lambda^-(z) \left\{ \mathbb{E}\left[H(t, x, s - \epsilon_{0,z}^-, z, q) \right] - H(\cdot) \right\} \\
& + \gamma^+(z) \left\{ \mathbb{E}\left[H(t, x, s + \eta_{0,z}^+, z, q) \right] - H(\cdot) \right\} \\
& + \gamma^-(z) \left\{ \mathbb{E}\left[H(t, x, s - \eta_{0,z}^-, z, q) \right] - H(\cdot) \right\} \\
& + \sum_l \beta_l(z) \left[H(t, x, s, z + l, q) - H(\cdot) \right]
\end{aligned} \tag{2}$$

In our case, the Markov chain Z can only be in three possible states $-1, 0, 1$, so summing over the possible jump sizes l is identical (up to reordering) of summing over the possible destination states k with the corresponding transition rates. Thus, plugging this back into our DPE equation 1, we obtain:

$$\begin{aligned}
\varphi q^2 = \partial_t H(t, x, s, z, q) + \lambda^+(z) \sup_{\delta \in \mathcal{A}} \Big\{ & e^{-\kappa\delta} \left(\mathbb{E}\left[H(t, x + (s+\delta), s + \epsilon_{0,z}^+, z, q-1) \right] - H(\cdot) \right) \\
& + (1 - e^{-\kappa\delta}) \left(\mathbb{E}\left[H(t, x, s + \epsilon_{0,z}^+, z, q) \right] - H(\cdot) \right) \Big\} \\
+ \lambda^-(z) & \left\{ \mathbb{E}\left[H(t, x, s - \epsilon_{0,z}^-, z, q) \right] - H(\cdot) \right\} \\
+ \gamma^+(z) & \left\{ \mathbb{E}\left[H(t, x, s + \eta_{0,z}^+, z, q) \right] - H(\cdot) \right\} \\
+ \gamma^-(z) & \left\{ \mathbb{E}\left[H(t, x, s - \eta_{0,z}^-, z, q) \right] - H(\cdot) \right\} \\
+ \sum_{k=-1,0,1} & G_{z,k} \left[H(t, x, s, k, q) - H(\cdot) \right]
\end{aligned} \tag{3}$$

where expectations are over the random variables $\epsilon_{0,z}^\pm$ and $\eta_{0,z}^\pm$. We also obtain two boundary conditions from the two possible stopping times:

$$\begin{cases} H(T, x, s, z, q) = x + q(s - \alpha q) & \text{if } \tau = T \\ H(t, x, s, z, 0) = x & \text{if } q = 0 \text{ (inventory has been liquidated by } \tau < T) \end{cases}$$

The terminal and boundary conditions suggest a familiar ansatz for H:

$$H(t, x, s, z, q) = x + qs + h(t, z, q) \tag{4}$$

where

$$\begin{aligned} h(T, z, q) &= -\alpha q^2 \\ h(t, z, 0) &= 0 \end{aligned}$$

Substituting the ansatz 4 back into the PDE 3, we obtain:

$$
\begin{aligned}
\varphi q^2 = \partial_t h(t,z,q) &+ \lambda^+(z) \sup_{\delta\in\mathcal{A}} \Big\{ e^{-\kappa\delta}\mathbb{E}\left[x + (s+\delta) + (q-1)(s+\epsilon^+_{0,z}) + h(t,z,q-1)\right. \\
&\qquad\qquad \left. - x - qs - h(t,z,q)\right] \\
&\qquad + (1-e^{-\kappa\delta})\mathbb{E}\left[x + q(s+\epsilon^+_{0,z}) + h(t,z,q)\right. \\
&\qquad\qquad \left. - x - qs - h(t,z,q)\right] \Big\} \\
&+ q\left(-\lambda^-(z)\mathbb{E}\left[\epsilon^-_{0,z}\right] + \gamma^+(z)\mathbb{E}\left[\eta^+_{0,z}\right] - \gamma^-(z)\mathbb{E}\left[\eta^-_{0,z}\right]\right) \\
&+ \sum_{k=-1,0,1} G_{z,k}\left[h(t,k,q) - h(t,z,q)\right] \\
= \partial_t h(t,z,q) &+ \lambda^+(z)\sup_{\delta\in\mathcal{A}}\left\{ e^{-\kappa\delta}\left(\delta - \mathbb{E}\left[\epsilon^+_{0,z}\right] + h(t,z,q-1) - h(t,z,q)\right)\right\} \\
&+ \mu(z)q + \sum_{k=-1,0,1} G_{z,k}\left[h(t,k,q) - h(t,z,q)\right]
\end{aligned}
$$

where

$$
\mu(z) = \lambda^+(z)\mathbb{E}\left[\epsilon^+_{0,z}\right] - \lambda^-(z)\mathbb{E}\left[\epsilon^-_{0,z}\right] + \gamma^+(z)\mathbb{E}\left[\eta^+_{0,z}\right] - \gamma^-(z)\mathbb{E}\left[\eta^-_{0,z}\right]
$$

To find the supremum over δ, consider the first-order constraint:

$$
\begin{aligned}
0 &= \partial_\delta\left\{ e^{-\kappa\delta}\left(\delta - \mathbb{E}\left[\epsilon^+_{0,z}\right] + h(t,z,q-1) - h(t,z,q)\right)\right\} \\
&= -\kappa e^{-\kappa\delta}\left(\delta - \mathbb{E}\left[\epsilon^+_{0,z}\right] + h(t,z,q-1) - h(t,z,q)\right) + e^{-\kappa\delta} \\
&= e^{-\kappa\delta}\left[-\kappa\left(\delta - \mathbb{E}\left[\epsilon^+_{0,z}\right] + h(t,z,q-1) - h(t,z,q)\right) + 1\right]
\end{aligned}
$$

which implies the term inside the square brackets must equal zero, and hence the optimal depth δ^* is given by:

$$
\boxed{\delta^* = \frac{1}{\kappa} + \Delta_q h + \mathbb{E}\left[\epsilon^+_{0,z}\right]} \tag{5}
$$

where $\Delta_q h = h(t,z,q) - h(t,z,q-1)$. Substituting this optimal depth back into the PDE, we obtain:

$$
\boxed{\begin{aligned}
0 = -\,\varphi q^2 + \partial_t h(t,z,q) &+ \frac{1}{\kappa}\lambda^+(z)e^{-\kappa\left(\frac{1}{\kappa} + \Delta_q h + \mathbb{E}\left[\epsilon^+_{0,z}\right]\right)} \\
&+ \mu(z)q + \sum_{k=-1,0,1} G_{z,k}\left[h(t,k,q) - h(t,z,q)\right]
\end{aligned}} \tag{6}
$$

Thus ends **Part 1**. It is exactly this above system of coupled PDEs on which we'll use some sort of numerical scheme to obtain a solution in **Part 2**. Adios!

Looking Ahead

It's 6pm. I spent the entire day on this, with the exception of going for lunch with Bill. Interestingly we had the worst lab attendance turn out all week today, so just Bill and I chilled at lunch - being a good Chinaman he took me to a sweet place in Chinatown. Anyway, I'm beat. If this doesn't count as > 3 hours of work I don't know what does. I sent a pared down version of this to Sebastian because the results I obtained are different from his and I think he's wrong, though he was quick to deny that that might be the case, though regardless I don't care because it's now in his territory and I can simply chill for the weekend. Farewell old friend!