1 Introduction

This work contains a derivation of the HJB equation for Buy Low Sell High, and a mathematical explanation of the models required to calibrate the work it presented, and HFT models in general. Though BLSH provides a general derivation, I am going to focus on the case where the fill intensities take the form $\lambda \exp^{-k\delta}$. Example code for the super maximum-likelihood esimtator and some sample datasets and fitting will be published on my Github and/or personal site, but the large majority of the code involved in this project will not be published becuase it is specific to the data provider that I am using and the asset that I am trading.

Please note that this is a *work in progress* post that is not in its final form and

2 Background of the Trading Problem

Trading is a complex stocahstic optimal control problem, and to make markets optimally we have to understand how the market evovles, or the dynamicsöf the market. In other words we need to understand how liquidity (our competition) changes over time, how the predictable compnent of the mid-price changes over time, and how the rate at which we receive orders changes over time. We also need to understand how all of these quantities changes when we receive market orders.

2.1 Intuition Behind Orderflow Model

2.2 Intuition Behind Liquidity Model

2.3 Intuition Behind Alpha Model

2.4 Model Dynamics Summary

Summarizing Buy Low Sell High, we model

$$d\lambda^+ = \beta(\theta - \lambda^+)dt + nd\overline{M_t^+} + vd\overline{M_t^-} \tag{1}$$
$$d\lambda^- = \beta(\theta - \lambda^-)dt + vd\overline{M_t^+} + nd\overline{M_t^-} \tag{2}$$
$$d\alpha_t = -\zeta\alpha_t dt + \sigma_\alpha dB_t + \varepsilon^+ d\overline{M_t^+} + \varepsilon^- d\overline{M_t^-} \tag{3}$$

We also have

$$\begin{cases} dk^+ = \beta_k(\theta_k - k^+)dt + n_k d\overline{M_t^+} + v_k d\overline{M_t^-} \\ dk^- = \beta_k(\theta_k - k^-)dt + v_k d\overline{M_t^+} + n_k d\overline{M_t^-} \end{cases} \tag{4}$$

3 Super-MLE Derivation

Our objective is going to be to derive MLE estimators for each of our processes individually, and then to combine them in such a way that allows to get joint estimates for all of our parameters.

Multivariate Hawkes Process

We have the coupled SDE's

$$\begin{cases} d\lambda^+ = \beta(\theta - \lambda^+)dt + nd\overline{M_t^+} + vd\overline{M_t^-} \\ d\lambda^- = \beta(\theta - \lambda^-)dt + vd\overline{M_t^+} + nd\overline{M_t^-} \end{cases} \tag{5}$$

Heuristic interpretation of the conditional intensity function:

$$\lambda^*(t)dt = \frac{f^*(t|\mathscr{H}_{t_n})dt}{1-F^*(t|\mathscr{H}_{t_n})} \tag{6}$$

$$= \frac{\mathbb{P}(t_{n+1} \in [t, t+dt]|\mathscr{H}_{t_n})}{\mathbb{P}(t_{n+1} \notin (t_n, t)|\mathscr{H}_{t_n}))} \tag{7}$$

$$= \frac{\mathbb{P}(t_{n+1} \in [t, t+dt], t_{n+1} \notin (t_n, t)|\mathscr{H}_{t_n})}{\mathbb{P}(t_{n+1} \notin (t_n, t)|\mathscr{H}_{t_n})} \tag{8}$$

$$= \mathbb{P}(t_{n+1} \in [t, t+dt]|t_{n+1} \notin (t_n, t)|\mathscr{H}_{t_n}) \tag{9}$$

$$= \mathbb{P}(t_{n+1} \in [t, t+dt]|\mathscr{H}_{t_-}) \tag{10}$$

$$= \mathbb{E}[N([t, t+dt))|\mathscr{H}_{t_-}) \tag{11}$$

where $N(A)$ is the number of arrivals falling in an interval A and $f^*(t_{n+1}|\mathscr{H}_{t_n})$ is the conditional density function of the next arrival time given the processes history up to time t_n. So already a joint density can be specified as $\prod_n f^*(t_n|\mathscr{H}_{t_{n-1}})$. Rewriting (1) we have

$$\lambda^*(t) = -\frac{d}{dt}\log(1-F^*(t)) \tag{12}$$

Integrating (7), we see that

$$-\int_{t_k}^{t} \lambda^*(u)du = \log(1-F^*(t)) - \log(1-F^*(t_k)) \tag{13}$$

$$= \log(1-F^*(t)) \tag{14}$$

, as $F^*(t_k) = 0$ since a Hawkes process is simple.

$$\implies F^*(t) = 1 - \exp(-\int_{t_k}^{t} \lambda^*(u)du), f^*(t) = \lambda^*(t)\exp(-\int_{t_k}^{t} \lambda^*(u)du) \tag{15}$$

. Substituting into our expression for the log-likelihood, I get

$$L = \prod_n f^*(t_n) \tag{16}$$

$$= \prod_{i=1}^{k} \lambda^*(t)\exp(-\int_{t_{i-1}}^{t_i} \lambda^*(u)du) \tag{17}$$

$$= (\prod_{i=1}^{k} \lambda^*(t))\exp(-\int_{0}^{t_k} \lambda^*(u)du) \tag{18}$$

. Now remember that since the process is observed from 0 to T, we must add a final term $(1-F^*(T))$, the probability of seeing no hits in $(t_k, T]$ to our product, yielding:

$$L = (\prod_{i=1}^{k} \lambda^*(t))\exp(-\int_{0}^{T} \lambda^*(u)du) \tag{19}$$

for a single-variate point process. To arrive at the likelihood we need we must extend this.

$$\mathscr{L} = (\prod_{i=1}^{k} \lambda^+(t))\exp(-\int_{0}^{T} \lambda^+(u)du)(\prod_{i=1}^{k} \lambda^-(t))\exp(-\int_{0}^{T} \lambda^-(u)du) \tag{20}$$

. Taking the log, we have

$$\log(\mathscr{L}) = \sum_{i=1}^{k}\log(\lambda_{t_i}^+) + \sum_{j=1}^{n}\log(\lambda_{t_j}^-) - \int_{0}^{T} \lambda^-(u)du - \int_{0}^{T} \lambda^+(u)du \tag{21}$$

Starting from the definition of our Hawkes process (4), consider a time interval over which no influential trades occur. Over this interval our dynamics are

$$\begin{cases} d\lambda^+ = \beta(\theta - \lambda^+)dt \\ d\lambda^- = \beta(\theta - \lambda^-)dt \end{cases} \tag{22}$$

Integrating this between trade times, we see that

$$\begin{cases} \lambda^+(t_n) = \theta - e^{-\varepsilon(t_n - t_{(n-1)+})}(\theta - \lambda^+(t_{(n-1)} - \mathscr{B}^+_{n-1}(n,v)) \\ \lambda^-(t_n) = \theta - e^{-\varepsilon(t_n - t_{(n-1)+})}(\theta - \lambda^-(t_{(n-1)} - \mathscr{B}^-_{n-1}(n,v)) \end{cases} \tag{23}$$

Where $\mathscr{B}^+_{n-1}(n,v) = n$ if the influential buy order that arrived was an influential buy and $\mathscr{B}^+_{n-1}(n,v) = v$ if it was an influential sell, and vice-versa for $\mathscr{B}^-_{n-1}(n,v)$. Armed with this, we see that

$$\lambda_t^{\pm} = \theta + \sum_{i=1}^{n} H_i^{\pm} e^{-\beta(t-t_i)} \tag{24}$$

, with $H_i^{\pm} = (B_i n + (1-B_i)v, B_i v + (1-B_i)n)$ where B_i is 1 if the order is a buy and 0 if it is a sell by induction on the trades. Looking at

$$\int_0^t \lambda_t^{\pm}(u)du \tag{25}$$

Subsituting our form for lambda, we get

$$\begin{cases} \int_0^T \lambda^+(u)du = \theta T + \sum_{infBUYS} \frac{-n}{\beta} e^{-\beta(T-t_{buy})} + \sum_{infSELLS} \frac{-v}{\beta} e^{-\beta(T-t_{sell})} \\ \int_0^T \lambda^-(u)du = \theta T + \sum_{infBUYS} \frac{-v}{\beta} e^{-\beta(T-t_{buy})} + \sum_{infSELLS} \frac{-n}{\beta} e^{-\beta(T-t_{sell})} \end{cases} \tag{26}$$

Substituing into our likelihood, we finally get

$$\log(\mathscr{L}) = \sum_{i=1}^{k} \log(\lambda^+_{t_i}) + \sum_{j=1}^{n} \log(\lambda^-_{t_j}) - 2\theta T + \sum_{infTRADES} \frac{n+v}{\beta} e^{-\beta(T-t_i)} \tag{27}$$

3.1 Ornstein-Uhlenbeck Jump Process

Examine (3). This is a mean-reverting Ornstein-Uhlenbeck process with long term mean 0 and jumps when influential market orders arrive. Das (1), gives a discrete-time MLE estimator for a related process, where

$$dr = k(\theta - r)dt + vdz + Jd\pi(h) \tag{28}$$

, where J plays a role similar to ε except is normally disitrubted with mean μ and variance γ^2 and $d\pi(h)$ a role similar to $\overline{dM_t^{\pm}}$. Das finds the transition probabilities to be

$$\begin{aligned} f[r(s)|r(t)] = q\exp(&\frac{-(r(s)-r(t)-k(\theta-r(t))\Delta t - \mu)^2}{2(v^2\Delta t + \gamma^2)})\frac{1}{\sqrt{2\pi(v^2\Delta t + \gamma^2)}} \\ &+ (1-q)\exp(\frac{-(r(s)-r(t)-k(\theta-r(t))\Delta t)^2}{2(v^2\Delta t)})\frac{1}{\sqrt{2\pi(v^2\Delta t)}} \end{aligned} \tag{29}$$

Modifying Das' work, I treat ε similar to J, except ε is a constant and does not follow a distribution. Additionally, it is key to note that unlike in the interest rate case, we know exactly when the market orders arrived in general (later I will develop a criterion for showing whether a MO is influential), and conviently the form of our transition probabilties changes. The transitiion probabilties for our α process can be given as $f[\alpha(t)|\alpha(t+\Delta t)]$, where is s is sufficiently close to t such that only one or no orders arrive during the time period (noting that our MO processes are mutlivariate hawkes processes and simple point processes so MOs almost surely never arrive at the same time) are now given by

$$
\begin{aligned}
f[\alpha(t+\Delta t)|\alpha(t)] = & \mathbb{1}M^{+}_{[t,t+\Delta t)} exp(\frac{-(\alpha(s)-\alpha(t)-k(\theta-\alpha(t))\Delta t-\varepsilon)^2}{2\sigma_{\alpha}^2\Delta t})\frac{1}{\sqrt{2\pi\sigma_{\alpha}^2\Delta t}} \\
& + \mathbb{1}M^{-}_{[t,t+\Delta t)} exp(\frac{-(\alpha(s)-\alpha(t)-k(\theta-\alpha(t))\Delta t+\varepsilon)^2}{2\sigma_{\alpha}^2\Delta t})\frac{1}{\sqrt{2\pi\sigma_{\alpha}^2\Delta t}} \\
& + \mathbb{1}M^{0}_{[t,t+\Delta t)} exp(\frac{-(\alpha(s)-\alpha(t)-k(\theta-\alpha(t))\Delta t)^2}{2\sigma_{\alpha}^2\Delta t})\frac{1}{\sqrt{2\pi\sigma_{\alpha}^2\Delta t}}
\end{aligned} \tag{30}
$$

Then the log-likelihood for our model is given by $\log(\prod_{t_i} f[\alpha(t_{i+1})|\alpha(t_i)])$. Notice that in the interest rate case since we don't know exactly when the jumps occurred, this log-likelihood must be maximized numerically. However since we when trades arrive and thus when we expect our jumps to occur, we are left with a log-likelihood that can be maximized by taking the gradient and solving algebraically.

3.2 Orderbook Shape Process

Our orderbook shape process follows the same dynamics as the conditional intensity for the hawkes process. We have

$$
\begin{cases}
dk^{+} = \beta_k(\theta_k - k^{+})dt + n_k d\overline{M_t^{+}} + v_k d\overline{M_t^{-}} \\
dk^{-} = \beta_k(\theta_k - k^{-})dt + v_k d\overline{M_t^{+}} + n_k d\overline{M_t^{-}}
\end{cases} \tag{31}
$$

Using our model for k that we will call $\hat{k}$ (refer to signal measurment, we will look at a bunch of orderbook and trade data as well to describe how liquidity changes as distance from the mid price changes), we seek to find the parameters in (31) to fit our emperical measurements the best. Using our work from the Hawkes Processes, I define the objective

$$
\sum_{OBS}(\hat{k} - \theta_k - \sum_{t_i<t} B_i e^{-\beta(t-t_{iinf})})^2 \tag{32}
$$

3.3 Pareto Distribution

This is a sample way of fitting our last parameter ρ. I model the distribution of trades using a pareto distribution. Given some sample x, the likelihood can be given by

$$
\log\mathscr{L} = \log(\prod_{i=1}^{n}\alpha\frac{\beta^{\alpha}}{x_i^{\alpha+1}}) \tag{33}
$$

$$
= n\log(\alpha) + n\alpha\log(\beta) - (\alpha+1)\sum_{i=1}^{n}\log(x_i) \tag{34}
$$

So we see that our maximum likelihood estimates for α, β are given by setting $\hat{\beta} = \min x_i$, and by differentiating WRT to α we see $\hat{\alpha} = \frac{n}{(\sum_{i=1}^{n} log(x_i)) - n\log(\beta)}$.

4 Super MLE

Combining our work from above, we see that the final objective for our trading algorithm is

$$\log(\mathscr{L}) = \sum_{alphameasurements} (\mathbb{1}M^{+}_{[t,t+\Delta t)}(\frac{-(\alpha(s)-\alpha(t)-k(\theta-\alpha(t))\Delta t-\varepsilon)^2}{2\sigma_\alpha^2\Delta t})\frac{1}{\sqrt{2\pi\sigma_\alpha^2\Delta t}}$$
$$+\mathbb{1}M^{-}_{[t,t+\Delta t)}(\frac{-(\alpha(s)-\alpha(t)-k(\theta-\alpha(t))\Delta t+\varepsilon)^2}{2\sigma_\alpha^2\Delta t})\frac{1}{\sqrt{2\pi\sigma_\alpha^2\Delta t}}$$
$$+\mathbb{1}M^{0}_{[t,t+\Delta t)}(\frac{-(\alpha(s)-\alpha(t)-k(\theta-\alpha(t))\Delta t)^2}{2\sigma_\alpha^2\Delta t})\frac{1}{\sqrt{2\pi\sigma_\alpha^2\Delta t}})$$
$$+\sum_{OBS}(\hat{k}-\theta_k-\sum_{t_i<t}B_i e^{-\beta(t-t_i)})^2+\sum_{i=1}^{k}\log(\lambda^{+}_{t_i})+\sum_{j=1}^{n}\log(\lambda^{-}_{t_j})-2\theta T+\sum_{influentialtrades}\frac{n+\nu}{\beta}e^{-\beta(T-t_i)} \tag{35}$$

. Where the effect of ρ is seen through its changes in the domains of the sums (which trades we label as influential or not.) So one way of performing a full calibration is to do an initial calibration without ρ being considered. Then while holding all parameters constant from this initial calibration, re-calculate the log-likelihood with different values of influential trade size thresholds, until we settle on new parameters and a most likely influential trade size-threshold. Once we have this influential trade size threshold, we can use are estimates of α and β to calculate ρ from the observed trade size distribution. In principle our calculation for ρ can be extended to incldue more factors, including the time since last trade. A more compicated maximization would have to be performed.

5 Signal Measurement

5.1 Trade Size Measurement

To measure ρ over a period $[0,T]$, record the volume of each trade that ocucrs during this period to fit α and β, and then once we have α and β we can compute ρ from the CDF of our pareto distribution and the trade threshold.

5.2 Alpha Process Measurement

α is the componenet of the dynamics of the mid price that we are able to predict. So one way of fitting α is to look at mid-price data for the duration of a time period that we want to trade on, and different features to predict changes in the mid price. Looking at Ait-Sahalia et al. gives us at least several possible sources that I will guess have at least some relevance to cryptocurrency trading as well. Looking at one second intervals, we can either predict one second ahead or same second returns from

1. Orderbook imbalance on the first n-levels
2. Aggregate trading volume in the past n-seconds
3. Aggregate signed trading volume in the past n-seconds
4. Number of trades in the past n-seconds (breadth)
5. Average per-trade volume during the past n-seconds
6. Quoted-spread
7. Immeadiacy ($\frac{1}{breadth}$)
8. Max-trade volume in the past n seconds
9. Past return during the last period

10. Transaction imbalance

Performing a linear regression on same period returns with these independent produces a fairly high R^2 of 0.2 - 0.5 which is to be expected because trades drive the changes in price. Whereas at least for cryptocurrency data this regression produces an R^2 in the low to mid single digits. To generate the signals that we put into our $d\alpha$ estimator, first fit this linear regression on historicla data (this regression could be refitted periodically, the best scheme to do so should be determined through a backtest). Then using our α signals and our trade time indicators, we go through the process highlighted above to fit our super-MLE estimator.

5.3 K Process Measurement

We ultimately want to determine the impact of the arrival of orders on the probability that our quotes get filled. We need to be able to update our quotes immeadiately, so need to have some idea of the impact that a MO arrival has on liquidity. The first step of doing this is getting estimates for k^+ and k^- over periods that are long enough that enogh orders arrive, but short enough to allow intra-hourly changes. I did my fitting on a 5 minute interval. One way of estimating k from this data is to record hits at two deltas δ_1 and δ_2, and solve the system of equations that is presented in source.

$$\begin{cases} \lambda_n(\delta_1) = Ae^{-k\delta_1} \\ \lambda_n(\delta_2) = Ae^{-k\delta_2} \end{cases} \tag{36}$$

An additional way of doing this is to record fills at more deltas, and then to perform a linear regression / least squares fitting of $Ae^{-k\delta}$ to the imperical lambdas. Over each 5 minute interval, record measurements for k using this process. Over these same intervals, we also want to include several other quantities that we expect to have an impact on the value of k that we meausre. We want to record:

1. The average spread during this interval
2. Bid / ask density
3. The coefficients a, b in $v(\delta) = a\delta^{\pm} + b$, where v denotes the volume of orders present on the order book.

These are some rudimentary features that we can extract from the orderbook that describe it's shape. Intuitively, we expect $\hat{k}$ to be inversely related to spread. I also expect a roughly linear relationship between bid and ask density, and $a^{\pm}, b^{\pm}$ and $\hat{k}$. Performing a rudimentary least-squares curve fitting, I get a high R^2 of need to find the R^2. Using these same features, a residual neural network can be fitted to predict the residuals of this model, causing a marginal but measurable improvement in R^2. Ultimately, we can perform this same procedure without the rudimentary feature extraction and feed entire orderbooks (or top n levels) of the orderbook to some neural network to predict $\hat{k}$ in a similar fashion. Once we have this model fitted, we will fit our super MLE estimator on its outputs.

6 Hamilton-Jacobi-Bellman Derivation

Market making is a stochastic optimal control problem where we aim to maximize a functional of the form

$$V(x,t) = \max_{\delta_a \delta_b} \mathbb{E}\Big[\int_t^T L(t, x, \delta_a, \delta_b) dt + V(T, \cdot)\Big] \tag{37}$$

. Many market-making papers have no running-cost, but the paper we are considering has a value function of the form

$$\Phi(t, X_t, S_t, q_t, \alpha_t, \boldsymbol{\lambda}_t, \boldsymbol{k}_t) = \sup_{\delta_u^+, \delta_u^- t \leq u \leq T \in \mathscr{A}} \mathbb{E}\Big[X_T + q_T S_T - \phi \int_t^T q_s^2 ds | \mathscr{F}_t\Big] \tag{38}$$

Since market making has optimal substructure, we can write

$$\Phi(t, X_t, S_t, q_t, \alpha_t, \boldsymbol{\lambda}_t, \boldsymbol{k}_t) = \sup_{\delta_u^+, \delta_u^- t \leq u \leq T \in \mathscr{A}} \mathbb{E}\left[\int_t^{t+\Delta t} -\phi q_s^2 ds + \Phi(t + \Delta t, X_{t+\Delta t}, S_{t+\Delta t}, q_{t+\Delta t}, \alpha_{t+\Delta t}, \boldsymbol{\lambda}_{t+\Delta t}, \boldsymbol{k}_{t+\Delta t})\right] \tag{39}$$

Subtracting Φ from both sides, dividing by Δt, and taking the limit as Δt goes to zero / computing the stochastic derivative / using Ito's Lemma, we are left with the Hamilton Jacobi Bellman equation presented in BLSH.

6.1 Resources

1. Das, Sanjiv R. "The surprise element: Jumps in interest rates." Journal of Econometrics, vol. 106, no. 1, 2002, pp. 27–65, https://doi.org/10.1016/s0304-4076(01)00085-9.

2. Rasmussen, Jakob Gulddahl. "Lecture Notes: Temporal Point Processes and the Conditional Intensity Function." arXiv.Org, 1 June 2018, arxiv.org/abs/1806.00221.

3. Laub, Patrick J., et al. "Hawkes Processes." arXiv.Org, 10 July 2015, arxiv.org/abs/1507.02822.

4. Aït-Sahalia, Yacine, et al. "How and When Are High-Frequency Stock Returns Predictable?" NBER, 22 Aug. 2022, www.nber.org/papers/w30366.

5. Cartea, Álvaro, et al. "Buy Low Sell High: A High Frequency Trading Perspective." SSRN, 26 Nov. 2011, `deliverypdf.ssrn.com/delivery.php?ID=70102006811110210409411909407311510605005102600703401000 amp;EXT=pdf&INDEX=TRUE.`

6. Sophie Laruelle - Events.Chairefdd.Org, `events.chairefdd.org/wp-content/uploads/2013/06/CAHIER_MICRO_1.pdf`. Accessed 18 Oct. 2023.