Limit Order Book Imbalance

Anton Rubisov

2014.07.17

Research Direction

To predict future stock price movement based on current price and order book imbalance. That is,

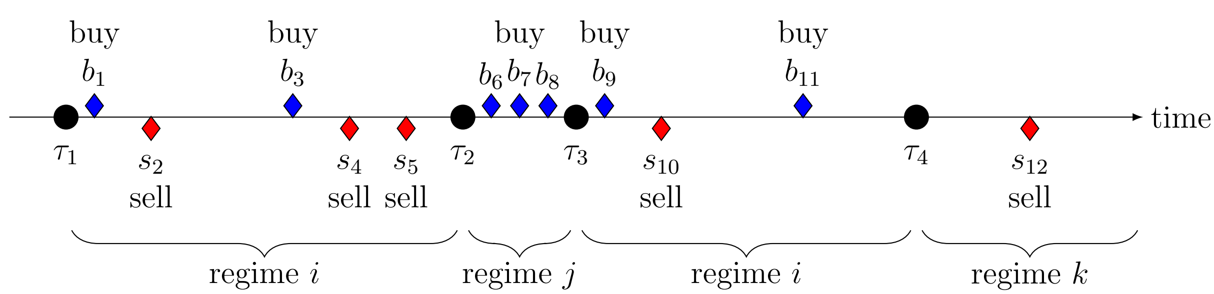
$$\dot{p} = f(p, I)$$

Framework

Limit order book imbalance is a ratio of limit order volumes between the bid and ask side, and can be calculated for example as $I(t) = \dfrac{V_b(t) - V_a(t)}{V_b(t) + V_a(t)} \in [-1, 1]$.

- We bin the bid/ask volume imbalances in the Limit Order Book into K bins, each being dubbed a "regime" of the limit order book.
- Z_t is a continuous-time Markov chain that tracks which regime we're in. Z_t takes values in $\{1, \ldots, K\}$, and has an infinitesimal generator matrix G.
- Conditional on being in some regime k, the arrival of buy and sell market orders follow independent Poisson processes with intensities $\lambda_k^{\pm}$.

We have observations of arrivals of buy/sell market orders and of regime switches occurring, all of which are timestamped. Pictorially, a timeline might look like:



Maximum Likelihood Estimation of G

Let G be the generator matrix for Z_t, so $G = \{q_{ij}\} \in \mathbb{R}^{K \times K}$ where q_{ij} are the transition rates from regime i to regime j for $i \neq j$, and $q_{ii} = -\sum_{j \neq i} q_{ij}$ so that the rows of G sum to 0.

When Z_t enters regime i, the amount of time it spends in regime i is exponentially distributed with rate $v_i = \sum_{j \neq i} q_{ij}$, and when it leaves regime i it will to go regime j with probability $p_{ij} = \dfrac{q_{ij}}{v_i}$.

From our observations we want to estimate the components of G. The holding time in a given regime i is exponentially distributed with pdf $f(t; v_i) = v_i e^{-v_i t}$. For the fictional events in the timeline above, the likelihood function (allowing for repetition of terms) would therefore be:

$$
\begin{aligned}
\mathcal{L}(G) &= (v_i e^{-v_i(\tau_2-\tau_1)} p_{ij})(v_j e^{-v_j(\tau_3-\tau_2)} p_{ji})(v_i e^{-v_i(\tau_4-\tau_3)} p_{ik}) \ldots \\
&= \prod_{i=1}^{K} \prod_{i \neq j} (v_i p_{ij})^{N_{ij}(T)} e^{-v_i H_i(T)} \\
&= \prod_{i=1}^{K} \prod_{i \neq j} (q_{ij})^{N_{ij}(T)} e^{-v_i H_i(T)}
\end{aligned}
$$

where:

$$
\begin{aligned}
N_{ij}(T) &\equiv \text{number of transitions from regime } i \text{ to } j \text{ up to time } T \\
H_i(T) &\equiv \text{holding time in regime } i \text{ up to time } T
\end{aligned}
$$

So that the log-likelihood becomes:

$$
\begin{aligned}
\ln \mathcal{L}(G) &= \sum_{i=1}^{K} \sum_{i \neq j} \left[N_{ij}(T) \ln(q_{ij}) - v_i H_i(T) \right] \\
&= \sum_{i=1}^{K} \sum_{i \neq j} \left[N_{ij}(T) \ln(q_{ij}) - \left(\sum_{i \neq k} q_{ik} H_i(T) \right) \right]
\end{aligned}
$$

To get a maximum likelihood estimate $\hat{q}_{ij}$ for transition rates and therefore the matrix G, we take the partial derivative of $\ln \mathcal{L}(G)$ and set it equal to zero:

$$
\begin{aligned}
\frac{\partial \ln \mathcal{L}(G)}{\partial q_{ij}} &= \frac{N_{ij}(T)}{q_{ij}} - H_i(T) = 0 \\
\Rightarrow \hat{q}_{ij} &= \frac{N_{ij}(T)}{H_i(T)}
\end{aligned}
$$

Maximum Likelihood Estimation of $\lambda_k^{\pm}$

Now we want to derive an estimate for the intensity of the Poisson process of market order arrivals conditional on being in some regime k. We'll look first at just the market buys for some regime k. In the above timeline, the market order buy arrival times are indexed by b_i. Since we're assuming that the arrival process is Poisson with the same intensity throughout trials, we can consider the inter-arrival time of events conditional on being in regime k. Then the MLE derivation follows just as for the CTMC:

$$\begin{aligned}\mathcal{L}(\lambda_k^+; b_1, \ldots, b_N) &= \prod_{i=2}^{N} \lambda_k^+ e^{-\lambda_k^+(b_i - b_{i-1})} \\ &= (\lambda_k^+)^{N_k^+(T)} e^{-\lambda_k^+ H_k(T)}\end{aligned}$$

where:

$$\begin{aligned}N_k^+(T) &\equiv \text{number of market order arrivals in regime } k \text{ up to time } T \\ H_k(T) &\equiv \text{holding time in regime } k \text{ up to time } T\end{aligned}$$

So that the log-likelihood becomes:

$$\ln \mathcal{L}(\lambda_k^+) = N_k^+(T) \ln(\lambda_k^+) - \lambda_k^+ H_k(T)$$

And the ML estimate for $\hat{\lambda}_k^+$ is:

$$\frac{\partial \ln \mathcal{L}}{\partial \lambda_k^+} = \frac{N_k^+(T)}{\lambda_k^+} - H_k(T) = 0$$

$$\Rightarrow \hat{\lambda}_k^+ = \frac{N_k^+(T)}{H_k(T)}$$