Homework #8

Junwu Zhang
CME 241: Reinforcement Learning for Finance

January 31, 2020

Problem 1.
Work out (in LaTeX) the solution to the Linear Impact model we covered in class

Solution.
First of all, there are a number of terms that we should discuss and define here.

The Trading Order Book (TOB) describes the relationship between volume, price per share, and the type of limit order (LO). A typical TOB is shown in Figure 1 below.

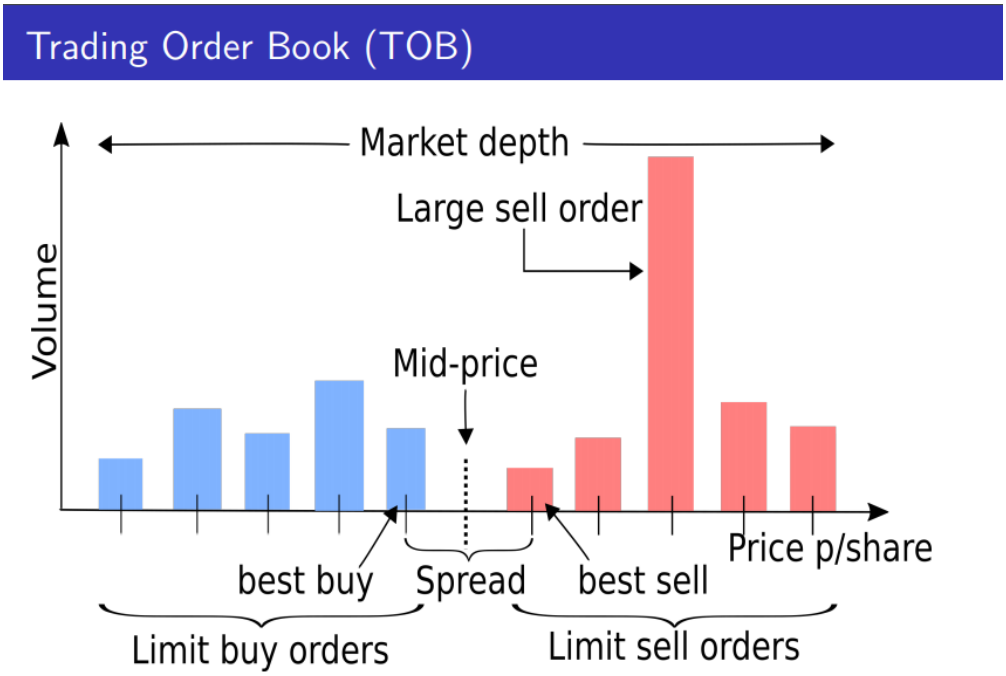

Figure 1: Typical Trading Order Book (TOB)

A Market Order (MO) alters the TOB, and *price impact* refers to how large-scale MO moves the bid/ask/mid price of shares in the market. In this problem, we're looking at a linear price impact model, where the differences in MO and prices are linearly related.

Second of all, we realize that we can formulate the entire optimal trade order execution problem as an Markov Decision Process (MDP). Our overall goal is to sell a large amount N of shares, where trading happens in T discrete time steps. Only MOs are allowed, and we need to balance the speed of selling with the market price of the shares. More specifically, the MDP can be constructed as following:

States are (t, P_t, R_t) where t represents the time step, P_t denotes Bid Price at start of time step t, and $R_t = N - \sum_{i=1}^{t-1} N_i$ denotes shares remaining at start of time step t

Actions is N_t which denotes the number of shares sold in time step t

Reward is $U(N_t \cdot Q_t)$

Price transition dynamics is $P_{t+1} = f_t(P_t, N_t, \epsilon_t)$

Given all these elements, the objective is to find a policy $\pi^*(t, P_t, R_t)$ that maximizes:

$$\mathbb{E}[\sum_{t=1}^{T} \gamma^t \cdot U(N_t \cdot Q_t)] \tag{1}$$

where γ is the discount factor of the MDP.

Once we have such MDP problem definition, we can consider a model with linear price impact, where N, N_t, P_t are all continuous. The price dynamics is given by:

$$P_{t+1} = P_t - \alpha N_t + \epsilon_t \tag{2}$$

where $\alpha \in \mathbb{R}_{\leq 0}$ and ϵ_t is i.i.d. with $\mathbb{E}[\epsilon_t | N_t, P_t] = 0$.

As a first step to solving this, we can denote the value function with policy π as:

$$V^{\pi}(t, P_t, R_t) = \mathbb{E}_{\pi}\left[\sum_{i=t}^{T} N_i \left(P_i - \beta N_i\right) | \, (t, P_t, R_t)\right] \tag{3}$$

Since optimal value function can be denoted as $V^*(t, P_t, R_t) = \max_{\pi} V^{\pi}(t, P_t, R_t)$, we can expand it to write:

$$V^*(t, P_t, R_t) = \max_{N_t}\Big(N_t \left(P_t - \beta N_t\right) + \mathbb{E}\left[V^*(t+1, P_{t+1}, R_{t+1})\right]\Big) \tag{4}$$

We can infer $V^*(T-1, P_{T-1}, R_{T-1})$ as:

$$\max_{N_{T-1}}\Big\{N_{T-1}\left(P_{T-1} - \beta N_{T-1}\right) + \mathbb{E}\left[R_T\left(P_T - \beta R_T\right)\right]\Big\} \tag{5}$$

$$= \max_{N_{T-1}}\Big\{R_{T-1}P_{T-1} - \beta R_{T-1}^2 + (\alpha - 2\beta)\left(N_{T-1}^2 - N_{T-1}R_{T-1}\right)\Big\} \tag{6}$$

Taking derivative and setting to zero, we have:

$$(\alpha - 2\beta)\left(N_{T-1}^2 - N_{T-1}R_{T-1}\right) = 0 \rightarrow N_{T-1}^* = \frac{R_{T-1}}{2} \tag{7}$$

Note that this is for the non-trivial case $\alpha < 2\beta$.

Plug the result from Equation (7) in the expression for V^*, we have:

$$V^*(T-1, P_{T-1}, R_{T-1}) = R_{T-1}P_{T-1} - R_{T-1}^2\left(\frac{\alpha+2\beta}{4}\right) \tag{8}$$

Similarly, continuing backwards in time, we have:

$$\begin{aligned} N_t^\star &= \tfrac{R_t}{T-t+1} \\ V^*(t, P_t, R_t) &= R_tP_t - \tfrac{R_t^2}{2}\left(\tfrac{2\beta+(T-t)\alpha}{T-t+1}\right) \end{aligned} \tag{9}$$

■

Problem 2.
Model a real-world Optimal Trade Order Execution problem as an MDP

Solution. The MDP can be modeled as:

- *States*: $(t, P_{b_t}, P_{s_t}, R_{b_t}, R_{s_t})$ where t represents the time step, P_{b_t} denotes buy price at start of time step t, P_{s_t} denotes sell price at start of time step t, R_{b_t} denotes shares available to buy at start of time step t, and R_{s_t} denotes shares available to sell at start of time step t
- *Actions*: N_t which denotes the number of shares sold in time step t
- *Reward*: $U(N_t \cdot Q_t)$
- *Price Impact*: In real-world scenario, price impact might be purely temporary and can be represented as:

$$P_{t+1} = P_te^{Z_t}, X_{t+1} = \rho X_t + \eta_t, Q_t = P_t(1 - \alpha N_t - \theta X_t) \tag{10}$$

■