Trade Order Execution Problem

Miao-Chin Yen

1 General Problem Formulation

Our task is to sell N shares by submitting orders in T discrete time steps to maximize the expected total utility of sales proceeds. For simplicity, our model only considers best bid price dynamics and we are only allowed to submit market orders (MOs). We then first give some notations as follows:

- $t = 0, 1, ..., T$ are time steps.
- P_t is the best bid price at the start of time step t.
- N_t is the number of shares sold in time step t.
- $R_t = N - \sum_{i=0}^{t-1} N_i$ is the number of shares remaining to be sold at the start of time step t.
- $P_{t+1} = f_t(P_t, \cdots)$ represents the dynamic of the best bid price.
- $N_t \cdot Q_t$ is the sales proceeds at time step t.
- $U(\cdot)$ is the utility function.

Note that the sales proceeds is not $N_t \cdot P_t$ since there is a temporary price impact effect. Q_t is specified to characterize this effect.

We can model this problem into a discrete time and finite horizon MDP with horizon T. Notice that all states at time step T are terminal states. We use backward induction technique to solve this problem. The formulation of MDP is as follows (for $0 \leq t < T$):

- State $s_t := (P_t, R_t) \in S_t$
- Action $a_t := N_t \in A_t$
- Reward $r_{t+1} := U(N_t \cdot Q_t)$
- Price Dynamics $P_{t+1} = f_t(P_t, \cdots)$

Our goal is to find an optimal policy $\pi^* = (\pi_t^*)_{0 \leq t < T}$ s.t

$$\pi^* = \arg\max_\pi E[\sum_{t=0}^{T-1} \gamma^t \cdot U(N_t \cdot Q_t)] \qquad \text{where } \pi_t^*(P_t, R_t) = N_t^*$$

and γ is the MDP discount factor. We then define optimal value function and invoke Bellman equation. We consider two models with different price dynamics in this report. The first one is a simple linear impact model with no risk-aversion.

2 Simple Linear Impact Model with No Risk-Aversion

First, we give the formulation of this scenario:

- State $s_t := (P_t, R_t) \in S_t$
- Action $a_t := N_t \in A_t$

- Reward $r_{t+1} := U(N_t \cdot Q_t) = U(N_t \cdot (P_t - \beta \cdot N_t))$ where $\beta \in \mathbb{R} \geq 0$. $\beta \cdot N_t$ is the temporary price impact.
- Price Dynamics $P_{t+1} = f_t(P_t, N_t, \epsilon_t) = P_t - \alpha N_t + \epsilon_t$ where $\alpha \in \mathbb{R}$ and ϵ_t is i.i.d. with $\mathbb{E}[\epsilon_t | N_t, P_t] = 0$. f_t is a function consisting of three elements:
  1. Permanent price impact αN_t: price impact of selling N_t shares
  2. Impact-independent market movement of best bid price for time step t: P_t
  3. Source of randomness: ϵ_t
- Utility function $U(x) = x$ is no risk-aversion function.
- Discount factor $\gamma = 1$

We then denote the value function for policy π as:

$$V_t^\pi((P_t, R_t)) = \mathbb{E}_\pi \left[\sum_{i=t}^{T} N_i \cdot (P_i - \beta \cdot N_i) \mid (P_t, R_t) \right]$$

and optimal value function as $V_t^*((P_t, R_t)) = \max_\pi V_t^\pi((P_t, R_t))$
Optimal value function satisfies the Bellman equation ($\forall\ 0 \leq t < T-1$) :

$$V_t^*((P_t, R_t)) = \max_{N_t} \left\{ N_t \cdot (P_t - \beta \cdot N_t) + \mathbb{E}\left[V_{t+1}^*((P_{t+1}, R_{t+1})) \right] \right\}$$

$$V_{T-1}^*((P_{T-1}, R_{T-1})) = N_{T-1} \cdot (P_{T-1} - \beta \cdot N_{T-1}) = R_{T-1} \cdot (P_{T-1} - \beta \cdot R_{T-1})$$

We then infer $V_{T-2}^*((P_{T-2}, R_{T-2}))$ as

$$\max_{N_{T-2}} \{ N_{T-2}(P_{T-2} - \beta N_{T-2}) + \mathbb{E}[R_{T-1}(P_{T-1} - \beta R_{T-1})] \}$$

$$= \max_{N_{T-2}} \{ N_{T-2}(P_{T-2} - \beta N_{T-2}) + \mathbb{E}[(R_{T-2} - N_{T-2})(P_{T-1} - \beta(R_{T-2} - N_{T-2}))] \}$$

$$= \max_{N_{T-2}} \{ N_{T-2}P_{T-2} - \beta N_{T-2}^2 + \mathbb{E}[(R_{T-2} - N_{T-2})(P_{T-2} - \alpha N_{T-2} + \epsilon_{T-2} - \beta R_{T-2} + \beta N_{T-2})] \}$$

$$= \max_{N_{T-2}} \{ N_{T-2}P_{T-2} - \beta N_{T-2}^2 + R_{T-2}P_{T-2} - \alpha R_{T-2}N_{T-2} - \beta R_{T-2}^2 + \beta N_{T-2}R_{T-2} - N_{T-2}P_{T-2}$$
$$+ \alpha N_{T-2}^2 + \beta N_{T-2}R_{T-2} - \beta N_{T-2}^2 \}$$

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

We then divide it into two subcases:

- If $\alpha \geq 2\beta$, we have the solution $N_{T-2}^* = 0$ or R_{T-2} and the optimal value function for time $T-2$ is:

  $$V_{T-2}^*((P_{T-2}, R_{T-2})) = R_{T-2}(P_{T-2} - \beta R_{T-2})$$

  The continuing backwards in time in this manner gives:

  $$N_t^* = 0 \text{ or } R_t$$
  $$V_t^*((P_t, R_t)) = R_t(P_t - \beta R_t)$$

  Hence, the optimal strategy for the case $\alpha \geq 2\beta$ is to sell all N shares at any one of the time steps $t = 0, \ldots, T-1$ and the optimal expected total sale proceeds is $N(P_0 - \beta N)$.
- If $\alpha < 2\beta$, we utilize first order condition (F.O.C) w.r.t. N_{T-2}:

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

We substitute N^*_{T-2} in the expression for $V^*_{T-2}\left((P_{T-2}, R_{T-2}\right)$:

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

We then infer one more term $V^*_{T-3}\left((P_{T-3}, R_{T-3})\right)$ to find the solution pattern:

$$\max_{N_{T-3}}\{N_{T-3}(P_{T-3} - \beta N_{T-3}) + E\left[R_{T-2}P_{T-2} - \frac{\alpha + 2\beta}{4}R^2_{T-2}\right]\}$$

$$= \max_{N_{T-3}}\{N_{T-3}P_{T-3} - \beta N^2_{T-3} + E[(R_{T-3} - N_{T-3})(P_{T-3} - \alpha N_{T-3} + \epsilon_{T-3}) - \frac{\alpha + 2\beta}{4}(R_{T-3} - N_{T-3})^2\}$$

$$= \max_{N_{T-3}}\{N_{T-3}P_{T-3} - \beta N^2_{T-3} + R_{T-3}P_{T-3} - \alpha R_{T-3}N_{T-3} - N_{T-3}P_{T-3} + \alpha N^2_{T-3}$$

$$-\frac{\alpha + 2\beta}{4}(R^2_{T-3} - 2R_{T-3}N_{T-3} + N^2_{T-3})\}$$

$$= \max_{N_{T-3}}\{R_{T-3}P_{T-3} + \frac{2\beta - \alpha}{2}R_{T-3}N_{T-3} + (\alpha - \beta - \frac{\alpha + 2\beta}{4})N^2_{T-3} - \frac{\alpha + 2\beta}{4}R^2_{T-3}\}$$

$$= \max_{N_{T-3}}\{R_{T-3}P_{T-3} - \frac{2\beta + \alpha}{4}R^2_{T-3} + \frac{2\beta - \alpha}{2}R_{T-3}N_{T-3} + N^2_{T-3}(\frac{3\alpha - 6\beta}{4})\}$$

We utilize F.O.C w.r.t. N_{T-3} and substitute N^*_{T-3} in the expression for $V^*_{T-3}\left((P_{T-3}, R_{T-3})\right)$:

$$N_{T-3} = \frac{R_{T-3}(\alpha - 2\beta)}{3\alpha - 6\beta} = \frac{R_{T-3}}{3}$$

$$V^*_{T-3}(P_{T-3}, R_{T-3}) = R_{T-3}P_{T-3} + R^2_{T-3}(\frac{-\alpha - \beta}{3})$$

The continuing backwards in time in this manner gives:

$$N^*_t = \frac{R_t}{T - t}$$

$$V^*_t\left((P_t, R_t)\right) = R_tP_t - \frac{R^2_t}{2}\left(\frac{2\beta + \alpha(T - t - 1)}{T - t}\right)$$

Hence, if we roll forward in time, we notice that the optimal strategy at time t is to uniformly split the remaining shares for selling.

We further study this problem with linear percentage temporary price impact model.

3 Linear Percentage Temporary (LPT) Price Impact Model with No Risk-Aversion

LPT is a model with purely temporary price impact. The formulation of this scenario is as follows:

- State $s_t := (P_t, R_t, X_t) \in S_t$
  1. We can view X_t as a variable providing personal information or market conditions.
- Action $a_t := N_t \in A_t$
- Reward $r_{t+1} := U(N_t \cdot Q_t) = U(N_t \cdot P_t \cdot (1 - \beta \cdot N_t - \theta \cdot X_t))$
  1. Q_t represents percentage price impact.
  2. $X_{t+1} = \rho \cdot X_t + \eta_t$ where η_t are independent and identically distributed random variables with mean 0 for all $t = 0, 1, \ldots, T - 1$.
- Price Dynamics $P_{t+1} = P_t \cdot e^{Z_t}$ where Z_t are independent and identically distributed random variables with mean μ_Z and variance σ^2_Z for all $t = 0, 1, \ldots, T - 1$,

- Utility function $U(x) = x$ is no risk-aversion function.
- Discount factor $\gamma = 1$

Note that Z_t and η_t are independent of each other for all $t = 0, 1, \ldots, T-1$, and ρ, β, θ are given constants. We then denote the value function for policy π as:

$$V_t^\pi\left((P_t, X_t, R_t)\right) = \mathbb{E}_\pi\left[\sum_{i=t}^{T} N_i \cdot P_i\left(1 - \beta \cdot N_i - \theta \cdot X_i\right) \mid (P_t, X_t, R_t)\right]$$

and optimal value function as $V_t^*\left((P_t, X_t, R_t)\right) = \max_\pi V_t^\pi\left((P_t, X_t, R_t)\right)$
Optimal value function satisfies the Bellman equation ($\forall\ 0 \le t < T-1$) :

$$V_t^*\left((P_t, X_t, R_t)\right) = \max_{N_t}\left\{N_t \cdot P_t\left(1 - \beta \cdot N_t - \theta \cdot X_t\right) + \mathbb{E}\left[V_{t+1}^*\left((P_{t+1}, X_{t+1}, R_{t+1})\right)\right]\right\}$$

$$V_{T-1}^*\left((P_{T-1}, X_{T-1}, R_{T-1})\right) = N_{T-1} \cdot P_{T-1}\left(1 - \beta \cdot N_{T-1} - \theta \cdot X_{T-1}\right) = R_{T-1} \cdot P_{T-1}\left(1 - \beta \cdot R_{T-1} - \theta \cdot X_{T-1}\right)$$

We then infer $V_{T-2}^*\left((P_{T-2}, X_{T-2}, R_{T-2})\right)$ as

$$\max_{N_{T-2}}\left\{N_{T-2}P_{T-2}\left(1 - \beta N_{T-2} - \theta X_{T-2}\right) + \mathbb{E}\left[R_{T-1}P_{T-1}\left(1 - \beta R_{T-1} - \theta X_{T-1}\right)\right]\right\}$$

$$= \max_{N_{T-2}}\left\{N_{T-2}P_{T-2}\left(1 - \beta N_{T-2} - \theta X_{T-2}\right) + \mathbb{E}\left[(R_{T-2} - N_{T-2})P_{T-1}\left(1 - \beta(R_{T-2} - N_{T-2}) - \theta X_{T-1}\right)\right]\right\}$$

Since $P_{T-1} = P_{T-2}e^{Z_{T-2}}$ and $X_{T-1} = \rho X_{T-2} + \eta_{T-2}$, we can rewrite the objective function as

$$\begin{gathered}\max_{N_{T-2}}\{N_{T-2}P_{T-2}\left(1 - \beta N_{T-2} - \theta X_{T-2}\right) + \\ \mathbb{E}\left[(R_{T-2} - N_{T-2})P_{T-2}e^{Z_{T-2}}\left(1 - \beta(R_{T-2} - N_{T-2}) - \theta(\rho X_{T-2} + \eta_{T-2})\right)\right]\}\end{gathered}$$

We further rewrite our objective function based on $\mathbb{E}[e^{Z_{T-2}}] = e^{\mu_z + \frac{{\sigma_z}^2}{2}}$ and $\mathbb{E}[\eta_{T-2}] = 0$ (denote $e^{\mu_z + \frac{{\sigma_z}^2}{2}}$ as q):

$$\begin{gathered}\max_{N_{T-2}}\{(1-q)N_{T-2}P_{T-2} - \beta(1+q)N_{T-2}^2P_{T-2} - \theta(1-\rho q)X_{T-2}P_{T-2}N_{T-2} \\ +qR_{T-2}P_{T-2} - \beta q R_{T-2}^2P_{T-2} + 2\beta q N_{T-2}R_{T-2}P_{T-2} - \rho\theta q X_{T-2}R_{T-2}P_{T-2}\}\end{gathered}$$

We utilize F.O.C w.r.t N_{T-2}:

$$\begin{aligned}P_{T-2}(1-q) - 2\beta(1+q)N_{T-2}^*P_{T-2} - \theta X_{T-2}P_{T-2}(1-\rho q) + 2\beta q R_{T-2}P_{T-2} &= 0 \\ \Rightarrow 2\beta(1+q)N_{T-2}^* = (1-q) + 2\beta q R_{T-2} - \theta(1-\rho q)X_{T-2} \\ \Rightarrow N_{T-2}^* = \frac{(1-q) + 2\beta q R_{T-2} - \theta(1-\rho q)X_{T-2}}{2\beta(1+q)} \\ = \frac{1-q}{2\beta(1+q)} + \frac{q}{1+q}R_{T-2} - \frac{\theta(1-\rho q)}{2\beta(1+q)}X_{T-2} \\ := c_{T-2}^{(1)} + c_{T-2}^{(2)}R_{T-2} + c_{T-2}^{(3)}X_{T-2}\end{aligned}$$

We then substitute N_{T-2}^* in the expression for $V_{T-2}^*((P_{T-2}, X_{T-2}, R_{T-2}))$:

$$\begin{aligned}V_{T-2}^*\left(P_{T-2}, X_{T-2}, R_{T-2}\right) =& P_{T-2}\left[a_{T-2} + b_{T-2}X_{T-2} + c_{T-2}X_{T-2}^2\right. \\ &\left. + d_{T-2}X_{T-2}R_{T-2} + e_{T-2}R_{T-2} + f_{T-2}R_{T-2}^2\right]\end{aligned}$$

where

$$\begin{aligned}a_{T-2} &= (1-q)\,c_{T-2}^{(1)} - \beta\,(1+q)\,{c_{T-2}^{(1)}}^2 \\ b_{T-2} &= (1-q)c_{T-2}^{(3)} - \theta\,(1-\rho q)\,c_{T-2}^{(1)} - 2\beta\,(1+q)\,c_{T-2}^{(1)}c_{T-2}^{(3)} \\ c_{T-2} &= -\theta\,(1-\rho q)\,c_{T-2}^{(3)} - \beta\,(1+q)\,{c_{T-2}^{(3)}}^2 \\ d_{T-2} &= -\theta\,(1-\rho q)\,c_{T-2}^{(2)} - \rho\theta q + 2q\beta c_{T-2}^{(3)} - 2\beta\,(1+q)\,c_{T-2}^{(2)}c_{T-2}^{(3)} \\ e_{T-2} &= (1-q)\,c_{T-2}^{(2)} + q + 2q\beta c_{T-2}^{(1)} - 2\beta\,(1+q)\,c_{T-2}^{(1)}c_{T-2}^{(2)} \\ f_{T-2} &= -q\beta + 2q\beta c_{T-2}^{(2)} - \beta\,(1+q)\,{c_{T-2}^{(2)}}^2\end{aligned}$$

The continuing backwards in time in this manner gives:

$$N^*_{T-k} = c^{(1)}_{T-k} + c^{(2)}_{T-k} R_{T-k} + c^{(3)}_{T-k} X_{T-k}$$

where

$$c^{(1)}_{T-k} = \frac{1 - q e_{T-k+1}}{2\left(\beta - q f_{T-k+1}\right)}, \qquad c^{(2)}_{T-k} = \frac{-2q f_{T-k+1}}{2\left(\beta - q f_{T-k+1}\right)}, \quad c^{(3)}_{T-k} = \frac{-q\rho d_{T-k+1} - \theta}{2\left(\beta - q f_{T-k+1}\right)}.$$

Hence

$$\begin{aligned} V^*_{T-k}\left(P_{T-k}, X_{T-k}, R_{T-k}\right) = & P_{T-k}\left[a_{T-k} + b_{T-k} X_{T-k} + c_{T-k} X^2_{T-k}\right. \\ & \left. + d_{T-k} X_{T-k} R_{T-k} + e_{T-k} R_{T-k} + f_{T-k} R^2_{T-k}\right] \end{aligned}$$

where

$$\begin{aligned} a_{T-k} &= (1-q)\, c^{(1)}_{T-k} - \beta\,(1+q)\, {c^{(1)}_{T-k}}^2 \\ b_{T-k} &= (1-q) c^{(3)}_{T-k} - \theta\,(1-\rho q)\, c^{(1)}_{T-k} - 2\beta\,(1+q)\, c^{(1)}_{T-k} c^{(3)}_{T-k} \\ c_{T-k} &= -\theta\,(1-\rho q)\, c^{(3)}_{T-k} - \beta\,(1+q)\, {c^{(3)}_{T-k}}^2 \\ d_{T-k} &= -\theta\,(1-\rho q)\, c^{(2)}_{T-k} - \rho\theta q + 2q\beta c^{(3)}_{T-k} - 2\beta\,(1+q)\, c^{(2)}_{T-k} c^{(3)}_{T-k} \\ e_{T-k} &= (1-q)\, c^{(2)}_{T-k} + q + 2q\beta c^{(1)}_{T-k} - 2\beta\,(1+q)\, c^{(1)}_{T-k} c^{(2)}_{T-k} \\ f_{T-k} &= -q\beta + 2q\beta c^{(2)}_{T-k} - \beta\,(1+q)\, {c^{(2)}_{T-k}}^2 \end{aligned}$$

Note that

$$\begin{aligned} a_{T-1} &= 0 \\ b_{T-1} &= 0 \\ c_{T-1} &= 0 \\ d_{T-1} &= -\theta \\ e_{T-1} &= 1 \\ f_{T-1} &= -\beta \end{aligned}$$

Therefore, the optimal strategy (N^*_t) is a linear function of two state variables X_t and R_t, and the optimal value function is a quadratic function of X_t and R_t proportional to P_t.

References

[1] A. Rao, T. Jelvis. *Foundations of Reinforcement Learning with Applications in Finance*

[2] D. Bertsimas, A. W. Lo. *Optimal control of execution costs.* In Journal of Financial Markets 1 (1998) 1—50.