Review of Optimal Execution Problem

Xingjian Xue, Yiwen Liu, Yujie Hong, Yuqing Bu

Abstract

We introduce the optimal execution problem, and classify the methods into price impact based models which focus on modelling temporary and permanent impact from the dynamics of market price and microstructure of the market, and order book based models which concentrate on the dynamics of the limit order book system. Nonlinear impact model use a power law temporary impact function giving more influence on the trading speed. From the simulation, we can tell that it is more suitable for rapid execution. Cartea-Jaimungal strategy incorporates order-flow into model, and adjusted linear model by by adding a weighted-average term of future expected net order-flow. Stochastic volatility is a discrete-time model following a coupled and finite Markov chain.And we use the forward-backward equations and Bellman equations to solve the discrete-time model.For three order book based models, The first one models the state of order book as a Markov Chain Process. The second model provides a fluid-diffusion model for discrete-time one-sided limit order book.At last, LSTM is suitable for predicting events with relatively long interstitions and delays in time series.

Contents

1 Introduction

This paper investigates different methods for the optimal execution problem. We classify the recent studies into two main streams: price impact streams and order book streams. We analyse these models and make comparisons under certain conditions.

Either for institutional investors or for traders, when they are trying to execute a large amount of order, execution cost is one of the most important factor that should be taken into consideration. Simply putting large orders into the market can often make the execution cost unacceptable. 'Execution cost' can be divided into the following subjects: commissions, bid-ask spreads, opportunity costs of waiting, and price impacts from trading[11]. Most papers focus on modelling order book and price impact.
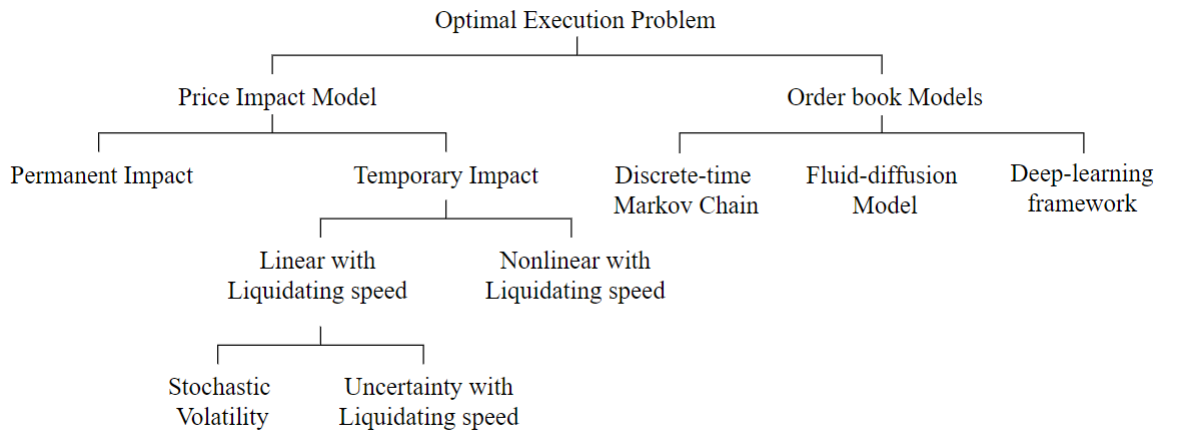
Figure 1: Types of methods

First is price impact. Market price (MO) could have its own dynamics, which can be modelled as a Geometric Brownian Motion in the famous Black-Scholes model[12]. However, it will also be affected by the liquidating process: for a large buy (sell) order, MO often goes up (down) and has fluctuation as the consequence. The basic reason behind this kind of dynamic change of price is that large orders will exhaust the supply of liquidity at each price level. In other word, the best price traders want their large orders not to be filled by the limit order books placed in the market. Besides, the large sell (buy) action provides more information to the market, then other investors may have response. This kind of influence to the price can be divided into two categories: temporary impact and permanent impact. The goal for optimal execution problem is trying to include all execution cost above in the model, especially price impact, and give the optimal execution strategy under performance criteria.

Another perspective is from order book[7]. During trading process, market orders are executed against the best price provided by the limit order book system. With the large data from order book, researchers can build models of high-frequency order book dynamics that contain information about market order flow, security liquidity and price dynamics. The goal for such modelling is

that we can use the information on the current state of the order book to predict its short-term behaviour, the future order size and price. Therefore, we can decide our optimal execution speed for large orders.

The optimal execution problem was first addressed in[3]. In the paper, Bertsimas and Lo built a discrete-time model for price dynamics, and provided one clear definition of best execution, based on the goal of minimizing the expected cost of execution using stochastic dynamic programming. Then in 2001, Almgren and Chriss[1] introduced linear temporary impact function into the model and use the mean-variance for optimization target. After two years, Almgren et.al [2] modelled temporary impact function as a power law function, which gave more balance to the execution speed than the linear model. Li et.al[13] added Poisson process as the dynamic of stochastic price impact factor into the model. From the perspective of limit order book, Rama Cont[7] built a prediction model for LOB by assuming that the dynamic of order book is a Poisson process and can be accelerated by Laplace transform method, which is suitable for high-frequency model.

In section 2, we present four price impact based models: linear impact model, which is also known as Almgren-Chriss execution model, nonlinear impact model, incorporating order-flow execution model and stochastic volatility model. We discuss the construction and the solutions of these models. In section 3, we present three order book based models: Markov chain model, fluid-diffusion model, LSTM neural network model, and introduce the basic ideas about these models. In section 4, we use linear impact model and nonlinear impact model to simulate the solution for 2014 INTC stock data and give some analysis. We give a square root nonlinear temporary impact function. We also plot the efficient frontier for linear case. Finally, in Appendix, we present the codes for simulation.

2   Price Impact Based Model

The first model introduced is a linear price impact (both temporary and permanent) model, with the target of minimizing mean-variance function.

2.1   Linear Impact Model

Almgren-Chriss execution model[1] is a classic discrete-time execution model and has become one of the fundamental framework for related studies in High-Frequency Trading and Algorithm Trading areas. The paper set the goal of execution in a 'mean-variance' way, which is maximizing the expected revenue of trading (minimizing the trading cost, like quadratic utility function or VAR) under an acceptable level of penalty from variance (uncertainty) of the revenue. Temporary and permanent impact in price dynamics are formulated as linear function driven by execution speed, and in the original model there is no drift term in permanent impact by assuming no information is given about future price direction. Finally, the optimal trading trajectories is obtained using static optimization procedures, that means the optimal trading trajectories is determined before trading.

2

2.1.1 Price Dynamics

Given X as the initial units of stocks that will be fully executed at time T, time interval $\tau = T/N$, define discrete time $t_k = k\tau$, for $k = 0, 1, ..., N$, then our inventory is a decreasing list $x_k, x_0 = X$, the number of units stocks that we have at time t_k. Our trading strategy is defined as a list $n_k = x_{k-1} - x_k$. The relationship between x_k and n_k is:

$$x_k = X - \sum_{i=1}^{k} n_i = \sum_{j=k+1}^{N} n_j, \ k = 0, ..., N. \tag{1}$$

Suppose the initial stock price is S_0, the price dynamics can be influenced in two main factors: its own volatility and drift, and market impact, temporary and permanent, which caused by our execution action.

As for permanent impact, it is because our large order come into market, the initial balance is broken and is shifting to the new 'equilibrium'. We can formulate this kind of evolution in following discrete-time equation:

$$S_k = S_{k-1} + \sigma\sqrt{\tau}\xi_k - \tau g(v) \tag{2}$$

Here, σ is the volatility of the stock, τ is the time interval, ξ_k is a random variable follows normal distribution $\mathcal{N}(0, 1)$, $v = \frac{n_k}{\tau}$ is the liquidate speed, $g(v)$ denotes the permanent impact function. We may add the drift term later, but here in this model, there is no drift term, which means that we do no make any assumption about future stock price moving direction.

In this model, the permanent impact function is a linear function, given by:

$$g(v) = \gamma v \tag{3}$$

γ is a constant, describes that every unit we sell will depress the current price. So Eq. 2 can be written as:

$$S_k = S_{k-1} + \sigma\sqrt{\tau}\xi_k - \tau\gamma v \tag{4}$$

As for temporary impact, it is caused because the bid price and order size in limit order book can not always fit our large order, so the liquidity will be exhausted, we may assume this influence will only lasting for a short period. Given liquidate speed v, the temporary impact can be formulated by:

$$\tilde{S}_k = S_{k-1} - h(v) \tag{5}$$

Here, for temporary impact function $h(v)$, we take:

$$h(v) = h(\frac{n_k}{\tau}) = \epsilon sgn(n_k) + \frac{\eta}{\tau}n_k = \epsilon sgn(n_k) + \eta v \tag{6}$$

ϵ is the estimation of fixed cost of selling, such as commissions. η is a coefficient from market microstructure. $sgn(n_k)$ is a signum function: if $n_k > 0$, its value is 1, if $n_k < 0$, its value is -1, obviously, in a pure sell execution, $sgn(n_k) \equiv 1$.

2.1.2 Mean-variance Performance Criterion

Since we have modelled the two kinds of market impact, we can now calculate the cost during whole execution period. We can define the *capture* of a trading trajectory, which means the total revenue after the final trade. We can calculate it by summing all trading units times the execution price, $\sum_{k=1}^{N} n_k \tilde{S}_k$. Our initial asset value is XS_0, so the execution cost (which is also known as implementation shortfall[14]) is:

$$XS_0 - \sum_{k=1}^{N} n_k \tilde{S}_k = -\sum_{k=1}^{N} (\sigma\sqrt{\tau}\xi_k - \tau g(v))x_k + \sum_{k=1}^{N} (n_k h(v)) \tag{7}$$

Using $E(x)$ and $V(x)$ to denote the expectation and the variance of execution cost. We can then get:

$$E(x) = \sum_{k=1}^{N} (\tau g(v)x_k) + \sum_{k=1}^{N} (n_k h(v)) \tag{8}$$

$$V(x) = \sigma^2 \sum_{k=1}^{N} (\tau x_k^2) \tag{9}$$

We can plug Eq. 3 and Eq. 6 into Eq. 8 then yields:

$$E(x) = \frac{1}{2}\gamma X^2 + \epsilon \sum_{k=1}^{N} |n_k| + \frac{\eta - 1/2\gamma\tau}{\tau} \sum_{k=1}^{N} n_k^2 \tag{10}$$

Our mean-variance criterion is to minimize

$$\min_x U(x) = E(x) + \lambda V(x) \tag{11}$$

for any different values of λ. λ here can be interpreted as a measure of risk-aversion, which shows the penalty on variance of execution cost. And the result is that there will be an unique trading trajectory n_k that for every λ, $E(x)+\lambda V(x)$ is minimal. $E(x)$ is strict convex, $V(x)$ is convex. We assume in the execution process, there is no buying action, so for a pure selling execution, $n_k \geq 0$.

2.1.3 Solutions

Applying Eq. 10 and Eq. 9 to Eq. 11, we can get the following unconstrained problem:

$$U(x) = \frac{1}{2}\gamma X^2 + \epsilon \sum_{k=1}^{N} |n_k| + \frac{\eta - 1/2\gamma\tau}{\tau} \sum_{k=1}^{N} n_k^2 + \lambda \cdot \sigma^2 \sum_{k=1}^{N} (\tau x_k^2) \tag{12}$$

Since $U(x)$ above is a quadratic and strict convex (for $\lambda \geq 0$) function of the inventory control x_k. We can therefore get the extreme point for this target

function by setting its all partial derivatives to zero, which means:

$$\frac{\partial U}{\partial x_j} = 2\tau(\lambda\sigma^2 x_j - (\eta - \frac{1}{2}\gamma\tau)\frac{x_{j-1} - 2x_j + x_{j+1}}{\tau^2}) = 0 \tag{13}$$

for $j = 1, 2, ..., N - 1$. We get:

$$\tilde{\kappa}^2 x_j = \frac{x_{j-1} - 2x_j + x_{j+1}}{\tau^2} \tag{14}$$

which $\tilde{\kappa}^2 = \frac{\lambda\sigma^2}{\eta - \frac{1}{2}\gamma\tau}$.Since Eq. 14 is a linear difference equation, we can write its solution in a form $e^{\kappa t_j}$, where $\kappa = \tilde{\kappa} = \sqrt{\frac{\lambda\sigma^2}{\eta}}$ as $\tau \to 0$.

Then the trading trajectory solution with boundary condition $x_0 = X, x_N = 0$ is:

$$x_j = X \cdot \frac{sinh(\kappa(T - t_j))}{sinh(\kappa T)}, \ j = 0, ..., N \tag{15}$$

and the associated trading list is:

$$n_j = X \cdot \frac{2sinh(\frac{1}{2}\kappa\tau)}{sinh(\kappa T)}cosh(\kappa(T - t_{j-1/2})), \ j = 1, 2, ..., N \tag{16}$$

2.2 Nonlinear Impact Model

In the above model, we have made some assumptions that price impact is a linear function of liquidate speed, and the only source of uncertainty is from underlying assets' volatility. However, in practice, linearity in price impact is an unreal assumption[15]. Therefore, the following model[2] improves the original model by assuming that the price impact is a nonlinear function of trading speed.

2.2.1 Nonlinear Price Impact

Assume our initial units of stock is X, and we wish to fully liquidated our inventory at time T, time interval $\tau = \frac{T}{N}$, $t_k = k\tau \ for \ k = 0, 1, ..., N$. Our holdings at time t_k is x_k, and our trading strategy is defined as same as the first model $n_k = x_{k-1} - x_k$, we can also define our liquidate speed $v_k = \frac{n_k}{\tau}$. We can get the same equation as Eq. 1, which is:

$$x_k = X - \sum_{i=1}^{k} n_i = \sum_{j=k+1}^{N} n_i, \ k = 0, ..., N.$$

This time we still divide the price impact into permanent impact and temporary impact two parts.

For permanent impact, the price dynamics satisfies the same form as Eq. 2, and here we can rewrite as:

$$S_k = S_0 + \sigma\sqrt{\tau}\sum_{k=0}^{N} \xi_k - \tau\sum_{k=0}^{N} g(v_k) \tag{17}$$

Figure 2: $f(v)$ and $h(v)$

σ is the *absolute* volatility of the stock, $\xi_k \sim \mathcal{N}(0,1)$, *i.i.d*, $g(v)$ denotes the linear permanent impact function. The formulation is $g(v) = \gamma X$. However, at this time, we consider this permanent impact is relatively small and will have no effect on determining the optimal execution strategy. It is a bit different from the former model, this formulation means we assume permanent impact only depends on inventory. The reason we made such assumption is because we want to get a more explicit solution for final trading model.

For temporary impact, the price dynamics is determined by following:

$$\tilde{S}_k = S_k - h(v_k) + \frac{1}{\sqrt{\tau}} f(v_k)\tilde{\xi}_k, \ k = 1, 2, ..., N \tag{18}$$

$h(v_k)$ is a temporary impact function, $\tilde{\xi}_k \sim \mathcal{N}(0,1)$, *i.i.d*, $f(v_k)$ is a new function which represents the uncertainty of the trading execution. The term $\frac{1}{\sqrt{\tau}}$ have two meanings: if it is fixed and finite, this term is just a scaling of the parameters; if in an continuous-time model, we make $\tau \to 0$, then this term can make sure that we can not diversify the risk by breaking the order into several small orders due to the uncertainty of the trade execution.

As shown in Fig. 2(Also see [2, Fig.1]), $h(v)$ represents the temporary price impact which is increasing as trading speed grows. The dashed line $f(v) \pm h(v)$ represents one standard deviation of the mean.

For a pure selling model, with $v \geq 0$, we take:

$$h(v) = \eta v^k \tag{19}$$

$$f(v) = 0 \tag{20}$$

with $k > 0$.

- $k < 1$, we interpret with fast trading velocity is under-penalized compared to linear trading velocity, which means sell the same amount of stock at every execution.
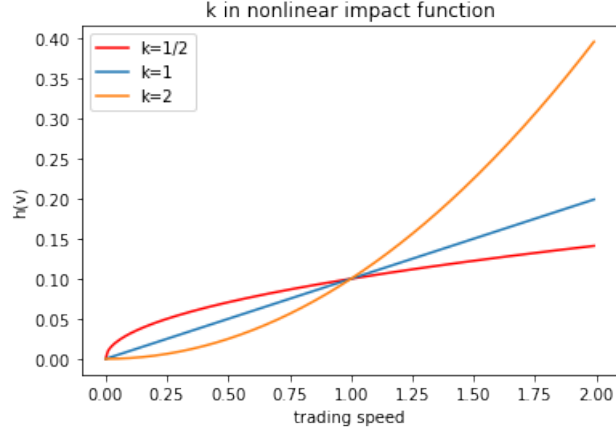
Figure 3: k in nonlinear temporary impact

- $k > 1$, we interpret with fast trading velocity is over-penalized compared to linear trading velocity. See Fig. 3.

The execution cost can be represented by the same as $XS_0 - \sum_{k=1}^{N} n_k \tilde{S}_k$, which equals:

$$- \sigma\sqrt{\tau} \sum_{k=1}^{N-1} x_k \xi_k + \tau \sum_{k=1}^{N} x_k g(v_k) + \tau \sum_{k=1}^{N} v_k h(v_k) - \sqrt{\tau} \sum_{k=1}^{N} v_k f(v_k) \xi_k \qquad (21)$$

2.2.2 Continuous-time Mean-variance Criteria

Same to the former performance criteria, we first show the discrete-time mean-variance criteria:

$$E(x) = \sum_{k=1}^{N} x_k g(v_k)\tau + \sum_{k=1}^{N} v_k h(v_k)\tau \qquad (22)$$

$$V(x) = \sum_{k=1}^{N} \sigma^2 x_k^2 \tau + \sum_{k=1}^{N} v_k^2 f(v_k)^2 \tau \qquad (23)$$

Taking the limit $\tau \to 0$, the trading strategy is a continuous path $x(t)$, and the execution speed $v(t) = -\dot{x}(t)$, because we are a pure selling process. Then the above equations have following continuous-time version:

$$E(x) = \int_0^T (x(t)g(v(t)) + v(t)h(v(t)))dt \qquad (24)$$

$$V(x) = \int_0^T (\sigma^2 x(t)^2 + v(t)^2 f(v(t))^2)dt \qquad (25)$$

and our target function is:

$$\min_x U(x) = E(x) + \lambda V(x) \qquad (26)$$

which is minimizing:

$$\min_{x(t)} \int_0^T F(x(t), v(t)) \, dt \tag{27}$$

where$F(x(t), v(t)) = xg(v) + vh(v) + \lambda\sigma^2 x^2 + \lambda v^2 f(v)^2$.

2.2.3   Solutions

In this calculus of variations, we can use Euler-Lagrange equation to get the stationary point:

$$\frac{d}{dt}\frac{\partial}{\partial \dot{x}}F(x, v) = \frac{\partial}{\partial x}F(x, v)$$

$$-\frac{d}{dt}F_v(x, v) = F_x(x, v)$$

$$0 = F_x(x, -\dot{x}) + \frac{d}{dt}F_v(x, -\dot{x})$$

$$0 = F_x(x, -\dot{x}) + \dot{x}F_{vx}(x, -\dot{x}) - \ddot{x}F_{vv}(x, -\dot{x})$$

To solve this second-order ordinary differential equation, we need to give boundary values $x(0)$ and $x(T)$. Since F does not explicitly depend on t, we can multiply by $\dot{x}$ and integrate for both sides, and get:

$$F(x, -\dot{x}) + \dot{x}F_v(x, -\dot{x}) = constant \tag{28}$$

We may obtain:

$$P(v) - P(v_0) = x(g(v) - vg'(v)) + \lambda\sigma^2 x^2 \tag{29}$$

where:

$$P(v) = v^2 h'(v) + \lambda v^2 (f(v)^2 + 2vf(v)f'(v)) = v^2 \frac{d}{dv}(h(v) + \lambda v f(v)^2)$$

The constant term is obtained when $v_0 = -\dot{x}|_{x=0}$, for a pure selling model, we have $v_0 \gg 0$, also we assume $P(v)$ is an increasing function of v and hence invertible to simplify the problem and get explicit solutions. Since $P(v)$ is increasing as our assumption, so as $P^{-1}(v)$. Thus, as $v_0 \to 0$, our final execution time T increase, then we can get our longest possible execution time by setting $v_0 = 0$, $P(0) = 0$. Now at this stage, we can plug in our price dynamic models for solutions.

The permanent impact is a linear function given:$g(v) = \gamma X$ which means $g(v)$ is independent of $x(t)$, so in Eq. 29, the first term in RHS vanished, and Eq. 29 becomes:

$$\int_{x(t)}^X \frac{dx}{P^{-1}(\lambda\sigma^2 x^2 + P(v_0))} = t \tag{30}$$

Since we setting $v_0 = 0$, we may get the quadrature problem:

$$\int_{x(t)}^{X} \frac{dx}{P^{-1}(\lambda\sigma^2 x^2)} = t \tag{31}$$

The temporary impact function and uncertainty function are $h(v) = \eta v^k$, $f(v) = 0$, we can get $P(v) = k\eta v^{k+1}$. Eq. 31 gets:

$$\int_{x(t)}^{X} (\frac{\lambda\sigma^2 x^2}{k\eta})^{-\frac{1}{k+1}} dx = t \tag{32}$$

Finally, our solution is:

$$x(t) = \begin{cases} X(1 + \dfrac{1-k}{1+k}\dfrac{t}{T_*})^{-\frac{1+k}{1-k}} & 0 < k < 1 \\[2mm] X exp(-\dfrac{t}{T_*}) & k = 1 \\[2mm] X(1 - \dfrac{k-1}{k+1}\dfrac{t}{T_*})^{\frac{k+1}{k-1}} & k > 1 \end{cases} \tag{33}$$

T_* is called 'characteristic time', and is defined as:

$$T_* = (\frac{k\eta X^{k-1}}{\lambda\sigma^2})^{\frac{1}{k+1}} \tag{34}$$

2.3 Incorporating Order-flow into Execution Model

2.3.1 The Cartea-Jaimungal Model

In the Cartea-Jaimungal model[4], it is assumed that the speed of investor's liquidation, by $v = \{v_t\}_{0 \le t \le T}$, and we denote the inventory by $X^v = \{X_t^v\}_{0 \le t \le T}$, which satisfies the equation:

$$dX_t^v = -v_t dt, \tag{35}$$

The impact of investor's speed of trading on prices could be divided into two parts: the permanent part and the temporary part. The permanent part influences the midprice, while the temporary part just impacts the immediate prices after the investor sells the shares.

Besides, we denote the speed sending buy and sell Market orders (MOs) of the rest traders: $\mu^+ = \{\mu_t^+\}_{0 \le t \le T}$, $\mu^- = \{\mu_t^-\}_{0 \le t \le T}$ respectively. And the midprice process $S^v = \{S_t^v\}_{0 \le t \le T}$, satisfies the stochastic differential equation (SDE)

$$dS_t^v = (g(\mu_t^+) - g(\mu_t^- + v_t))dt + \sigma dW_t, S_0^v = S, \tag{36}$$

where $g(\cdot)$ is a function denoting the permanent impact of order-flow, $W = \{W_t\}_{0 \le t \le T}$ is a standard Brownian motion, and μ^+, μ^- are independent of $W = \{W_t\}_{0 \le t \le T}$.

As for the temporary impact, the average price per share obtained will be

lower than the current bid price,

$$\tilde{S}_t^v = S_t^v - \frac{1}{2}\delta - f(v_t), \tag{37}$$

where $\delta \geq 0$ is the bid-ask spread which is assumed to be a constant, and $f(\cdot)$ is a function denoting the temporary impact of the investor's trading action.

Finally, we get the investor's cash process, C_t^v satisfies the SDE

$$dC_t^v = \tilde{S}_t^v v_t dt, C_0^v = x. \tag{38}$$

2.3.2 Performance Criterion and Value Function

The performance criteria is

$$H^v(t, c, S, \boldsymbol{\mu}, x) = \mathbb{E}_{t,c,S,\boldsymbol{\mu},x}\left[C_T + X_T^v(S_T^v - \frac{1}{2}\delta - \alpha X_T^v) - \phi \int_t^T (X_u^v)^2 du\right], \tag{39}$$

where $\boldsymbol{\mu} = \{\mu^+, \mu^-\}$, $C_t = c$, $S_{t-} = S$, and $X_t = x$ in the expectation operator. And the value function is

$$H(t, c, S, \boldsymbol{\mu}, x) = sup_{v \in \mathcal{A}} H^v(t, c, S, \boldsymbol{\mu}, x), \tag{40}$$

where $\mathcal{A}$, the set of admissible strategies, consists of $\mathcal{F}$-predictable processes where $\int_t^T |v_u| du < +\infty, \mathbb{P}-$a.s.

In the expectation operator, C_T means the investor's terminal cash from liquidating the shares throughout the trading horizon, $X_T^v(S_T^v - \frac{1}{2}\delta - \alpha X_T^v)$ means that at the end of the strategy, the proceeds the investor would receive while liquidating the remaining inventory X_T^v, and $\phi \int_t^T (X_u^v)^2 du$ is the running penalty , with the running penalty parameter $\phi \geq 0$.

Now, we simplify the problem by assuming that $g(x) = bx, f(x) = kx$, with non-negative constants b, k. To simplify the problem, we define the net order-flow $\mu_t = \mu_t^+ - \mu_t^-$, excluding the investor's trading rate, and we obtain the new midprice process

$$dS_t^v = b(\mu_t - v_t)dt + \sigma dW_t. \tag{41}$$

And the $H(t, c, S, \boldsymbol{\mu}, x)$ satisfies the dynamic programming equation (DPE):

$$(\partial_t + \frac{1}{2}\sigma^2 \partial_{SS})H + \mathcal{L}^{\boldsymbol{\mu}}H - \phi x^2 + sup_v\left\{\left(v(S - \frac{1}{2}\delta - kv)\partial_c + b(\mu - v)\partial_S - v\partial_x\right)H\right\} = 0, \tag{42}$$

for $x > 0$, subject to the following condition

$$H(T, c, S, \boldsymbol{\mu}, x) = c + x(S - \frac{1}{2}\delta) - \alpha x^2, \tag{43}$$

and $\mathcal{L}^{\boldsymbol{\mu}}$ is the generator of the process $\boldsymbol{\mu}$.

2.3.3 The DPE Solution

And the solution to the DPE(42)

$$H^v(t, c, S, \boldsymbol{\mu}, x) = c + x(S - \frac{1}{2}\delta) + h_0(t, \boldsymbol{\mu}) + xh_1(t, \boldsymbol{\mu}) + x^2 h_2(t), \tag{44}$$

$$h_2(t) = \sqrt{k\phi} \frac{1 + \zeta e^{2\gamma(T-t)}}{1 - \zeta e^{2\gamma(T-t)}} - \frac{1}{2}b, \tag{45a}$$

$$h_1(t, \boldsymbol{\mu}) = b \int_t^T \left(\frac{\zeta e^{\gamma(T-u)} - e^{-\gamma(T-u)}}{\zeta e^{\gamma(T-t)} - e^{-\gamma(T-t)}} \right) \mathbb{E}_{t,\boldsymbol{\mu}}[\mu_u] du, \tag{45b}$$

$$h_0(t, \boldsymbol{\mu}) = \frac{1}{4k} \int_t^T \mathbb{E}_{t,\boldsymbol{\mu}}[h_1^2(t, \boldsymbol{\mu}_u)] du, \tag{45c}$$

with the constants

$$\gamma = \sqrt{\frac{\phi}{k}}, \quad and \quad \zeta = \frac{\alpha - \frac{1}{2}b + \sqrt{\phi k}}{\alpha - \frac{1}{2}b - \sqrt{\phi k}}. \tag{45d}$$

The term $c + x(S - \frac{1}{2}\delta)$ is the book-value of the investor's remaining inventory. The function $h_0(t, \boldsymbol{\mu}) + xh_1(t, \boldsymbol{\mu}) + x^2 h_2(t)$ is independent of the midprice and represents the excess value of the optimal trading.

Suppose $\mathcal{F}_t^{\boldsymbol{\mu}}$ denotes the natural filtration generated by the $\boldsymbol{\mu}$. And the optimal trading speed is $v_t^* = -\frac{1}{2k}(bx + \partial_x h)$, with $h(t, \boldsymbol{\mu}, x) = h_0(t, \boldsymbol{\mu}) + xh_1(t, \boldsymbol{\mu}) + x^2 h_2(t)$,

$$v_t^* = \gamma \frac{1 + \zeta e^{2\gamma(T-t)}}{1 - \zeta e^{2\gamma(T-t)}} Q_t^{v^*} - \frac{b}{2k} \int_t^T \left(\frac{\zeta e^{\gamma(T-u)} - e^{-\gamma(T-u)}}{\zeta e^{\gamma(T-t)} - e^{-\gamma(T-t)}} \right) \mathbb{E}[\mu_u | \mathcal{F}_t^{\boldsymbol{\mu}}] du. \tag{46}$$

2.3.4 Comparison with the Almgren–Chriss Model

Compared with the Cartea-Jaimungal model, we think the Almgren–Chriss strategy in continuous-time is the optimal strategy, overlooking the impact of order-flow on the midprice dynamics. However, the agent's own trading rate does affect the midprice in a permanent way, denoted by

$$dS_t^v = -bv_t dt + \sigma dW_t, \tag{47}$$

and the execution price receives a linear impact, so that $\tilde{S}_t^v = S_t^v - kv_t$. In this case the optimal liquidation speed is

$$v_t^{AC} = \gamma \frac{1 + \zeta e^{2\gamma(T-t)}}{1 - \zeta e^{2\gamma(T-t)}} X_t^{v^*}, \tag{48}$$

with the same constants γ and ζ in (45d). which is the first term of the optimal speed of trading in the Cartea-Jaimungal model(46), while the second term of the DPE solution adjusts the speed according to the weighted average of the expected net order-flow in the future. The strategy in the Cartea-Jaimungal model(46)

places more emphasis on the expected net order-flow if it is near the current time t.The impact of order-flow would become less significant if maturity approaches $(t \to T)$, and the investor just needs to use the Almgren–Chriss strategy.

2.4   Stochastic Volatility and Liquidity

We then model the problem of liquidating some fixed shares of a certain financial asset in a discrete-time way [6] optimally with stochastic volatility and liquidity. Our goal is to find strategies that minimize the expectation and an expected exponential of the implementation cost. We just take the stochastic volatility and the temporary price impact into account. To make the model practical, we would make the Markov assumptions in the following part.

2.4.1   Discrete-time Model

We consider the problem of liquidating an asset position of $X \in \mathbb{R}_+$ shares until a given time $T \in \mathbb{R}_+$. We divide the interval $[0, T]$ into N subintervals of length Δt and decide at every time $t_{n-1} = (n-1)\Delta t$ how many shares y_n to sell in the interval $(t_{n-1}, t_n]$. We assume the execution price

$$\tilde{S}_n = S_{n-1} - \eta_n y_n, \tag{49}$$

where S_n follows the dynamics

$$S_n = S_{n-1} + \sigma_n \xi_n - \gamma y_n, \tag{50}$$

We think of S_n as a fixed convex combination of the bid and ask price: $S_n = \lambda S_n^b + (1 - \lambda)S_n^a$. For $\lambda = 1/2$, S_n is the mid-price. But depending on the particular asset to be traded, it could be closer to the bid or ask. ξ_n is a sequence of random variables independent normally distributed with mean 0 and variance Δt. σ_n is a stochastic volatility. $\gamma \in \mathbb{R}_+$ is a constant describing a permanent price impact and $\{\eta_n\}$ a stochastic liquidity process modelling temporary price impact.

We suppose that S_n, σ_n and η_n are attainable and denote the filtration they generate by $\{\mathcal{F}_n\}$. The proceeds from selling the asset shares are $\sum_{n=1}^{N} y_n \tilde{S}_n$. We describe a liquidation strategy in terms of remaining shares $x_n = X - \sum_{i=1}^{n} y_i$ and call it admissible if the following conditions are satisfied:

$$x_0 = X, x_n \leq x_{n-1}, X_N = 0, \tag{51}$$

Thus, $\{x_n\}$ is predictable with respect to $\{\mathcal{F}_n\}$. An admissible strategy is completely specified by $\{x_n\}$ for $1 \leq n \leq N-1$. We denote the set of all admissible strategies by $\mathcal{A}$. And we define implementation cost as the difference value between the initial value and the proceeds: $C(x) := XS_0 - \sum_{n=1}^{N} y_n \tilde{S}_n$. Since $x_N = 0$, it can be written as

$$C(x) = \frac{\gamma X^2}{2} + \sum_{n=1}^{N} [(x_{n-1} - x_n)^2 (\eta_n - \frac{\gamma}{2}) - \sigma_n \xi_n x_n].$$

Now, to minimize the expectation criterion of $C(x)$. We can just ignore the constant $\frac{cX^2}{2}$ from $C(x)$ and get the same optimal solution for the quadratic form

$$Q(x) = \sum_{n=1}^{N}[(x_{n-1} - x_n)^2 \tilde{\eta}_n - \sigma_n \xi_n x_n], \tilde{\eta}_n = \eta_n - \frac{\gamma}{2} > 0. \tag{52}$$

To minimize the expectation of $Q(x)$, the dynamics of $(S_n, \sigma_n, \tilde{\eta}_n)$ can be more general. Also, we assume that ξ_n is independent of $\sigma(\mathcal{F}_{n-1}, \sigma_n, \tilde{\eta}_n)$, which is equal to $\sigma(\mathcal{F}_{n-1}, \sigma_n, \eta_n)$, with $(\sigma_n, \tilde{\eta}_n)$ taking values in a finite subset $V \subseteq \mathbb{R}_+^2$. Besides, conditioned on $\mathcal{F}_{n-1}$, the distribution of $(\sigma_n, \tilde{\eta}_n)$ just depends on $(\sigma_{n-1}, \tilde{\eta}_{n-1}, \varphi_{n-1})$, while φ_{n-1} is defined as $\varphi_{n-1} := \varphi_{n-1}(\sigma_{n-1}, \xi_{n-1})$ for a positive integer K and a measurable function $\varphi : \mathbb{R} \to \{1, \ldots, K\}$. It follows that $(\sigma_n, \tilde{\eta}_n, \varphi_n)$ is a Markov chain with finite state space $V^K := V \times \{1, \ldots, K\}$ and time-dependent transition probabilities are given by

$$p_{n-1}^{st} = \mathbb{P}\left[(\sigma_n, \tilde{\eta}_n, \varphi_n) = t \mid (\sigma_{n-1}, \tilde{\eta}_{n-1}, \varphi_{n-1}) = s\right], s, t \in V^K.$$

Moreover, if we assume that $(\sigma_n, \tilde{\eta}_n)$ is a Markov chain with state space V which is independent of the sequence of innovations $\{\xi_n\}$. Its transition probabilities are given by

$$p_{n-1}^{st} = \mathbb{P}\left[(\sigma_n, \tilde{\eta}_n) = t \mid (\sigma_{n-1}, \tilde{\eta}_{n-1}) = s\right], s, t \in V.$$

2.4.2 Risk-neutral Objective

$$p_{n-1}^{st} = \mathbb{P}\left[(\sigma_n, \tilde{\eta}_n, \varphi_n) = t \mid (\sigma_{n-1}, \tilde{\eta}_{n-1}, \varphi_{n-1}) = s\right], s, t \in V^K.$$

The model is

$$\min_{x \in \mathcal{A}} \mathbb{E}_0^s[Q(x)] = X^2 a_0^s, \tag{53}$$

and the unique optimal liquidation strategy conditioned on x_{n-1}^* and $(\sigma_{n-1}, \tilde{\eta}_{n-1}, \varphi_{n-1}) = s$ is given by

$$x_n^* = x_{n-1}^* \frac{\mathbb{E}_{n-1}^s[\tilde{\eta}_n]}{\mathbb{E}_{n-1}^s[\tilde{\eta}_n] + \sum_{t \in V^K} p_{n-1}^{st} a_n^t}, n = 1, \ldots, N-1, \tag{54}$$

where the coefficients a_n^s satisfy the backwards recursion:

$$a_{N-1}^s = \mathbb{E}_{N-1}^s[\tilde{\eta}_N], a_{n-1}^s = \frac{\mathbb{E}_{n-1}^s[\tilde{\eta}_n] \sum_{t \in V^K} p_{n-1}^{st} a_n^t}{\mathbb{E}_{n-1}^s[\tilde{\eta}_n] + \sum_{t \in V^K} p_{n-1}^{st} a_n^t}, n \leq N-1, \tag{55}$$

if $\tilde{\eta}_n = \eta$, then

$$a_n^s = \frac{\eta}{N-n}, x_n^* = x_{n-1}^* \frac{N-n}{N-n+1} = X \frac{N-n}{N}.$$

2.4.3 Expected Exponential Cost

It is assumed that $(\sigma_n, \tilde{\eta}_n)$ is a Markov chain with state space V which is independent of the sequence of innovations $\{\xi_n\}$ and transition probabilities are given

by

$$p^{st}_{n-1} = \mathbb{P}\left[(\sigma_n, \tilde{\eta}_n) = t | (\sigma_{n-1}, \tilde{\eta}_{n-1}) = s\right], s, t \in V.$$

The model for Expected Exponential Cost is $\mathbb{E}^s_0\left[exp(\alpha Q(x))\right]$, where $Q(x) = \sum_{n=1}^N [(x_{n-1} - x_n)^2 \tilde{\eta}_n - \sigma_n \xi_n x_n]$. And we consider the value function

$$J^s_n(z) := \min_{x \in \mathcal{A}_n(z)} \mathbb{E}^s_n\left[exp(\alpha Q_n(x))\right], \tag{56}$$

where $\mathbb{E}^s_n$ denotes the conditional expectation $\mathbb{E}\left[.|(\sigma_{n-1}, \tilde{\eta}_{n-1}) = s\right]$, $Q_n(x) = \sum_{i=n+1}^N [(x_{i-1} - x_i)^2 \tilde{\eta}_i - \sigma_i \xi_i x_i]$. And the value function J satisfies the Bellman equation

$$J^s_{N-1}(x_{N-1}) = \sum_{t \in V} p^{st}_{N-1} exp(\alpha x^2_{N-1}(w_2 - \frac{\gamma}{2})),$$

$$J^s_{n-1}(x_{n-1}) = \min_{0 \le x_n \le x_{n-1}} \left[\sum_{t \in V} p^{st}_{n-1} exp(\alpha(x_{n-1} - x_n)^2(w_2 - \frac{\gamma}{2}) + \frac{1}{2}x^2_n \alpha^2 w^2_1 \Delta t) J^n_t(x_n)\right],$$

for $n \le N - 1$ and the minimizing x^*_n n form the unique optimal strategy for the problem. Moreover, if we assume that $\sigma_n = \sigma, \tilde{\eta}_n = \tilde{\eta}$. The minimization of the expected exponential reduces to the deterministic problem

$$\min_{0 \le x_n \le x_{n-1}} \left[\sum_{n=1}^N (x_{n-1} - x_n)^2 \tilde{\eta} + \frac{1}{2}x^2_n \alpha \sigma^2 \Delta t\right],$$

which is a convex problem. And it can be reduced to the first-order condition

$$x_n \alpha \sigma^2 \Delta t = 2\tilde{\eta}(x_{n-1} - 2x_n + x_{n+1}), n = 1, \ldots, N - 1.$$

$x^*_n = X \frac{\sinh(\beta(T - n\Delta t))}{\sinh(\beta T)}$, for the unique $\beta \ge 0$ satisfying $\cosh(\beta \Delta t) - 1 = \frac{\alpha \sigma^2 \Delta t}{4\tilde{\eta}}$.

3   Order Book Based Model

The following models are based on the direction of modelling order book. We briefly introduce the main ideas of three models.

3.1   Discrete-time Markov Chain

3.1.1   Basis Setup

Rama Cont [8] there is a market where limit orders can be represented as a price sequence $1, 2, \ldots, n$. The upper bound n can be large enough in order to put all possible price into our consideration. We can consider the state of the order book $X(t)$ as a continuous- time process $X(t) = (X_1(t), \ldots, X_n(t))$ $t \ge 0.|X_p(t)|, 1 \le p \le n$ is called the number of best limit orders (lowest price for sell and highest price for buy) at price p.

Define the bid price $p_B(t)$ as

$$p_B(t) \equiv \sup \{p = 1, \ldots, n, X_p(t) < 0, \} \vee 0$$

and ask price $p_A(t)$ at time t as

$$p_A(t) = \inf \{p = 1, \ldots, n, X_p(t) > 0\} \wedge (n+1)$$

which we define price 0 and n+1 for bid/ask price to make sure that there always have bid/ask orders in the book. Moreover, the mid-price $p_M(t)$ and the bid-ask spread $s(t)$ are:

$$p_M(t) \equiv \frac{p_B(t) + p_A(t)}{2} \quad \text{and} \quad s(t) \equiv p_A(t) - p_B(t)$$

3.1.2 Dynamics of the Order Book

Then we can define dynamics of the order book. We denote the order book by $x \in \mathbb{Z}^n x^{p \pm 1} = x \pm e^p$, where $e^p \in \mathbb{Z}^n$ is a zero vector with only $p - th$ value is 1. Assuming all orders are of unit size(take average size in empirical examples), the state transitioning can be formulated by:

- a limit buy order at level $p \le p_A(t) : x \to x_{p-1}$

- a limit sell order at level $p \ge p_B(t) : x \to x_{p+1}$

- a market buy order: $x \to x_{p_A(t)-1}$

- a market sell order : $x \to x_{p_B(t)+1}$

- a cancellation of an oustanding limit buy order at price $p \le p_A(t) : x \to x_{p+1}$

- a cancellation of an oustanding limit sell order at price $p \ge p_B(t) : x \to x_{p-1}$

Moreover, they propose a stochastic model using independent Poisson processes. Given the above settings, X is a continuous-time Markov chain with known transition rates.

3.1.3 Model Estimation

Zovko and Farmer [16] suggested a power law for the arrival rates λ

$$\lambda(i) = \frac{k}{i^\alpha}$$

The arrival rate of market orders can be estimated by

$$\hat{\mu} = \frac{N_m}{T} \frac{S_m}{S_l}$$

where T is the length of our sample (in minutes) and N_m is the number of market orders.

The cancellation rate function is then given by

$$\hat{\theta}(i) = \frac{N_c(i)}{TQ_i} \frac{S_c}{S_l}$$

3.2 Fluid-diffusion Model

3.2.1 LOB Dynamics

Ulrich Horst and Dörte Kreher[10] specified a one-sided LOB model. It is a discrete-time model and the process is $\tilde{S}^{(n)} = (B^{(n)}, v^{(n)})$. $B^{(n)}$ is the dynamics of the best bid price and $v^{(n)}$ is the dynamics of the bid-side volume process. The grids they set are $\Delta x^{(n)}, \Delta v^{(n)}$ and $\Delta t^{(n)}$. The initial value of best bid price is $B_0^{(n)} = b_n \Delta x^{(n)}$. The initial value of the volume density function is $v_0^{(n)}$. Then the dynamics of the model is as following:

$$B_k^{(n)} = B_{k-1}^{(n)} + \Delta x^{(n)} \left[\mathbb{I}_B \left(\phi_k^{(n)} \right) - \mathbb{I}_A \left(\phi_k^{(n)} \right) \right]$$

$$v_k^{(n)} = v_{k-1}^{(n)} + \Delta v^{(n)} M_k^{(n)}$$

here

$$M_k^{(n)}(\cdot) := \mathbb{I}_C \left(\phi_k^{(n)} \right) \frac{\omega_k^{(n)}}{\Delta x^{(n)}} \mathbb{I}_{I^{(n)}} \left(\pi_k^{(n)} \right) (\cdot)$$

and $\phi_k^{(n)}$ is an indicator function which indicates three events - price increases, price decreases and limit order placements(cancellations). $\omega_k^{(n)}$, $\pi_k^{(n)}$ are the size and location of a replacement or cancellation respectively.

3.2.2 Price Process

In this part, they discretize price $\Delta p^{(n)} = o(1)$ and define the normalized increments of $B^{(n)}$ by

$$\delta Z_k^{(n)} := \frac{\delta \bar{B}_k^{(n)}}{r^{(n)} \left(S_{k-1}^{(n)} \right)}, \quad Z_k^{(n)} := \sum_{j=1}^{k} \delta Z_j^{(n)} \quad \text{for } k = 1, \ldots, T_n.$$

Then,

$$B^{(n)}(t) = B_0^{(n)} + \sum_{k=1}^{\lfloor t/\Delta t^{(n)} \rfloor} \delta B_k^{(n)}$$

$$= B_0^{(n)} + \sum_{k=1}^{\lfloor t/\Delta t^{(n)} \rfloor} \left[p^{(n)} \left(S_{k-1}^{(n)} \right) \Delta t^{(n)} + r^{(n)} \left(S_{k-1}^{(n)} \right) \delta Z_k^{(n)} \right]$$

Finally Ulrich Horst[10] obtained the continuous process $Z_k^{(n)}, k = 1, \ldots, T_n$ by linear interpolation which is weakly to convergent to a standard Brownian motion

when n goes to infinity.

$$Z^{(n)}(t) := \sum_{k=0}^{T_n} Z_k^{(n)} \mathbb{I}_{\left[t_k^{(n)}, t_{k+1}^{(n)}\right)}^{(t),} \quad t \in [0, T]$$

.

3.2.3 Volume Process

Similar to the analysis of price process, the infinite dimensional volume process $V(n)$ converge as $n \to \infty$. Therefore, we can represent it as a solution to a stochastic differential equations driven by infinite dimensional martingale which is convergent in distribution to a cylindrical Brownian motion as $n \to \infty$.

3.2.4 Infinite Dimensional SDE

The sequence of LOB models is relatively compact in a localized sense and any accumulation point is the solution to a certain infinite dimensional SDE which is designed by a Brownian motion and a cylindrical Brownian motion. Moreover, under certain assumptions, the LOB dynamics will converge to a unique limit.

Since the finite dimensional SDE

$$\bar{B}^m(t) = B_0 + \int_0^t p\left(\bar{S}^m(u)\right) du + \int_0^t r\left(\bar{S}^m(u)\right) dZ_u$$

$$\bar{V}_i^m(t) = \langle V_0^m, f_i \rangle + \int_0^t \mu_i\left(\bar{S}^m(u)\right) du + \sum_{j \leq i} \int_0^t d_{ij}\left(\bar{S}^m(u)\right) dW_u^j,$$

$$i \in \mathcal{I}_m(l_0),$$

with $\bar{V}^m := \sum_{i \in \mathcal{I}_m(l_0)} \bar{V}_i^m f_i$ and $\bar{S}^m = \left(\bar{B}^m, \bar{V}^m\right)$ has a unique strong solution $\bar{S}^m$.

Considering the above results, define

$$\underline{V}^m(t) := V_0^m + \sum f_i \int_0^t \mu_i\left(\bar{S}^m(u)\right) du + \sum f_i \sum \int_0^t d_{ij}\left(\bar{S}^m(u)\right) dW_u^j$$

Then $S^m := \left(\bar{B}^m\left(\cdot \wedge \tau_m\right), \underline{V}^m\left(\cdot \wedge \tau_m\right)\right)$ is the unique strong solution of

$$\hat{S}^m(t) = S_0^m + \int_0^{t \wedge \tau_{m,m}} G_m\left(\hat{S}^m(u)\right) dY(u), \quad t \in [0, T],$$

$$\tau_{m,l} := \inf\left\{t \geq 0 : \hat{B}^m(t) \geq l\right\} \wedge T.$$

3.3 LSTM Model

3.3.1 Recurrent Neural Network

We firstly introduce RNN model which a basic deep learning method. RNN is very useful in analyzing times series data such as stock prices. The difference between vanilla Neural Network and Recurrent Neural Network is that RNN have loops and it allows information sharing and preserving. The output of hidden layer are stored in the memory which can also be used as input again.

However, if we have a long-term time series data, there is a problem that the gradients tend to vanish or explode after propagating many times. Therefore, Sepp Hochreiter and Jurgen Schmidhuber[9] proposed LSTM in 1997.

3.3.2 Long Short-term Memory Model

Now, we considered the LSTM model. In LSTM, there is a special structure called gate unit which is used to decide whether keep or override information in a memory cell. There are three kinds of gates - input gate, output gate and forget gate. These gates usually consist of two parts - a sigmoid neural net layer and a pointwise multiplication operation. The sigmoid output which is between 0 and 1 gives an idea of how component can go through .

First, we need to determine what to erase from the memory cell. This step is dependent on the "forget gate" sigmoid layer. It first takes the value of h_{t-1} and x_t and gives out a number between 0 and 1 for each component in memory cell C_{t-1}. If the output is 1, we keep all the information while if the output is 0, we will throw away all of the information.

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f)$$

Next, we determine what new information to store in the memory cell. This consists of two parts - an "input gate" sigmoid layer and a tanh layer. The "input gate" layer tell us what to update and the tanh layer will create a new $\tilde{C}_t$ which might be include in the memory cell.

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i)$$
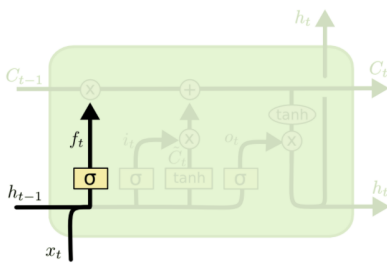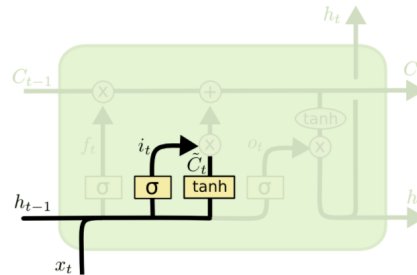$$\tilde{C}_t = \tanh(W_C [h_{t-1}, x_t] + b_C)$$

Figure 4: Step 1

Figure 5: Step 2

Then, we can update the old memory cell C_{t-1} by a new one C_t by combining the values we obtained before. We multiply the old cell with f_t and plus $i_t * \tilde{C}_t$ to get a C_t.

$$C_t = f_t * C_{t-1} + i_t * \tilde{C}_t$$

In the end, we produce the output with the "ouput gate" sigmoid layer because we will filter the cell state to give the final output. We first use a sigmoid layer to determine what part to reveal. Then the value will go through the tanh. Then we multiply it by the output of the "ouput gate" which gives us the final h_t.

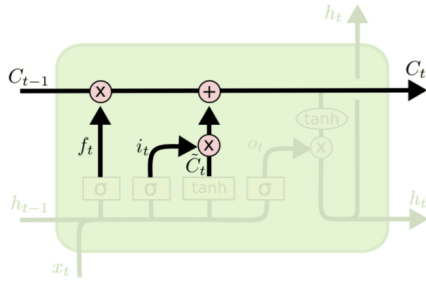$$o_t = \sigma(W_o[h_{t-1}, x_t] + b_o)$$
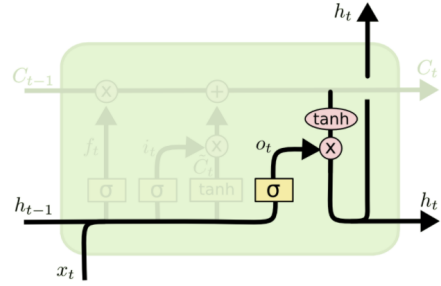$$h_t = o_t * \tanh C_t$$

Figure 6: Step 1

Figure 7: Step 2

4 Simulation

We use the linear impact model and nonlinear impact model for simulation and try to find the optimal execution inventory list under the given real stock data. However, we have to admit that some of the estimations may not be fully correct. The data we choose is NASDAQ 2014 INTC stock data, because we want to be consistent with the third model, Catea-Jaimungal model[5].

We estimate the parameters in the following way. We get 2014 daily stock data by *yfinance* yahoo-finance python package in Table. 1:

Then we calculate the realized volatility for parameter estimation, and plot the log return of the stock and the monthly realized volatility in Fig. 8.

For the linear impact model, we assume that bid-ask spread is $\frac{1}{8}$, our expected annual return is 10%, initial price $S_0 = 53.41$ and initial inventory $X = 4042$. The calculated annual volatility is 0.23 and we can get daily volatility by $\frac{0.23}{\sqrt{240}}$. To obtain the absolute σ and α, we must scale the price, so $\sigma = \frac{0.23}{\sqrt{240}} * S_0$ and $\alpha = \frac{0.1 * S_0}{240}$.

Suppose we want to liquidate all our inventory in an hour, so we take $T = N = 60(mins)$, $\tau = 1(min)$. For temporary impact function, we choose $\epsilon = \frac{1}{16}$,

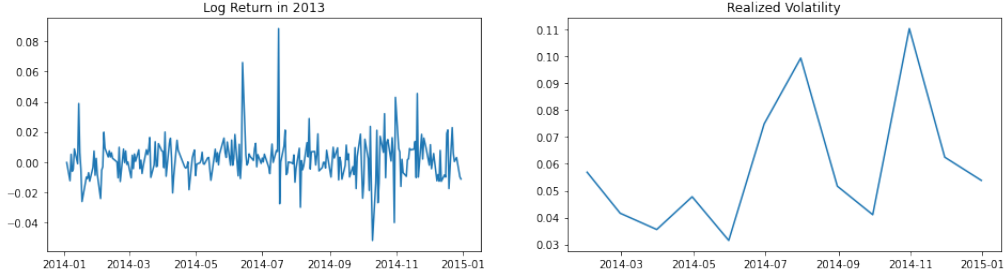|	Open	High	Low	Close	Adj Close	Volume
Date						
2014-01-02	25.780001	25.820000	25.469999	25.790001	20.523907	31833300
2014-01-03	25.860001	25.900000	25.600000	25.780001	20.515947	27796700
2014-01-04	25.770000	25.790001	25.450001	25.459999	20.261290	28682300
...	...	...	...	...	...	...
2014-12-26	37.520000	37.740002	37.520000	37.549999	30.830029	14037200
2014-12-29	37.450001	37.520000	37.169998	37.180000	30.526237	12203300
2014-12-30	37.080002	37.189999	36.759998	36.759998	30.181400	15214100

Table 1: 2014 INTC data

Figure 8: realized volatility

which is one-half of the bid-ask spread and $\eta = 2.5 \cdot 10^{-4}$. For permanent impact function, we have $\gamma = 2.5 \cdot 10^{-7}$.

We plot the efficient frontier for searching the best estimation of risk-aversion parameter λ.

Each point in Fig. 9 represents a trading strategy under a specific λ. The tangent line can indicate the optimal solution for λ and we estimate $\lambda = 5 \cdot 10^{-6}$. Then we plot the optimal solution in Fig. 10:

The red dots represent the inventory at each time. We can see the inventory decreases fast at first up to five minutes, and gradually slow until 30 minutes approximately. The linear model has a significant long tail after 30 minutes,
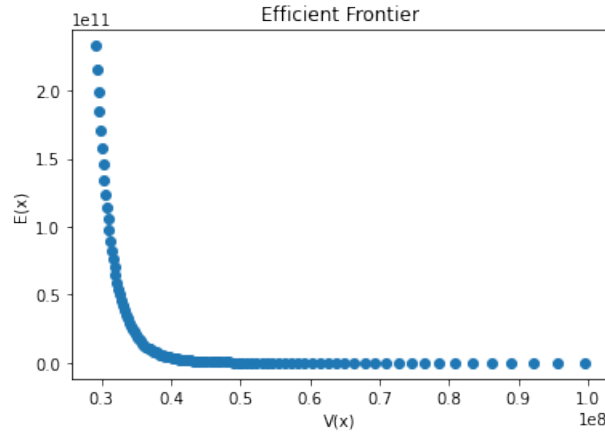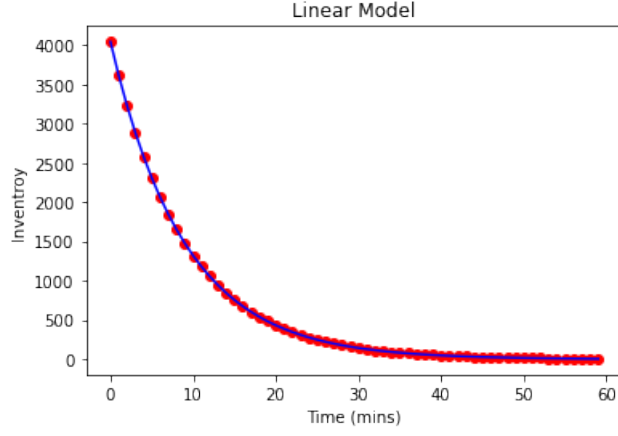

Figure 9: Efficient Frontier
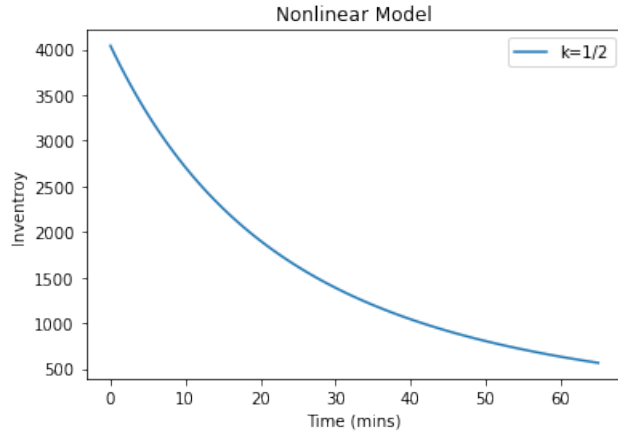
Figure 10: Linear Impact Model

Figure 11: Nonlinear Impact Model

which shows that the decreasing speed of the inventory is very slow.

For the nonlinear imapct model, our initial inventory is $X = 4042$, and $\sigma = \frac{0.23}{\sqrt{240}}$. As the same in previous strategy, we want to liquidate all our inventory in one hour, so we choose $k = \frac{1}{2}$, the square root of trading speed. For parameter λ, it is quite difficult to choose for real situations, so we decide to let $\lambda = 1$ for simplification. For temporary impact function, $\eta = 0.1$, which gives a heavy influence for temporary impact. Then from the optimal solution, we plot the optimal inventory in Fig. 11.

We can see the liquidate speed increases at first and gradually slows down similarly. However, compared to Fig. 10, the difference is, for the nonlinear impact model, it does not have a long tail at the last trading time.

5   Conclusion

In this paper, we introduce the optimal execution problem, and classify the methods into two main streams with several models: price impact based models and order book based models. Price impact models focus on modelling temporary

and permanent impact from the dynamics of market price and microstructure of
the market. Order book based models capture the dynamics of the limit order
book system.

Linear impact model is a simple and straightforward method and gives ex-
plicit solution for the problem under the assumption that both temporary and
permanent impact functions are linear functions. Both of the functions use mean-
variance criterion for optimization. Alternatively, nonlinear impact model pro-
vides a power law temporary impact function, which gives more weight on the
trading speed. Besides, it defines a 'characteristic time', which depends on the
initial portfolio size. As the consequence, from the simulation, we can see the
nonlinear model is more suitable for rapid execution because it does not have
long tails compared with linear model.

The Cartea-Jaimungal strategy contains an Almgren–Chriss execution strat-
egy adjusted by a weighted-average of the expected net order-flow and propor-
tional to the ratio of permanent to temporary linear impacts. Stochastic volatility
is a discrete-time model following a coupled and finite Markov chain. We also de-
rive Bellman equations and forward-backward equations to solve the risk-neutral
problem and the expected exponential problem.

For three order book based models, the parameters for Markov chain model
are easy to estimate. When applying diffusion model, we usually need use a soft-
ware such as Matlab to solve the stochastic differential equations. And LSTM is
suitable for handling and predicting important events with relatively long inter-
stitions and delays in time series.

6   Appendix

```python
1  # import packages
2  import numpy as np
3  import yfinance as yf
4  import pandas as pd
5  import matplotlib.pyplot as plt
6
7  def realized_vol(x):
8      return(np.sqrt(np.sum(x**2)))
9
10 data = yf.download('INTC',start='2014-01-01',end='
       2014-12-31')
11 data['Log_return'] = np.log(data['Adj_Close'] / data['Adj_
       Close'].shift(1))
12
13 data = data.loc[:,['Log_return']]
14 data = yf.download('INTC',start='2014-01-01',end='
       2014-12-31')
15 data_rv = data.groupby(pd.Grouper(freq = 'M')).apply(
       realized_vol)
```

```python
16  data_rv.rename(columns={'Log_return': 'realized_vol'},
        inplace=True)
17
18  plt.figure(figsize=(16,4))
19  plt.subplot(1,2,1)
20  plt.plot(data)
21  plt.title('Log_Return_in_2013')
22  plt.subplot(1,2,2)
23  plt.plot(data_rv)
24  plt.title('Realized_Volatility')
25
26  # Nonlinear model simulation for INTEL stock 2014
27  k = 1 / 2
28  sigma = 0.23 / np.sqrt(240) # change to daily volatility
29  X = 4042   # our initial inventroy
30  lmda = 1
31  eta = 1e-1
32  tau = 1 # an hour
33  T_star = ((k * eta * (X**(k-1))) / (lmda * sigma ** 2)) **
        (1 / (k+1))
34
35  def nonlinear_x1(t,k): # 0<k<1
36      T_star = ((k * eta * (X**(k-1))) / (lmda * sigma ** 2)
            ) ** (1 / (k+1))
37      return((1 + (1-k) / (1+k) * t / T_star)**(-(1+k) / (1-
            k)))
38
39  def nonlinear_x2(t,k): # k=1
40      T_star = ((k * eta * (X**(k-1))) / (lmda * sigma ** 2)
            ) ** (1 / (k+1))
41      return(np.exp(-t / T_star))
42
43  def nonlinear_x3(t,k): # k>1
44      T_star = ((k * eta * (X**(k-1))) / (lmda * sigma ** 2)
            ) ** (1 / (k+1))
45      return((1 - (k-1) / (k+1) * t / T_star)**((k+1) / (k
            -1)))
46
47  t1 = np.linspace(0,60,100) * 0.10838061547025504 # T_star
        at k = 1/2
48
49  plt.plot(t1,X*nonlinear_x1(t1,1/2),label = 'k=1/2')
50  plt.xticks(np.arange(0,7,1),np.arange(0,70,10))
51  plt.title('Nonlinear_Model')
52  plt.xlabel('Time_(mins)')
53  plt.ylabel('Inventroy')
54  plt.legend()
```

```
55
56  def h(v,k,eta): # nonlinear impact function
57      return(eta*v**k)
58  v = np.arange(0,2,0.01)
59
60  plt.plot(v,h(v,1/2,1e-1),label='k=1/2',color='red')
61  plt.plot(v,h(v,1,1e-1),label='k=1')
62  plt.plot(v,h(v,2,1e-1),label='k=2')
63  plt.ylabel('h(v)')
64  plt.xlabel('trading speed')
65  plt.title('k in nonlinear impact function')
66  plt.legend()
67
68  # linear model for INTEL stock in 2014
69  # bid-ask spread 1/8
70  # expected return 10%
71
72  S_0 = 53.41
73  X = 4042
74  T = 60
75  N = 60
76  tau = 1
77  sigma = 0.23*S_0 / np.sqrt(240) # absolute volatility =
        vol* initial price
78  alpha = 0.1*S_0/240   # alpha = expected return* initial
        price/date
79  epsilon = 0.0625 # constant for temporary impact
80  gamma = 2.5e-7
81  eta = 2.5e-4
82  lambda_u = 2e-7
83  lamda = lambda_u
84  lambda_v = 1.645
85
86  def x(t,lambda_u): # solution for 0<k<1
87      kappa = np.sqrt(lambda_u* sigma ** 2 / eta)
88      return(X * np.sinh(kappa * (T - t)) / np.sinh(kappa *
        T))
89
90  def E(x,lambda_u): # expectation of cost
91      kappa = np.sqrt(lambda_u* sigma ** 2 / eta)
92      return(gamma * x**2 / 2 + epsilon*x + (eta-gamma*tau
        /2)*x**2*(np.tanh(kappa*tau/2)*(tau*np.sinh(2*kappa
        *T)+2*T*np.sinh(kappa*tau)))/(2*tau**2*np.sinh(
        kappa*tau)**2))
93
94  def V(x,lambda_u): # variance of cost
95      kappa = np.sqrt(lambda_u* sigma ** 2 / eta)
```

24

```
96        return(sigma**2*x**2/2*(tau*np.sinh(kappa*T)*np.cosh(
             kappa*(T-tau))-T*np.sinh(kappa*tau))/(np.sinh(kappa
             *T)**2*np.sinh(kappa*tau)))

97
98   t = np.arange(0,60,1)

99
100  la = np.linspace(9e-7,9e-6,100)
101  plt.scatter(V(X,la),E(X,la))
102  plt.xlabel('V(x)')
103  plt.ylabel('E(x)')
104  plt.title('Efficient_Frontier')

105
106  plt.scatter(t,x(t,5e-6),color = 'red')
107  plt.plot(t,x(t,5e-6),'-',color = 'blue',label = 'lambda=2e
         -6')
108  plt.xlabel('Time_(mins)')
109  plt.ylabel('Inventroy')
110  plt.title('Linear_Model')
```

References

[1] Robert Almgren and Neil Chriss. Optimal execution of portfolio transactions. *Journal of Risk*, 3:5–40, 2001.

[2] Robert F Almgren. Optimal execution with nonlinear impact functions and trading-enhanced risk. *Applied mathematical finance*, 10(1):1–18, 2003.

[3] Dimitris Bertsimas and Andrew W Lo. Optimal control of execution costs. *Journal of Financial Markets*, 1(1):1–50, 1998.

[4] Alvaro Cartea and Sebastian Jaimungal. Incorporating order-flow into optimal execution. *Mathematics and Financial Economics*, 10(3):339–364, 2016.

[5] Alvaro Cartea and Sebastian Jaimungal. Incorporating order-flow into optimal execution. *Mathematics and Financial Economics*, 10(3):339–364, 2016.

[6] Patrick Cheridito and Tardu Sepin. Optimal trade execution under stochastic volatility and liquidity. *Applied Mathematical Finance*, 21(4):342–362, 2014.

[7] Rama Cont, Sasha Stoikov, and Rishi Talreja. A stochastic model for order book dynamics. *Operations research*, 58(3):549–563, 2010.

[8] Rama Cont, Sasha Stoikov, and Rishi Talreja. A stochastic model for order book dynamics. *Operations Research*, 58(3):549–563, 2010.

[9] Sepp Hochreiter and Jürgen Schmidhuber. Long Short-Term Memory. *Neural Computation*, 9(8):1735–1780, 11 1997.

[10] Ulrich Horst and Dörte Kreher. A diffusion approximation for limit order book models. *Stochastic Processes and their Applications*, 129(11):4431–4479, 2019.

[11] Thomas F Loeb. Trading cost: the critical link between investment information and results. *Financial Analysts Journal*, 39(3):39–44, 1983.

[12] Robert C. Merton. Theory of rational option pricing. *The Bell Journal of economics and management science*, pages 141–183, 1973.

[13] Somayeh Moazeni, Thomas F Coleman, and Yuying Li. Optimal execution under jump models for uncertain price impact. *Journal of Computational Finance*, 16(4):1–44, 2013.

[14] Andre F Perold. The implementation shortfall: Paper versus reality. *Journal of Portfolio Management*, 14(3):4, 1988.

[15] André F Perold and Robert S Salomon Jr. The right amount of assets under management. *Financial Analysts Journal*, 47(3):31–39, 1991.

[16] Ilija Zovko and J Doyne Farmer. The power of patience: a behavioural regularity in limit-order placement. *Quantitative Finance*, 2(5):387–392, 2002.