ATQS HW2

Yifan Li, Jingsheng Huang, Caiyi (Anna) Zhang

March 2020

1 Building Impact Model from Public Data

Our first step is to select enough stocks to calculate the required variables, here we create 4 Python files as follows:

1.1 getOriginStockList.py

This file is used to select valid stocks from file *All-CRSP-2007.xlsx*, we have criteria as below

1. stock must have valid cumulative factor data in the excel
2. the Trading Symbol and Ticker Symbol must be the same
3. stock must have full data (with 251 length) in the year of 2007

With the criteria above, we get 353 valid stocks, the names of which are stored in the txt file *ListOfMoreStock.txt*

IMPORTANT: since reading data from *All-CRSP-2007.xlsx* takes much time, we do not provide unit test for this file.

1.2 FileExistenceChecker.py

This file is used to check whether we have the data of the stocks from the stock list, we have criteria as below

1. stock must have full 65-day trade data in our database
2. stock must have full 65-day quote data in our database

With the criteria above and the list from *ListOfMoreStock.txt*, we get 3055 valid stocks, the names of which are stored in the txt file *NewListOfStock.txt*

1.3 StockDivider.py

This file is used to divide the whole stocks into two parts: stocks with high volume and stocks with low volume. Our method is: collecting the total volumes of stocks in *NewListOfStock.txt*, ranking the stock based on total volumes and evenly dividing the stocks into two part. High-Volume stocks list is stored in the txt file *HighVolumeStockList.txt*, Low-Volume stocks list is stored in the txt file *LowVolumeStockList.txt*

1.4 getBasicData.py

This file is used to get access to our database and calculate the required and useful variables for the next model fitting section. The formulas and criteria for calculating different variables are below

1. VWAP_to_4_00PM: calculated by VWAP class
2. VWAP_to_3_30PM: calculated by VWAP class
3. terminal price: the average of the last five mid-quote prices, if the total number of daily quotes is less than five, we set "None" as the value

4. arrival price: the average of the first five mid-quote prices, if the total number of daily quotes is less than five, we set "None" as the value

5. 2-min mid-quote return: calculated by ReturnBuckets class

6. daily volume: the sum of all trading sizes in a day

7. imbalance value (in dollars) : we firstly get the sign for each trade with tickerTester class, then multiply each trading size by its assigned sign and sum them up, finally get imbalance value by multiplying the sum number by VWAP_to_3_30PM, if VWAP_to_3_30PM is invalid number, we set "None" as the imbalance number

8. daily value (in dollars) : calculated by multiplying VWAP_to_4_00Pm by daily volume

1.5 StockSelection.py

This file is used to filter stocks with qualified data from *getBasicData.py*. Since 2-mins returns are essential for the construction of sigma matrix and overall data completeness. We selected stocks with less than 5% missing returns and divide into active and less-active stocks for later analysis. In our case, we assume stocks remained from sp500 are active and the rest is labelled less-active.

1.6 Processing Statistics.py

The main purpose of this part is to process the statistics from the TAQ data for the use of later nonlinear regression. The data set on which we performed analysis contains, before filtering, about 8000 stock trades from NYSE TAQ data from June 2007 through September 2007. In terms of exploratory data analysis, we will perform stock selection according to data completeness and variable interpretation based on Almgren's model.

1.7 Variable Interpretation

We will perform a simple form of Almgren's model based on the following regression function:

$$h = \eta\sigma\left(\frac{X}{\frac{6}{6.5}V}\right)^{\beta} \tag{1}$$

where:

- X is the imbalance value from 9.30AM to 3.30PM
- V is the average daily value
- σ is the standard deviation of 2-mins returns scaled to 1 day
- η is the regression output, value we are looking for
- β is the regression output, value we are looking for

We will make one key departure from the Almgren's model. Almgren uses average daily volume traded, but we will use average daily value traded because this allows us to ignore splits.

1.8 Data Completeness and Processing

The preliminary analysis on the statistics matrix we obtain illustrates a highly completeness of the data except the 2-min returns. We will filter stocks that have more than 5% missing returns in a total 195*65 return window, and conduct further cleaning and computation on those remained stocks of total number 1212.

The three main pieces of data we need to compute across days are

- (i) a 10-day look-back on average daily value
- (ii) a 10-day look-back on the standard deviation of 2-minute mid-quote returns
- (iii) the days that will not be used in our regression

for (i), we simply take moving average of daily value on a window of ten days, which are the average value we will be using in regression.

for (ii), we will perform 10-day rolling standard deviation on 2-min returns, and because of the previous filter, in ten-day window of 195*10 returns, the missing points are relatively smaller than the existing values. Therefore, we will compute the standard deviation of available points in the ten-day window, which is a good estimate if the window contains missing values. The scaled rolling standard deviation is used as sigma matrix in regression.

for (iii), we will drop the first 9 days for the following reasons:

- (1) in terms of market impact, the average value and sigma are computed as rolling values, we should drop days that not in the matrix to match the date size
- (2) a preliminary analysis on 2-min returns illustrates a potential extreme event happened on the 9th day which caused market halted (see table).

	PRU	TJX	GPS	SPG	UNM	CLX	LXK	MS	GLW	ACS	...	ACAS	XLNX	JNY	PDCO	SWY	FIS	ED	HST	MHS	EXPE
1900	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
1901	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
1902	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
1903	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
1904	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
1905	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
1906	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
1907	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
1908	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
1909	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
1910	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
1911	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
1912	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
1913	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
1914	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
1915	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
1916	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
1917	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
1918	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
1919	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
1920	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
1921	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
1922	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
1923	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN

Figure 1: 2-min returns on the 9th day

1.9 doRegression.py

The formulas used to calculate the coefficients and their t-statistics (using heteroskedasticity-robust standard errors) are

$$SE^*(\hat{\beta}_k = \sqrt{\frac{1}{N}(S_{xx}^{-1}\hat{S}S_{xx}^{-1})_{kk}}$$

where

$$S_{xx} = \frac{1}{N}X'X = \frac{1}{N}\sum_{i=1}^{N} x_i x_i'$$

$$\hat{S} = \frac{1}{N}\sum_{i=1}^{N} \hat{\epsilon}_i^2 x_i x_i'$$

$$t_{\hat{\beta}} = \frac{\hat{\beta} - \beta_0}{S.E.(\hat{\beta})}$$

Since p-value is small (¡.01), the parameter η is statistically significant, but the p-value for β is 0.6, so it's not statistically significant.

First, assume heteroskedastic-robust standard error, then for less active stocks, we have $\eta = 0.7114$ and $\beta = .0222$; for active stocks, we have $\eta = 0.6251$ and $\beta = -0.0331$. Assuming regular standard errors, then for active stocks, we have $\eta = 7.6126$ and $\beta = 0.7356$; for less active stocks, we have $\eta = 2.5606$ and $\beta = 0.4071$. Both indicate parameters generated for liquid versus less liquid stocks are different.

Analysis of residuals are shown as scatter plots:

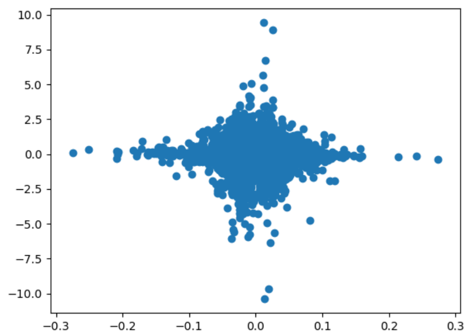

Figure 2: residuals vs $\hat{h}$ from NLS

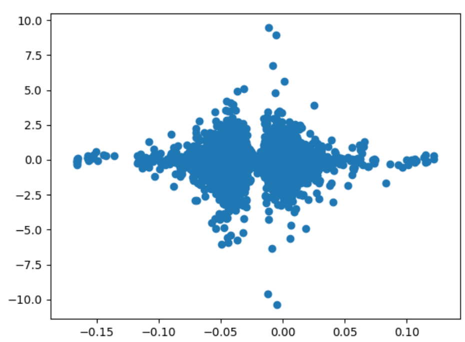

Figure 3: residuals vs $\hat{h}$ from WLS

The residual plots show a point cloud with a horizontal band, this is the preferred pattern as the residuals should be centered on zero throughout the range of fitted values. In other words, the model is correct on average for all fitted values.

2 Almgren-Chriss Meets HJB

1. Formulate Almgren-Chriss optimal execution problem as a stochastic control problem. Assume you have no alpha.

   We define the revenue of the trading strategy as:

$$R = XS_0 + \sigma \int_0^T x(t)dW_t - \frac{\gamma}{2}X^2 - \eta \int_0^T (\dot{x}(t))^2 dt$$

   We have

$$\mathbf{E}[R] = XS_0 - \frac{\gamma}{2}X^2 - \eta \int_0^T (\dot{x}(t))^2 dt$$

$$Var(R) = \sigma^2 \int_0^T x^2(t)dt$$

   where $XS_0 - \frac{\gamma}{2}X^2$ is a constant, so let $B = XS_0 - \frac{\gamma}{2}X^2$. With risk aversion parameter λ, define

$$\tilde{R} = \mathbf{E}[R] - \lambda Var(R) - B = -\int_0^T (\lambda\sigma^2 x^2(s) + \eta\dot{x}^2(s))ds$$

   $\tilde{R}$ can be seen as a function of x and t. We rewrite the formula

$$\tilde{R}(x,t) = -\int_t^T (\lambda\sigma^2 x^2(s) + \eta\dot{x}^2(s))ds$$

Define value function/gain function

$$H(x,t) = \sup_{x(t)\in\mathscr{A}_{t,T}} \mathbf{E}_{x,t}\left[-\int_t^T (\lambda\sigma^2 x^2(s) + \eta\dot{x}^2(s))ds\right]$$

where $\mathscr{A}_{t,T}$ is the set of admissible control laws and $\mathbf{E}_{x,t}$ is the expectation conditional on the state $x^u(t) = x$. Hence, we formally formulate the stochastic control problem with value function

$$H(x,t) = \sup_{x(t)\in\mathscr{A}_{t,T}} \mathbf{E}_{x,t}\left[-\int_t^T (\lambda\sigma^2 x^2(s) + \eta\dot{x}^2(s))ds\right]$$

we need to find optimal control $\tilde{x} \in \mathscr{A}_{t,T}$ to realize $H(x,t)$.

2. Derive HJB equation and solve for the control and value function. What solution do you obtain?

   With Theorem 3 (The Hamilton-Jacobi-Bellman Equation), H satisfies the HJB

   $$\partial_t H(t,x) + \sup_{x(t)} \left(\mathcal{L}_t^u H(t,x) + F(t,x,u)\right) = 0$$

   with condition $H(T,x) = G(x)$.

   With the model we build,

   $$F(t,x,u) = -\lambda\sigma^2 x^2(t) - \eta\dot{x}^2(t), \quad G(x) \equiv 0$$

   since $dx(t) = \dot{x}(t)dt$, then $\mathcal{L}_t^u H(t,x) = \dot{x}(t)\partial_x H$, and since the value of $x(t)$ can be uniquely determined by derivative $\dot{x}(t)$ with specific boundary conditions $x(0) = X, x(T) = 0$, we can rewrite the PDE as

   $$\partial_t H - \lambda\sigma^2 x^2 + \sup_{\dot{x}(t)\in\mathscr{A}_{t,T}^*} (\dot{x}\partial_x H - \eta\dot{x}^2) = 0$$

   when $\dot{x} = \frac{\partial_x H}{2\eta}$, $\dot{x}\partial_x H - \eta\dot{x}^2$ reach its supremum, then the PDE can be transformed to

   $$\partial_t H - \lambda\sigma^2 x^2 + \frac{(\partial_x H)^2}{4\eta} = 0$$

   with condition $H(T,x) = 0$.

   To solve the PDE, let $H(t,x) = x^2 K(T-t)$, then

   $$\begin{aligned}\partial_t H &= -x^2\partial_t K(T-t)\\ \partial_x H &= 2xK(T-t)\end{aligned}$$

   so we have

   $$-x^2\partial_t K(T-t) - \lambda\sigma^2 x^2 + \frac{1}{4\eta}4x^2K^2(T-t) = 0$$

   since $x^2 \geq 0$ and when $x = 0$ the equation still holds:

   $$\begin{aligned}-\partial_t K - \lambda\sigma^2 + \frac{1}{\eta}K^2 &= 0\\ -\partial_t K + \frac{1}{\eta}K^2 &= \lambda\sigma^2\end{aligned}$$

   We have the solution

   $$K(T-t) = \sqrt{\lambda\eta\sigma^2}\tanh\left[\frac{c_1\sqrt{\lambda\sigma^2}\eta + \sqrt{\lambda\sigma^2}(T-t)}{\sqrt{\eta}}\right]$$

   where c_1 an undefined constant. Therefore

   $$H(t,x) = x^2 K(T-t)$$

   since $H(T,x) = x^2K(0) = 0$, we should have $K(0) = 0$.

   Thus,

   $$\begin{aligned}K(0) = \tanh\left[\frac{c_1\sqrt{\lambda\sigma^2}\eta}{\sqrt{\eta}}\right] &= 0\\ c_1 &= 0\\ H(t,x) = x^2\sqrt{\lambda\eta\sigma^2}\tanh\left[\frac{\sqrt{\lambda\sigma^2}(T-t)}{\sqrt{\eta}}\right]\end{aligned}$$

with optimal control

$$\tilde{x}(t) = X\frac{\sinh\big(\kappa(T-t)\big)}{\sinh(\kappa T)}$$

where $\kappa = \sqrt{\frac{\lambda\sigma^2}{\eta}}$ and value function

$$H(t,x) = X^2\left[\frac{\sinh\big(\kappa(T-t)\big)}{\sinh(\kappa T)}\right]^2 \kappa\eta\cdot\tanh\big(\kappa(T-t)\big)$$

3. Solve the same problem assuming you have alpha. What solution do you obtain? How large does alpha have to be to have an impact?

   Taking α into consideration, we should insert another drift term into the SDE of the stock price $S(t)$, then it will be

$$dS = \sigma dW + (\alpha + \gamma\dot{x}(t) + \eta\ddot{x}(t))dt$$

   or equivalently,

$$S(t) = S_0 + \sigma W_t + \alpha t - \gamma(X - x(t)) + \eta\dot{x}(t)$$

   Then define the revenue of the trading strategy $x(t)$:

$$\begin{aligned}
R' &= \int_0^T v(t)S(t)dt \\
&= -\int_0^T S(t)dx(t) \\
&= -S(t)x(t)|_0^T + \int_0^T x(t)dS(t) \\
&= ① + ②
\end{aligned}$$

   where

$$\begin{aligned}
① &= -\left[S_0 + \alpha W_t + \alpha t - \gamma(X - x(t)) + \eta\dot{x}(t)\right] x(t)|_0^T \\
&= X[S_0 + \eta\dot{x}(0)] \\
② &= \sigma\int_0^T x(t)dW_t + \alpha\int_0^T x(t)dt - \frac{\gamma}{2}x^2 + \eta\int_0^T x(t)d\dot{x}(t) \\
&= \sigma\int_0^T x(t)dW_t + \alpha\int_0^T x(t)dt - \frac{\gamma}{2}x^2 - \eta X\dot{x}(0) - \eta\int_0^T (\dot{x}(t))^2dt \\
\Rightarrow R' &= XS_0 - \frac{\gamma}{2}X^2 + \sigma\int_0^T x(t)dW_t + \int_0^T \alpha x(t) - \eta(\dot{x}(t))^2dt \\
\mathbf{E}[R'] &= XS_0 - \frac{\gamma}{2}X^2 + \int_0^T \alpha x(t) - \eta(\dot{x}(t))^2dt \\
Var(R') &= \sigma^2\int_0^T x^2(t)dt
\end{aligned}$$

   Let $B = XS_0 - \frac{\gamma}{2}X^2$, which is a constant.
   Similarly, we define

$$\tilde{R}' = \mathbf{E}[R'] - \lambda Var(R') - B = \int_0^T \alpha x(t) - \lambda\sigma^2x^2(t) - \eta(\dot{x}(t))^2dt$$

   and

$$\tilde{R}'(t,x) = \int_t^T \alpha x(s) - \lambda\sigma^2x^2(s) - \eta(\dot{x}(s))^2ds$$

   then the value function should be

$$H'(t,x) = \sup_{x(t)\in\mathscr{A}_{t,T}} \mathbf{E}_{x,t}\left[\int_t^T \alpha x(s) - \lambda\sigma^2x^2(s) - \eta\dot{x}^2(s)ds\right]$$

   with Theorem 3 (HJB) we can derive the PDE for H'.

$$\partial_t H' + \sup_{x(t)\in\mathscr{A}_{t,T}}\Big(\dot{x}(t)\partial_x H' + \alpha x(t) - \lambda\sigma^2x^2(t) - \eta\dot{x}^2(t)\Big) = 0$$

With the same method, we transfer the PDE into

$$\partial_t H' + \alpha x(t) - \lambda\sigma^2 x^2(t) + \sup_{\dot{x}\in\mathscr{A}^*_{t,T}} \Big(\dot{x}(t)\partial_x H' - \eta\dot{x}^2(t)\Big) = 0$$

then $\dot{x}(t) = \frac{\partial_x H}{2\eta}$ reaches the supremum of the last part in LHS. Thus we have

$$\partial_t H' + \alpha x(t) - \lambda\sigma^2 x^2(t) + \frac{(\partial_x H')^2}{4\eta} = 0$$

Equivalently,

$$\partial_t H' - \lambda\sigma^2 \left[x - \frac{\alpha}{2\lambda\sigma^2}\right]^2 + \frac{\alpha^2}{4\lambda\sigma^2} + \frac{1}{4\eta}\left(\partial_x H'\right)^2 = 0$$

Let $x' = x - \frac{\alpha}{2\lambda\sigma^2}$, then $\partial_x H' = \partial_{x'} H'$, so we have

$$\partial_t H' - \lambda\sigma^2 x'^2 + \frac{\alpha^2}{4\lambda\sigma^2} + \frac{1}{4\eta}(\partial_{x'} H')^2 = 0$$

Let $H' = x'^2 K(T-t) + \frac{\alpha^2}{4\lambda\sigma^2}(T-t)$, then

$$\begin{aligned}
\partial_t H' &= -x'^2 \partial_t K(T-t) - \frac{\alpha^2}{4\lambda\sigma^2} \\
\partial_{x'} H' &= 2x' K(T-t) \\
\Rightarrow -x'^2\partial_t K(T-t) - \frac{\alpha^2}{4\lambda\sigma^2} - \lambda\sigma^2 x'^2 + \frac{\alpha^2}{4\lambda\sigma^2} &+ \frac{1}{4\eta} 4x'^2 K^2(T-t) = 0 \\
\Rightarrow -x'^2\partial_t K(T-t) - \lambda\sigma^2 x'^2 + \frac{1}{4\eta} 4x'^2 K^2(T-t) &= 0
\end{aligned}$$

since $x'^2 \geq 0$ and when $x' = 0$ the equation still holds, we have

$$-\partial_t K(T-t) - \lambda\sigma^2 + \frac{1}{\eta} K^2(T-t) = 0$$

Similarly, with the previous part, we can get

$$K(T-t) = \sqrt{\lambda\eta\sigma^2}\tanh\left(\frac{c_1\sqrt{\lambda\sigma^2}\eta + \sqrt{\lambda\sigma^2}(T-t)}{\sqrt{\eta}}\right)$$

$$H(t, x') = x'^2 K(T-t) + \frac{\alpha^2}{4\lambda\sigma^2}(T-t)$$

where c_1 is an undefined parameter.
Since $H(T,x) = 0$ for all $x \in \mathscr{A}_{t,T}$, we have $H(T,x') = 0 = x'^2 K(0)$, so $K(0) = 0$, $c_1 = 0$.

$$\Rightarrow H(t,x') = x'^2\sqrt{\lambda\eta\sigma^2}\tanh\left(\frac{\sqrt{\lambda\sigma^2}(T-t)}{\sqrt{\eta}}\right) + \frac{\alpha^2}{4\lambda\sigma^2}(T-t)$$

Since $x' = x - \frac{\alpha}{2\lambda\sigma^2}$, the value function is

$$H(t,x) = \left(x - \frac{\alpha}{2\lambda\sigma^2}\right)^2 \kappa\eta\tanh\left(\kappa(T-t)\right) + \frac{\alpha^2}{4\lambda\sigma^2}(T-t)$$

Now we need to find the optimal control function by solving the following problem:

$$\begin{aligned}
\underset{x(t)}{\text{maximize}} \quad & \int_0^T \alpha x(t) - \lambda\sigma^2 x^2(t) - \eta(\dot{x}(t))^2 dt \\
\text{subject to} \quad & x(0) = X \\
& x(T) = 0
\end{aligned}$$

Similar to the procedure from lecture notes, define

$$F(x, \dot{x}, t) = \alpha x - \lambda\sigma^2 x^2 - \eta\dot{x}^2$$

the answer if fully defined by the following ODE:

$$\alpha x - \lambda\sigma^2 x^2 + \eta\dot{x}^2 = 0 \tag{2}$$
$$x(0) = X \tag{3}$$
$$x(T) = 0 \tag{4}$$

Differentiate (2) we get

$$\alpha\dot{x} - 2\lambda\sigma x\dot{x} + 2\eta\dot{x}\ddot{x} = 0$$
$$\dot{x}(\alpha - 2\lambda\sigma^2 x + 2\eta\ddot{x}) = 0$$

ignore the trivial solution $\dot{x} = 0$ and consider $\alpha - 2\lambda\sigma^2 x + 2\eta\ddot{x} = 0$, it is a second order non-homogeneous linear ODE with solution

$$x(t) = \frac{\alpha}{2\lambda\sigma^2} + c_1 e^{-\kappa t} + c_2 e^{\kappa t}$$

where $\kappa = \sqrt{\frac{\lambda\sigma^2}{\eta}}$.

Let $\theta = \frac{\alpha}{2\lambda\sigma^2}$, then $x(t) = \theta + c_1 e^{-\kappa t} + c_2 e^{\kappa t}$ with conditions $x(0) = X$, $x(T) = 0$. We get

$$c_1 = \frac{\theta e^{\kappa T}}{e^{2\kappa T} - 1} + \frac{e^{2\kappa T}(x - \theta)}{e^{2\kappa T} - 1}$$
$$c_2 = \frac{-(x - \theta)}{e^{2\kappa T} - 1} - \frac{\theta e^{\kappa T}}{e^{2\kappa T} - 1}$$

then the optimal control is

$$\tilde{x}(t) = \theta + c_1 e^{-\kappa t} + c_2 e^{\kappa t}$$

the value function is

$$H(t, x) = (c_1 e^{-\kappa t} + c_2 e^{\kappa t})^2 \kappa\eta \tanh\big(\kappa(T - t)\big) + \frac{\alpha^2}{4\lambda\sigma^2}(T - t)$$

where $\kappa = \sqrt{\frac{\lambda\sigma^2}{\eta}}$, $\theta = \frac{\alpha}{2\lambda\sigma^2}$,

$$c_1 = \frac{\theta e^{\kappa T}}{e^{2\kappa T} - 1} + \frac{e^{2\kappa T}(x - \theta)}{e^{2\kappa T} - 1}$$
$$c_2 = \frac{-(x - \theta)}{e^{2\kappa T} - 1} - \frac{\theta e^{\kappa T}}{e^{2\kappa T} - 1}$$

indicating alpha has great impact on the form of value function and optimal control.

3 Standard concepts of statistical trading

1. 4 most commonly used high-frequency trading strategies:

   - passive market-making strategy: involves submission of non-marketable resting orders to provide liquidity, primary source of profits are from the spread and liquidity rebates offered by trading centers to liquidity-supplying orders
   - arbitrage strategy: capture pricing inefficiencies between related products or markets, usually involves taking liquidity
   - structural strategy: capture structural vulnerabilities in the markets or in certain market participants
   - directional strategy: a speculative strategy that anticipates price movement in a particular direction to capture short-term profit (e.g. order anticipation, momentum ignition)

   According to SEC's Concept Release on Equity Market Structure report, these are the 4 broad types of strategies often used by proprietary trading firms. (Reference: https://www.sec.gov/rules/concept/2010/34-61358.pdf)

2. Provide rough estimate of the profitability of high frequency traders in today's equity market. How do they use leverage? Motivate your assumptions, and how you come to the conclusion.

   According to estimates from consultancy firm TABB Group, HFT in the US have seen revenue from the equity markets collapse from a peak of \$7.2 billion in 2009 to below \$1 billion in 2017 for the first time since the financial crash. (Reference: https://www.ig.com/en/trading-strategies/high-frequency-trading-explained–why-has-it-decreased–181010)

   Our assumption is that HFT often use leverage on stocks with high liquidity. The motivation is that high liquidity is associated with large volume, which creates small spreads, and in order to obtain material profit, HFT need to use leverage so they can trade in large volumes.

3. Does high frequency trading impose risks of systemic nature?

   High frequency trading can impose systemic risks by deteriorating liquidity at times of crisis, according to a recent article on the coronavirus mayhem. The reason is that before HFT, it's mostly the banks that accounted for the market-making, but as HFT becomes more common and more regulations imposed on banks, HFT nowadays dominates much of the market-making activity, mainly by providing liquidity, capturing and narrowing the spreads. The most recent example is the market turmoil as a consequence of COVID-19. Volatility spikes, and high frequency traders pull back on their capital, deteriorating market depth. There are a lot of panic selling when market is volatile, only the liquidity once offered by HFT is not there, precisely when it's needed the most. I agree with the explanation in the article.

4. Propose your own intraday equity trading strategy. Describe what methodology and data you would use to research your idea. Provide a "guesstimate" of its performance (Sharpe ratio).

   An intraday equity trading strategy could be event-driven or based on market sentiment. The methodology involves using natural language processing (NLP) and sentiment analysis to analyze data such as market news, social media tweets, earnings conference calls. For example, we can obtain reliable data from twitter, Webhose, or news API, and we find the most impactful news (with max and min sentiment scores) of a given day, we can then experiment with different strategies such as long/short stocks with the most positive/negative sentiments, or excluding those and trade only stocks with lower volatilities. A guesstimate of its performance measured in Sharpe ratio could be around 20%.