Mathematical Market Microstructure – An Optimization Approach

Lecture III – On the optimal design of execution algorithms

FINM 37601 – Fall, 2023

By Hongsong Chou, Ph.D.

AGENDA

- Trade-off between execution risk and market impact – the concept of "execution efficient frontier"

- Discussion on market impact modeling

- A generalized quadratic optimization framework for optimal trade scheduling problem

AGENDA

- Trade-off between execution risk and market impact – the concept of “execution efficient frontier”

- Discussion on market impact modeling

- A generalized quadratic optimization framework for optimal trade scheduling problem

THE DEFINITION OF SLIPPAGE

- The performance of a VWAP algo is often measured as a difference between the execution price of the algo's trade and a market VWAP:

$$VWAP\ Slippage = (VWAP\ of\ the\ stock\ in\ market\ - VWAP\ of\ algo's\ executions) * SignOfTrade \quad (1.1)$$

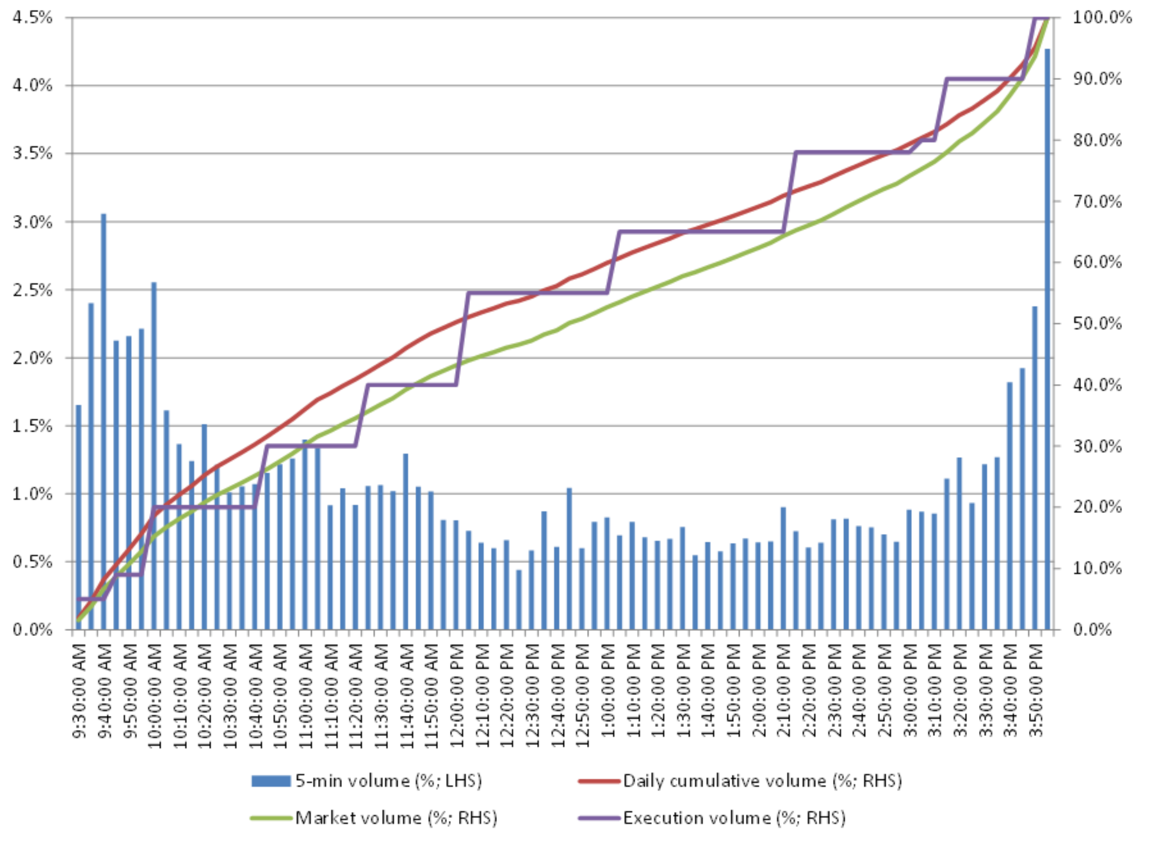
- It is often expressed in basis points by dividing the above slippage by the market VWAP, then multiplied by 10000;
- If we use i to index individual "slices" of a VWAP trade (i = 1, 2, …), P_i to index market's prices "bucketed around" the ith slice of the algo's execution, p_i the trade price of the ith execution of the algo, v_i the percentage of the market's trade volume "bucketed around" the ith slice of the algo's execution, w_i the percentage of the trade size of the ith execution of the algo among the total execution volume of the algo order, then we can re-write definition (1) above as:

$$VWAP\ Slippage\ (for\ Buy) = \sum_i P_i v_i - \sum_i p_i w_i$$

- Note that v_i and w_i are also called "realized market volume profile" and "realized execution profile".

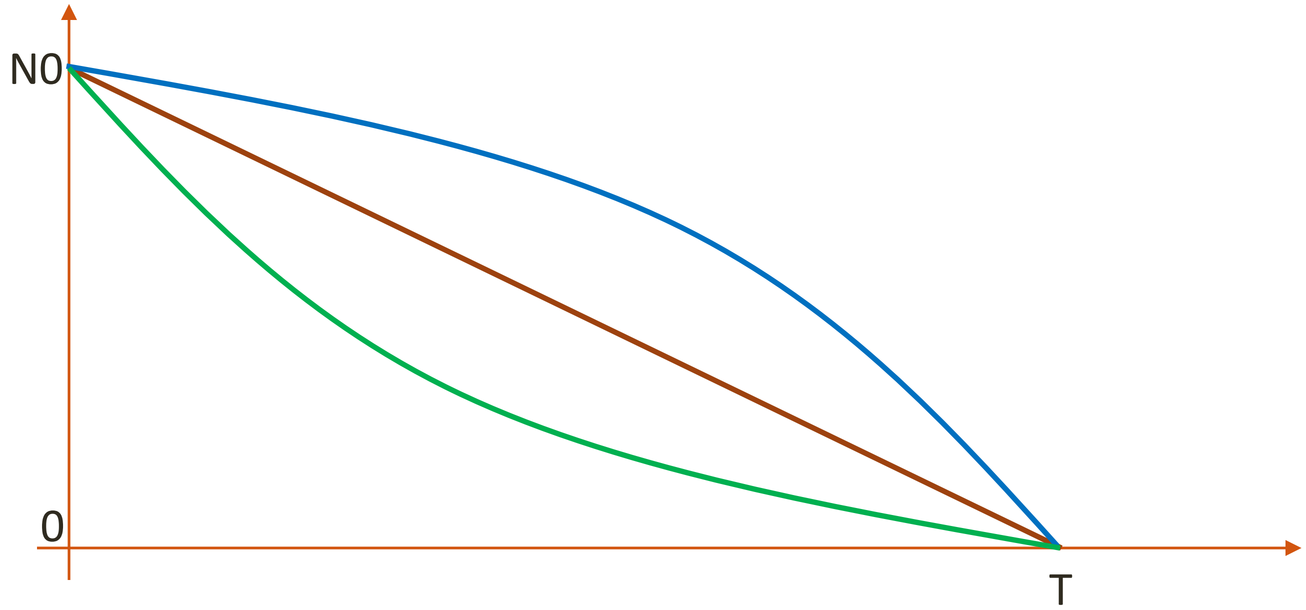
EXECUTION PROFILES (I)

- Trade schedules:



EXECUTION PROFILES (II)

- Liquidation schedules:



$$N(t) = N_0(1 - \frac{t}{T})$$

EXECUTION RISK DURING LIQUIDATION

- Assuming a linear liquidation profile:

$$N(t) = N_0(1 - \frac{t}{T})$$

- Therefore,

$$V(t) = N(t)P(t) = P(t)N_0(1 - \frac{t}{T})$$

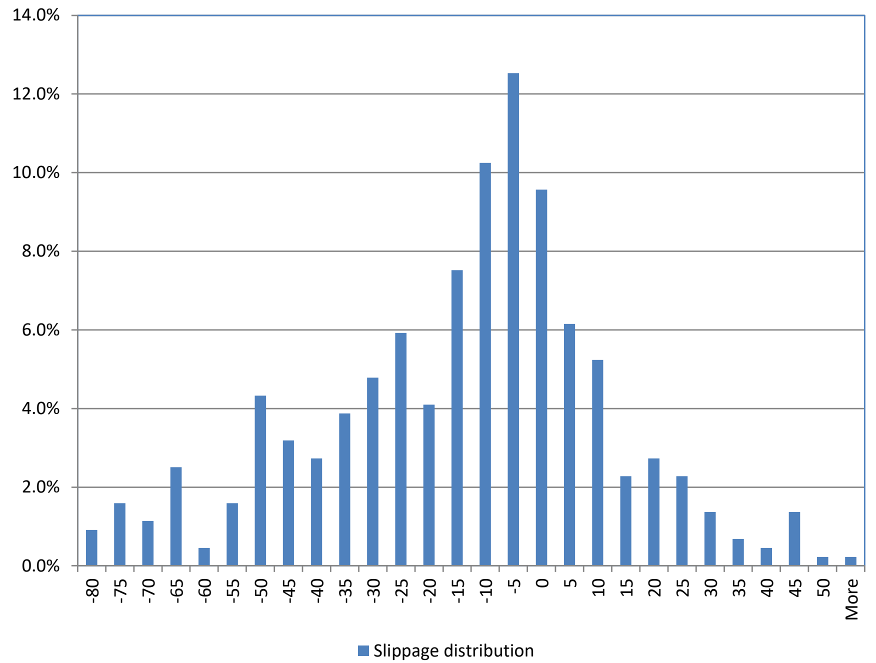
- The "holding" risk of a position that is being liquidated for a short duration is:

$$Var(V)dt \approx (P_0N_0)^2\sigma^2\, dt\left(1 - \frac{t}{T}\right)^2$$

- Therefore, we can calculate the total "holding risk" during liquidation as:

$$Var_{tot}(V) \approx (P_0N_0)^2\sigma^2 \int_0^T dt\left(1 - \frac{t}{T}\right)^2 = \frac{T}{3}V_0^2\sigma^2$$

MEAN VS. STDEV. OF SLIPPAGE



QUADRATIC OPTIMIZATION PROBLEM (I)

- The well-known Markowitz portfolio theory suggests to maximize a quadratic utility function (or, in optimization, an objective function) that combines portfolio return (excess risk free rate) and variance:

$$\sum_{i=1}^{n} w_i r_i - \lambda \sum_{i,j=1}^{n} w_i w_j \sigma_i \sigma_j \rho_{ij}$$

- In trading, the objective function has to add one more term: transaction cost:

$$\sum_{i=1}^{n} w_i r_i - \lambda \sum_{i,j=1}^{n} w_i w_j \sigma_i \sigma_j \rho_{ij} - \mu G(x_i, t_i | i = 1,2, \dots, n)$$

- Here *G* is the market impact function which is positive definite and a generally concave form of the execution profiles of individual stocks in the portfolio; there can be correlations incorporated but in practice often omitted due to both theoretical and practical complexity.

QUADRATIC OPTIMIZATION PROBLEM (II)

- For traders (and their algos), although short-term alphas are very important to good executions, the returns in the sense of "portfolio managers" are not critical; therefore, algo optimization problem often ignores the first term on previous page and keep only the risk and market impact terms:

$$\min\left\{\lambda \sum_{i,j=1}^{n} w_i w_j \sigma_i \sigma_j \rho_{ij} + \mu G(x_i, t_i | i = 1,2,\dots,n)\right\}$$

- The arguments for the minimization operation are the execution profiles, which are path dependent on market activities.

QUADRATIC OPTIMIZATION PROBLEM (III)

- Consider a (very much) simplified problem in the case of a single stock execution:

$$\min\left\{\lambda\sigma\sqrt{\frac{T}{3}}+\mu C\left(\frac{X}{ADV\cdot T}\right)^{\gamma}\right\}$$

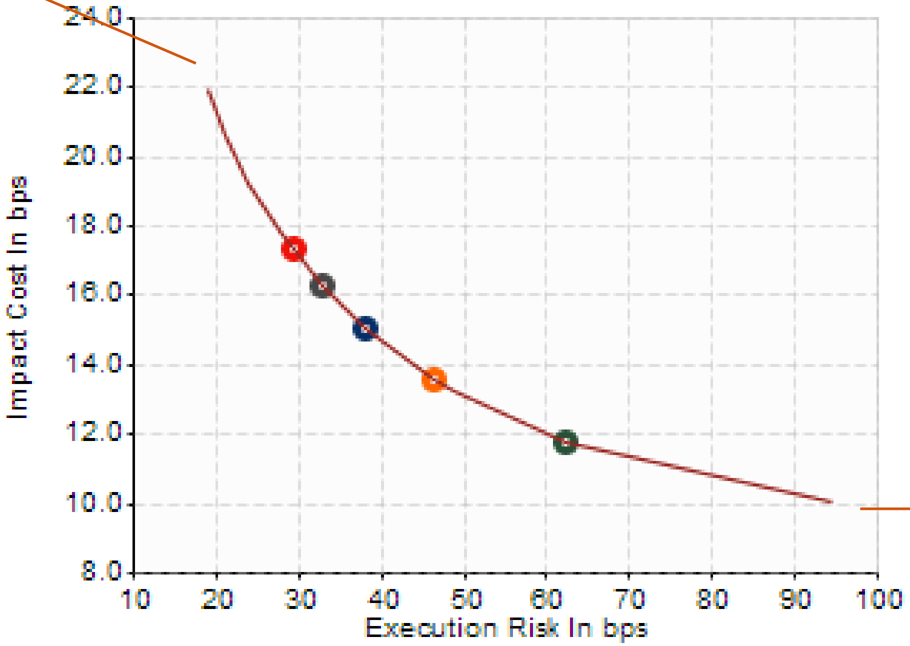
- This means that, assuming the only variable is time, we have

$$T_{*}^{-\frac{1}{2}-\gamma}=\frac{\lambda\sigma}{\sqrt{12}\cdot\mu C\gamma\cdot\left(\frac{X}{ADV}\right)^{\gamma}}\,.$$

QUADRATIC OPTIMIZATION PROBLEM (IV)

- Let’s look at the case of a stock:

In limit, market order is the one theoretically minimizes risk.



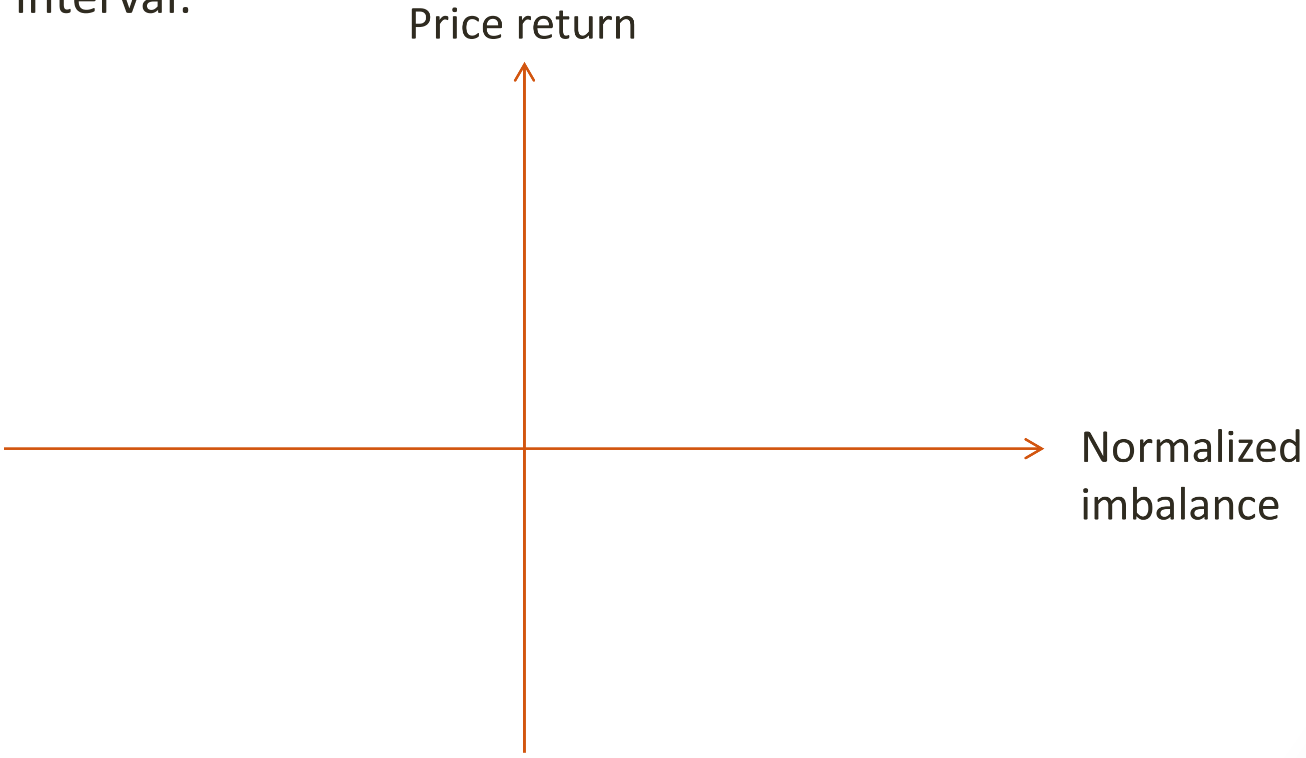
In limit, VWAP algo is the one theoretically minimizes impact.

AGENDA

- Trade-off between execution risk and market impact – the concept of “execution efficient frontier”
- Discussion on market impact modeling
- A generalized quadratic optimization framework for optimal trade scheduling problem

DISCUSSION ON MARKET IMPACT MODELS (I)

- Empirical studies have found for different assets and across different markets, there seems to hold a concave relation between net imbalance of transactions and the price movement measured in the same time interval:



DISCUSSION ON MARKET IMPACT MODELS (II)

- Price movement forecasting models vs. market impact models:
  - In many cases, a market impact model is also a price forecasting model;
  - But there is a key difference between the two: the impact of one's own trade on the price movement in the next time period;
  - Revisiting the Roll model: $\Delta p_t = u_t + (q_t - q_{t-1})c$;
  - Another key difference (esp. from practical point of view) is that a typical market impact model needs to span a certain period of time, which is often the execution horizon of an order; typical market microstructure price forecasting model tends to predict a rather short horizon.

DISCUSSION ON MARKET IMPACT MODELS (III)

- Key parameters that need to be considered in a market impact model:
  - Order size, X;
  - Expected execution horizon, T;
  - Microstructure variables, such as a stock's volatility, σ, bid-ask spread Spd, etc.;
- The three components of market impact:
  - Instantaneous impact;
  - Temporary impact;
  - Permanent impact;
- A figurative description of the three components of market impact:

DISCUSSION ON MARKET IMPACT MODELS (IV)

- Consider the following market impact model:

$$MarketImpact = C_1 \cdot Spd + C_2 \cdot \left(\sigma\sqrt{T}\right)^{\alpha} \cdot \left(\frac{X}{ADV \cdot T}\right)^{\beta} + C_3 \cdot \sigma^2 \cdot \left(\frac{X}{ADV}\right)^2 \tag{2.1}$$

- Here $C_{1,2,3}$, α, β are parameters of the model and need to be fitted from data; X is the size of the order, ADV is the average daily volume of the stock which is a measure of liquidity, T is the execution horizon, and σ is the realized volatility of the stock;
- Note that $\frac{\alpha}{2} - \beta < 0$ to ensure that market impact decreases as execution horizon is lengthened.

AGENDA

- Trade-off between execution risk and market impact – the concept of “execution efficient frontier”
- Discussion on market impact modeling
- A generalized quadratic optimization framework for optimal trade scheduling problem

A GENERAL FRAMEWORK (I)

- The key variable that we need to solve in execution algo problem is the optimal execution profile, $y_i(t)$, which is the "trade rate" of the algo on stock *i* and *i*=1,2,…,*n;* therefore, $y_i(t)dt$, is the number of shares of stock *i* that need to be executed in time dt;
- Assume that the corresponding "market trading rate" for stock *i* is $m_i(t)$, which means that, if we assume the total number of shares of stock *i* is $N_i\ \{i = 1,2,\dots,n\}$, then we can introduce two more variables:

$$Y_i(t) = N_i - \int_0^t y_i(\tau)d\tau\,, \tag{3.1}$$

$$M_i(t) = N_i - \int_0^t m_i(\tau)d\tau\,. \tag{3.2}$$

- Just a note that $m_i(t)$ can be quite general, not necessarily related to the market volume profiles, as we will discuss later;
- Here, *T* is the maximum duration allowed for the algo to trade; note that the optimal solution does not require the executions of child orders to be spread out from the whole duration; it can be significantly "front-loaded" if necessary.

A GENERAL FRAMEWORK (II)

- The execution slippage between our algo execution and a benchmark can be written as (using buy order as an example):

$$IS = \int_0^T m_i(t) S_i(t, y_{\{\le t\}}) dt \; - \; \int_0^T y_i(t) \widehat{S}_i(t, y_{\{\le t\}}) dt \qquad (3.3)$$

- Note the "path dependence" of the execution price, $\widehat{S}_i(t, y_{\{\le t\}})$, and the market price, $S_i(t, y_{\{\le t\}})$, the latter of which should includes the impact of our trades on the market;
- To frame the problem in the general Markowitz quadratic optimization format, we need to formulate $E[IS]$ and $Var[IS]$;
- To simplify the problem without the loss of generality, we need to establish the relationship between $\widehat{S}_i(t, y_{\{\le t\}})$ and $S_i(t, y_{\{\le t\}})$; we construct this relationship by taking into account market impacts from our trades that can be expressed in explicit formulas;
- To make things simple, we denote $\widehat{S}_i(t, y_{\{\le t\}})$ and $S_i(t, y_{\{\le t\}})$ as $\hat{S}_{i,t}$ and $S_{i,t}$.

A GENERAL FRAMEWORK (III)

- If we assume that the impact is generated by our own trades only, then a model for the market price can be expressed as:

$$S_{i,t} = S_i \left[Z_i + \int_0^t G_i(y_i(\tau), t-\tau) d\tau + F_i \left(\int_0^t y_i(\tau) d\tau \right) \right] \qquad (3.4)$$

- The second and the third terms are the temporary and permanent impacts, respectively; note that they are "deterministic"; the first term can be modeled as a Martingale, and S_i is the initial price of stock *i*; which is known at t=0;
- The execution price will be related to the market price through an instantaneous impact term like below:

$$\hat{S}_{i,t} = S_{i,t} - S_i \cdot sign((y_i(t)) \cdot H_i((y_i(t)). \qquad (3.5)$$

A GENERAL FRAMEWORK (IV)

- Now we can write both $E[IS]$ and $Var[IS]$ in the following way:

$$E[IS] = \sum_{i=1}^{n} \int_{t=0}^{T} S_i \cdot |y_i(t)| \cdot H_i dt -$$
$$\sum_{i=1}^{n} \int_{t=0}^{T} dt \cdot S_i \cdot \left(y_i(t) - m_i(t)\right) \left[\int_{0}^{t} G_i(y_i(\tau), t-\tau) d\tau + F_i \left(\int_{0}^{t} y_i(\tau) d\tau \right) \right];$$

$$VAR[IS] = \sum_{i,j=1}^{n} \int_{0}^{T} dt \cdot S_i S_j \cdot [Y_i(t) - M_i(t)] \Sigma_{ij} \left[Y_j(t) - M_j(t)\right]$$

- Here Σ_{ij} is the covariance matrix of the portfolio.

SIDE MARK: REVISITING THE SLIPPAGE BREAKDOWN

- The performance of a VWAP algo depends on three profiles: model profile, market realized profile and execution profile:

$$VWAP\ Slippage\ (for\ Buy) = \sum_i P_i v_i - \sum_i p_i w_i$$
$$= \sum_i P_i(v_i - m_i) + \sum_i P_i(m_i - w_i) + \sum_i (P_i - p_i) w_i;$$

- In the above equation, the **first term** indicates the slippage component that is due to the difference between realized market volume profile, v_i, and the theoretical model profile, m_i; the **second term** indicates the slippage component due to the difference between theoretical model profile and realized execution profile; the **third term** indicates the slippage component due to the difference between market trade price and the execution price of this individual VWAP order (at child order level).

ABOUT BENCHMARKS

- Note that the "trade rate" of the market, $m_i(t)$ as defined in equation (3.2) on page 19, can be quite general:
  - For VWAP benchmark, $m_i(t)$ essentially gives the *i*th stock's market volume profile (per unit time distribution, not cumulative);
  - For implementation shortfall, $m_i(t) = \delta(t)N_i$ where $\delta(t)$ is the Dirac delta function;
  - For TargetClose, $m_i(t) = \delta(t - T_{close})N_i$;
  - Etc.
- To generalize our discussion, $m_i(t)$ can be the "trade rate" of any other strategy as long as it is pre-determined. Therefore, the expected slippage and its standard deviation are both measured between our own trade schedules (i.e., the to-be-solved $y_i(t)$) and a deterministic trade schedule as a benchmark.

HANDLING MARKET IMPACT FUNCTIONS

- The instantaneous impact is rather simple: H_i=$C_1 \cdot Spd_i$;
- The temporary impact can be written in different ways; to reflect the "lingering" effect of this term, $G_i(y_i(\tau), t-\tau)$ can be written as

  $$G_i(y_i(\tau), t-\tau) = C_{temp} \cdot sign(y_i(\tau)) \cdot \left(\sigma(\tau)\sqrt{t-\tau}\right)^{\alpha} \cdot \left|\frac{y_i(\tau)}{(t-\tau)\cdot ADV}\right|^{\beta};$$
- The permanent impact can be written as: $F_i\left(\int_0^t y_i(\tau)d\tau\right) = C_{temp} \cdot \sigma(\tau)^2 \cdot \left(\frac{\int_0^t y_i(\tau)d\tau}{ADV}\right)^2$.

HANDLING RISK CALCULATIONS

- The execution variance term is given by:

$$VAR[IS] = \sum_{i,j=1}^{n} \int_0^T dt \cdot S_i S_j \cdot [Y_i(t) - M_i(t)]\Sigma_{ij} \left[Y_j(t) - M_j(t)\right]$$

- In practice, the variance-covariance matrix Σ_{ij} is often represented with a factor-based risk model and can be written as (using PCA risk model as an example):

$$\Sigma_{ij} = \sum_{m=1}^{K} L_{im} L_{mj} + \varepsilon_i \varepsilon_j \delta_{ij}$$

- Here L is the factor loading matrix with the number of factors to be set at K; and ε_i is the idiosyncratic risk for the ith stock;
- Note that the to-be-solved trade schedule, $y_i(t)$, is summed over time and represented by $Y_i(t)$.

CASTING IT AS AN OPTIMIZATION PROBLEM

- The objective function can be written as:

$$\min_{y_i(t), i=1,2,..,n} \{-E[IS] + \lambda \cdot VAR[IS]\} \qquad (3.6)$$

- Here, λ is a risk aversion factor that has to be set beforehand by trader;
- The boundary conditions to the "static" optimization problem (3.6) can be set generically as $LB_q \leq C(y) \leq UB_q$ where *q* indicates the *qth* constraints on the to-be-determined variables $y_i(t), i = 1,2,..,n$ for t in [0, t].

SOLVING THE OPTIMIZATION PROBLEM

- Usually, problem (3.6) can be solved in three steps:
  - Step 1, minimize the expected IS term so as to obtain a minimum impact cost, $Cost_{min}$;
  - Step 2, minimize the variance of IS term so as to obtain a maximum impact cost, $Cost_{max}$;
  - Step 3, solve the following optimization problem:

$$\min_{y_i(t), i=1,2,..,n} \{VAR[IS]\}$$

$$\text{s.t. } E[IS] \leq (1-\gamma) \cdot Cost_{min} + \gamma \cdot Cost_{max}.$$

THAT'S ALL FOR THIS LECTURE.

THANK YOU!