Optimal Execution Price for Off-exchange Trading*

Seiya Kuno†
Faculty of Commerce, Doshisha University

Abstract

We consider the optimal price to use in off-exchange transactions, taking into account both exchange and off-exchange transactions. Generally, when referring to the closing price of the exchange in off-exchange transactions, there is a possibility of market manipulation by both the institutional investor (principal) and the broker-dealer (agent), so the desirable price here is a price that eliminates the possibility of market manipulation. Specifically, we take the standpoint of the institutional investor in particular and show that VWAP is optimal under certain conditions, and consider the possibility of market manipulation when the conditions are relaxed.

1 Introduction

Institutional investors generally execute large volumes of securities. Therefore, when executing only on a stock exchange, it is necessary to take into account the price impact, which is the fluctuation in price per unit trading volume. Many optimal execution papers have shown that when executing on an exchange, the cost of this impact can be reduced by dividing a large amount of execution (see [1] and [3] for seminal papers). On the other hand, if there is a counterparty outside the exchange that will accept large volumes of securities without the price impact, many institutional investors will consider using the off-exchange trading venue first, and then execute the rest on the exchange. These off-exchange executions may be conducted under the system of an exchange as off-floor trading, or may take place in a trading venue such as a dark pool where no execution information is made public. Tradings between institutional investors and their counterparties (securities companies) are formulated as a principal-agent problem, with the institutional investors (clients) as the principals and the securities companies (dealers) as the agents. In agency trading, the agent is generally a broker, and the broker decides the execution price while considering the benefit of the client. In principal trading, the client forms the contract, and the agent is generally a dealer, so the client decides the transaction price while considering the profit of the agent. As stated in many publications, e.g. [7], using multiple trading venues makes it possible, in principle, to manipulate prices. For example, if a contract is made to purchase 100,000 shares at the closing price on the stock exchange on that day, it is theoretically possible to intentionally influence the closing price and significantly lower the price. Such tradings are prohibited by regulators, but it is practically difficult to penalize them when

*This research was supported in part by the Research Institute for Mathematical Sciences, a Joint Usage/Research Center, Kyoto University in 2024 and the Grant-in-Aid for Scientific Research (No. 23K01460) of the Japan Society for the Promotion of Science.

†Address: Karasuma-higashi-iru, Imadegawa-dori, Kamigyo-ku, Kyoto, 602-8580 Japan
E-mail address: skuno@mail.doshisha.ac.jp

they are not obvious. When considering off-exchange trading venues, (1) institutional investors can manipulate prices in exchange tradings to get an advantage in OTC(off-exchange) tradings, and (2) broker-dealers can manipulate prices to get an advantage in OTC transactions when procuring a specific amount on the exchange after a contract is conducted, so it is necessary to set an appropriate price for both parties. It is therefore desirable to include in the contract both positive and negative incentives to discourage market manipulation.

In this paper, an institutional investor (client) who wishes to make a large purchase first demand a specific amount of volume on a trading venue set up by a dealer outside the exchange. When the dealer does not have the securities, he or she procures the securities on the exchange. Then within the framework of principal trading, we consider the price to be used in an alternative (off-exchange) trading venue in which the client minimizes its own costs, taking into account the dealer's profits.

The remainder of this paper is organized as follows. We set up a market model on the exchange and formulate principal trading, and also discuss the VWAP strategy in section 2. In Section 3, we show that the dealer's optimal strategy on the exchange is the VWAP in the framework of Baldauf et. al (2022), which does not consider permanent impact, then derive a second-best trading strategy for off-exchange trading venue, and show that the optimal execution price off-exchange is the VWAP on the exchange. Then, we consider the effect of permanent impact on off-exchange prices. Finally, we conclude this paper in Section 4.

2 Model setup

We consider a principal trading between a single client, who is a single institutional investor, and a single dealer, who is a securities firm. Before the opening of trading on the exchange, the client and dealer trade a predetermined amount (1 unit for simplicity) at a pre-determined price after the closing time of trading on the exchange. Then, we assume that the client, an institutional investor, can only purchase 1 unit through OTC trading. The price determined by the client may be the closing price[6] and [10], VWAP[2], TWAP, etc., of that day. The exact pricing is not yet known, but the price to be used has been decided in advance. In a similar situation, [11] is considering agency transactions using the closing price and taking commission into consideration. We assume that the dealer procures one unit during the trading time on the exchange on that day, $t = 1 \sim T$, while taking into account the cost of market impact at the exchange, before selling one unit to a client in the OTC trading.

2.1 Market model

Let the execution price of the dealer at the exchange at time t be p_t, and the trading volume (purchase volume) be q_t. Then, for the trading volume q_t we have

$$\sum_{t=1}^{T} q_t = 1. \tag{2.1}$$

In [2], they define the trading volume of noise traders in the exchange at time t as a deterministic value η_t. Then, let ν be the trading volume at time t between a dealer who procures a large amount of securities and a noise trader at the exchange, we have

$$\nu_t = q_t + \eta_t. \tag{2.2}$$

Then,

$$\sum_{t=1}^{T} \nu_t = \sum_{t=1}^{T} \eta_t + \sum_{t=1}^{T} q_t = \sum_{t=1}^{T} \eta_t + 1 \tag{2.3}$$

In addition, it is assumed that large volumes of execution on the exchange can be executed instantly with price impact. Therefore, the execution price model on the exchange is given by

$$p_t = p_0 + f\left(\frac{q_t}{\eta_t}\right) + G\left(\frac{\mathbf{q}}{\eta}\right) + \epsilon_t, \tag{2.4}$$

where f represents the temporary impact on the exchange, G represents the cumulative permanent impact to which it remains to some extent, and ϵ represents the public news effect, and $N \sim (\mu, \sigma_\epsilon^2)$ for all t. Also, it is satisfied

$$\mathbb{E}\left[(\epsilon_t - \epsilon_{t+1})|\eta, \epsilon_{t-1}\right] = 0. \tag{2.5}$$

The simplest price model, used in [1], [8], and [9], when temporary and permanent impacts are both linear is:

$$p_t = p_0 + \lambda\left(\frac{q_t}{\eta_t}\right) + \sum_{k=1}^{t-1}(1-\alpha)\lambda\left(\frac{q_k}{\eta_k}\right) + \epsilon_t, \tag{2.6}$$

where λ represents the price fluctuation per unit execution (its inverse is depth) and α represents the rate at which the temporary impact returns to its previous level.

2.2 Client problem

The risk-neutral client determines the off-exchange execution price, denoted by S, so that the risk-averse dealer accepts the client's offer (IR constraint) and maximizes the dealer's expected profit (IC constraint), maximizing the client's own expected profit $-\mathbb{E}[S(\mathbf{p}, \nu)]$. At that time, the dealer's strategy is also determined. That is,

$$\min_{S, \mathbf{q}} \quad \mathbb{E}[S(\mathbf{p}, \nu)] \tag{2.7a}$$

$$\text{s.t.} \quad \mathbb{E}[u(S(\mathbf{p}, \nu) - \mathbf{p} \cdot \mathbf{q})] \geq u(0), \tag{2.7b}$$

$$\mathbb{E}[u(S(\mathbf{p}, \nu) - \mathbf{p} \cdot \mathbf{q})] \geq \mathbb{E}[u(S(\mathbf{p}, \nu) - \mathbf{p} \cdot \bar{\mathbf{q}})], \forall \bar{\mathbf{q}}. \tag{2.7c}$$

Where, (2.7b) represents individual rationality constraints (participation constraints), and (2.7c) represents incentive compatibility constraints. In addition, in the constraints, $S(\mathbf{p}, \nu) - \mathbf{p} \cdot \mathbf{q}$ represents the dealer's payoff, and this problem is a second-best problem.

2.3 VWAP

VWAP stands for volume weighted average price, and when a large amount of stock is executed on the exchange, it is represented by

$$VWAP = \frac{\sum_{t=1}^{T} \nu_t p_t}{\sum_{t=1}^{T} \nu_t}. \tag{2.8}$$

When a dealer executes at an exchange with the VWAP target strategy, he or she will use the POV strategy represented by

$$q_t = \frac{\mu_t}{\sum_{t=1}^{T} \mu_t}. \tag{2.9}$$

In the next section, we show that under certain conditions, similar to [4], the target VWAP minimizes costs at the exchange.

3 No permanent impact model

In this section, based on [2], we introduce a model for calculating OTC (off-exchange) trading prices using a price model that does not take permanent impact into account. From the price model in (2.4) of the previous section, if we set $G = 0$ and $p_0 = 0$, then the execution price at time t is

$$p_t = f\left(\frac{q_t}{\eta_t}\right) + \epsilon_t. \tag{3.1}$$

3.1 First-best trading strategy

Consider the case where a client can observe the dealer's execution on the exchange and the dealer's knowledge of prices. In this case, the optimal strategy is to find a strategy in which equality holds in the IR constraint (2.7b), that is, a strategy satisfies $S(\mathbf{p}, \nu) = \mathbf{p} \cdot \mathbf{q}$. Therefore, the problem of the client's optimal trading strategy is

$$\min_{S, \mathbf{q}} \quad \mathbb{E}[S(\mathbf{p}, \nu)],$$

and trading strategy $\mathbf{q}$ is first-best if it satisfies

$$\underset{\mathbf{q}}{\text{argmin}} \mathbb{E}\left[\sum_{t=1}^{T} p_t q_t | \eta\right]. \tag{3.2}$$

Therefore, we seek an execution strategy $\mathbf{q}^* = \{q_1, q_2, \cdots, q_T\}$ that minimizes the following expectation,

$$\mathbb{E}\left[\sum_{t=1}^{T}\left(f\left(\frac{q_t}{\eta_t}\right) q_t + \epsilon_t q_t\right) |\eta\right] = \sum_{t=1}^{T} \mathbb{E}\left[f\left(\frac{q_t}{\eta_t}\right) q_t |\eta\right] + \sum_{t=1}^{T} \mathbb{E}\left[\epsilon_t q_t |\eta\right] \tag{3.3}$$

Here, since

$$\mathbb{E}[(\epsilon_t - \epsilon_T) q_t | \eta, \epsilon_{\mathbf{t-1}}] = \mathbb{E}[(\epsilon_t - \epsilon_{t+1} + \epsilon_{t+1} - \cdots + \epsilon_{T-1} - \epsilon_T) q_t | \eta, \epsilon_{\mathbf{t-1}}] = 0, \tag{3.4}$$

then

$$\begin{aligned}\sum_{t=1}^{T}\mathbb{E}\left[\epsilon_t q_t|\eta\right] &= \mathbb{E}\left[\epsilon_T|\eta\right]+\sum_{t=1}^{T}\mathbb{E}\left[\left(\epsilon_t-\epsilon_T\right)q_t|\eta\right]\\ &= \mu+\sum_{t=1}^{T-1}\mathbb{E}[\mathbb{E}[(\epsilon_t-\epsilon_T)q_t|\eta,\epsilon_{\mathbf{t-1}}]|\eta]\\ &= \mu. \end{aligned} \tag{3.5}$$

Secondly, since $\frac{q_t}{\eta_t}f(\frac{q_t}{\eta_t})$ is a convex function, by Jensen's inequality,

$$\begin{aligned}\sum_{t=1}^{T}\mathbb{E}\left[f\left(\frac{q_t}{\eta_t}\right)q_t|\eta\right] &= \mathbb{E}\left[\left(\sum_{s=1}^{T}\eta_s\right)\left(\frac{1}{\sum_{s=1}^{T}\eta_s}\right)\sum_{t=1}^{T}\frac{q_t}{\eta_t}f\left(\frac{q_t}{\eta_t}\right)\eta_t|\eta\right]\\ &\geq \mathbb{E}\left[\left(\sum_{s=1}^{T}\eta_s\right)\left(\frac{1}{\sum_{s=1}^{T}\eta_s}\sum_{t=1}^{T}\frac{q_t}{\eta_t}\eta_t\right)f\left(\frac{1}{\sum_{s=1}^{T}\eta_s}\sum_{t=1}^{T}\frac{q_t}{\eta_t}\eta_t\right)|\eta\right]\\ &= \mathbb{E}\left[\left(\sum_{s=1}^{T}\eta_s\right)\left(\frac{1}{\sum_{s=1}^{T}\eta_s}\sum_{t=1}^{T}q_t\right)f\left(\frac{1}{\sum_{s=1}^{T}\eta_s}\sum_{t=1}^{T}q_t\right)|\eta\right]. \end{aligned} \tag{3.6}$$

Therefore, using (2.1), we get

$$\mathbb{E}\left[\sum_{t=1}^{T}\left(f\left(\frac{q_t}{\eta_t}\right)q_t+\epsilon_t q_t\right)|\eta\right]=\sum_{t=1}^{T}\mathbb{E}\left[f\left(\frac{q_t}{\eta_t}\right)q_t|\eta\right]+\mu\geq f\left(\frac{1}{\sum_{s=1}^{T}\eta_s}\right)+\mu \tag{3.7}$$

The equality in (3.7) holds when

$$\frac{q_t}{\eta_t}=\frac{1}{\sum_{s=1}^{T}\eta_s}\Leftrightarrow q_t=\frac{\eta_t}{\sum_{s=1}^{T}\eta_s}, \tag{3.8}$$

and in this case $\mathbf{q}$ is optimal. This first-best strategy $q_t^{fb}=\frac{\eta_t}{\sum_{s=1}^{T}\eta_s}$ is equivalent to (2.9) and is a POV strategy. Using the market trading volume $\nu_t=q_t+\eta_t$, also this can be expressed as follows:

$$q_t^{fb}=\frac{\nu_t}{\sum_{s=1}^{T}\nu_s}. \tag{3.9}$$

3.2 Second-best trading strategy

Similarly, consider problem (2.7). With $\nu^{fb}=\eta+\mathbf{q}^{fb}$, the necessary and sufficient condition for S to be optimal is that both of the following are satisfied,

$$(1)\qquad S(\mathbf{p},\nu)=\mathbf{p}\cdot\mathbf{q}^{fb} \tag{3.10}$$

$$(2)\qquad \forall\bar{\mathbf{q}},\mathbb{E}[u(S(\mathbf{p},\nu^{fb})-\mathbf{p}\cdot\mathbf{q}^{fb}]\geq\mathbb{E}[u(S(\mathbf{p},\bar{\nu})-\mathbf{p}\cdot\bar{\mathbf{q}})]. \tag{3.11}$$

For more details, see [2].

Proposition 1 ([2])
The optimal off-exchange trading price S is the intraday VWAP (S^{VP}) on the exchange.

We confirm that the previous conditions are satisfied. Then,

$$
\begin{aligned}
\mathbb{E}\left[u\left(S^{VP}-\mathbf{p}\cdot\mathbf{q}\right)\right] &= \mathbb{E}\left[u\left(\sum_{t=1}^{T}\frac{p_t\nu_t}{\sum_{s=1}^{T}\nu_s}-\sum_{t=1}^{T}p_tq_t\right)\right] \\
&= \mathbb{E}\left[u\left(\sum_{t=1}^{T}\frac{\left(f\left(\frac{q_t}{\eta_t}\right)+\epsilon_t\right)\nu_t}{\sum_{s=1}^{T}\nu_s}-\sum_{t=1}^{T}\left(f\left(\frac{q_t}{\eta_t}\right)+\epsilon_t\right)q_t\right)\right]. \quad (3.12)
\end{aligned}
$$

Now, using induction, we can show that

$$
\sum_{t=1}^{T}f\left(\frac{q_t}{\eta_t}\right)\nu_t \leq \left(\sum_{t=1}^{T}\nu_t\right)\left(\sum_{t=1}^{T}f\left(\frac{q_t}{\eta_t}\right)q_t\right). \quad (3.13)
$$

Using this and from Jensen's inequality for the concave dealer's utility function,

$$
\begin{aligned}
\mathbb{E}\left[u\left(S^{VP}-\mathbf{p}\cdot\mathbf{q}\right)\right] &= \mathbb{E}\left[u\left(\sum_{t=1}^{T}\frac{\left(f\left(\frac{q_t}{\eta_t}\right)+\epsilon_t\right)\nu_t}{\sum_{s=1}^{T}\nu_s}-\sum_{t=1}^{T}\left(f\left(\frac{q_t}{\eta_t}\right)+\epsilon_t\right)q_t\right)\right] \\
&\leq \mathbb{E}\left[u\left(\frac{\left(\sum_{t=1}^{T}\nu_t\right)\left(\sum_{t=1}^{T}f\left(\frac{q_t}{\eta_t}\right)q_t\right)}{\sum_{s=1}^{T}\nu_s}+\sum_{t=1}^{T}\frac{\epsilon_t\nu_t}{\sum_{s=1}^{T}\nu_s}-\sum_{t=1}^{T}\left(f\left(\frac{q_t}{\eta_t}\right)+\epsilon_t\right)q_t\right)\right] \\
&= \mathbb{E}\left[u\left(\sum_{t=1}^{T}\frac{\epsilon_t\nu_t}{\sum_{s=1}^{T}\nu_s}-\sum_{t=1}^{T}q_t\epsilon_t\right)\right] \\
&\leq u\left(\mathbb{E}\left[\sum_{t=1}^{T}\frac{\epsilon_t\nu_t}{\sum_{s=1}^{T}\nu_s}-\sum_{t=1}^{T}q_t\epsilon_t\right]\right). \quad (3.14)
\end{aligned}
$$

Here, from equation(3.5), and

$$
\begin{aligned}
\mathbb{E}\left[\sum_{t=1}^{T}\frac{\epsilon_t\nu_t}{\sum_{s=1}^{T}\nu_s}\right] &= \mathbb{E}\left[\mathbb{E}\left[\sum_{t=1}^{T}\frac{\epsilon_t\nu_t}{\sum_{s=1}^{T}\nu_s}|\eta\right]\right] \\
&= \mathbb{E}\left[\frac{\sum_{t=1}^{T}\mathbb{E}[\epsilon_tq_t|\eta]+\sum_{t=1}^{T}\mathbb{E}[\epsilon_t\eta_t|\eta]}{\sum_{s=1}^{T}(q_s+\eta_s)}\right] \\
&= \mathbb{E}\left[\frac{\mu+\mu\sum_{t=1}^{T}\eta_t}{1+\sum_{s=1}^{T}\eta_s}\right] \\
&= \mu, \quad (3.15)
\end{aligned}
$$

then, we have,

$$
\mathbb{E}\left[u\left(S^{VP}-\mathbf{p}\cdot\mathbf{q}\right)\right] \leq u\left(\mathbb{E}\left[\sum_{t=1}^{T}\frac{\epsilon_t\nu_t}{\sum_{s=1}^{T}\nu_s}-\sum_{t=1}^{T}q_t\epsilon_t\right]\right) = u(0). \quad (3.16)
$$

In other words, to maximize this expected payoff, $S^{VP} = \mathbf{p}\cdot\mathbf{q}$ is satisfied. Also, from the previous condition (1), it must be equal to $\mathbf{p}\cdot\mathbf{q}^{fb}$ for $S(\mathbf{p},\nu)$ to be optimal,

$$
\begin{aligned}
S^{VP} &= \mathbf{p}\cdot\mathbf{q}^{fb} \\
&= \sum_{t=1}^{T}\frac{p_t\nu_t}{\sum_{s=1}^{T}\nu_s}. \quad (3.17)
\end{aligned}
$$

We can also see that $(S^{VP}, \mathbf{q}^{fb})$ satisfies the IC constraint. Additionally, we find that even if there is drift in the public news effect, this does not affect strategies on the exchange (first-best trading strategies).

3.3 Permanent impact effects

In this subsection, we consider the effect of off-market trading prices when taking into account the permanent impact on the exchange, which we have set to 0. Consider the following price model:

$$p_t = f\left(\frac{q_t}{\eta_t}\right) + G\left(\frac{q_t}{\eta_t}\right) + \epsilon_t. \tag{3.18}$$

For simplicity, we assume a simple price model in which both the temporary and permanent impacts are linear, as shown in equation (2.6) . Then, up until now, we have set $\alpha = 1$, but from now on we will consider $0 \leq \alpha < 1$. In this case, as shown in [1] and [9], the optimal procurement strategy for the dealer at the exchange is a monotonically decreasing strategy that executes more at the initial stage. This is because the effects of execution are cumulative, so it is less costly to execute more early stage. At this time, the off-exchange contract price may be affected by drift due to public news on the exchange, so it is generally higher than when $\alpha = 1$. On the other hand, if we use the model of [3], which assumes that price fluctuations due to large volume executions cannot be reversed (therefore $\alpha = 0$), it is shown that equal volume execution is optimal. In particular, if the trading time interval is constant, this strategy is a TWAP strategy. Then, as $S = S(TWAP) = S^{TP}$,

$$S^{TP} = \mathbf{p} \cdot \mathbf{q}^{fb} = \sum_{t=1}^{T} \frac{p_t}{T}. \tag{3.19}$$

In this case, it can be seen that the off-exchange trading price will be relatively high.

4 Conclusion

In a simple pricing model that does not consider permanent impact, it can be shown that the optimal off-exchange contract price determined in advance for the principal trading is the VWAP. This means that the optimal execution strategy for a dealer to procure at the exchange at the lowest cost is the VWAP, and the solution to the first-best problem reduces to an optimal execution problem. When permanent impact is taken into account, additional costs are incurred when procuring, so the optimal off-exchange trading price is higher than when only temporary impact is considered. Generally, this additional amount is added as an off-exchange fee, but the derivation of an analytical solution for this is left as a future research. Furthermore, while we treated the noise trader's volume η_t for all t as deterministic, we also leave it as a future work to treat it as stochastic.

References

[1] Almgren, R. and Chriss, N. (2000), "Optimal Execution of Portfolio Transactions," *Journal of Risk*, **3**, 5–39.

[2] Baldauf, M., Frei, C., and Mollner, J. (2022), "Principal Trading Arrangements: When Are Common Contracts Optimal," *Management Science*, **68**, 4, 3112–3128.

[3] Bertsimas, D. and Lo, A. W. (1998), "Optimal Control of Execution Costs," *Journal of Financial Markets*, **1**, 1–50.

[4] Cartea, Á. and Jaimungal, S. (2016), "A Closed-Form Execution Strategy to Target Volume Weighted Average Price," *SIAM Journal on Financial Mathematics*, **7**, 760–785.

[5] Frei, C. and Westray, N. (2015), "Optimal Execution of a VWAP Order: A Stochastic Control Approach," *Mathematical Finance*, **25**, 3, 612–639.

[6] Frei, C. and Mitra, J. (2021), "Optimal Closing Benchmarks," *Finance Research Letters*, **40**, 101674.

[7] Harris, L. (1989), "A Day-End Transaction Price Anomaly," *Journal of Financial and Quantitative Analysis*, **24**, 1, 29–45.

[8] Huberman, G. and Stanzl, W.(2005), "Optimal Liquidity Trading," *Review of Finance*, **9**, 165–200.

[9] Kuno, S. and Ohnishi, M. (2015), "Optimal Execution in Illiquid Market under the Absence of Price Manipulation," *Journal of Mathematical Finance*, **15**, 1, 1–15.

[10] Kuno, S., Ohnishi, M., and Shimizu, P. (2017), "Optimal Off-Exchange Execution with Closing Price," *Journal of Mathematical Finance*, **17**, 1, 54–64.

[11] Kuno, S.(2019), "Pricing of Guaranteed Order Execution Contracts," *Research Institute for Mathematical Sciences (RIMS) Kôkyûroku*, **2111**, 114–124.

[12] Mitchell, D., Bialkowski, J. P., and Tompaidis, S. (2013), "Optimal VWAP Tracking," SSRN, http://ssrn.com/abstract=2333916.