Continuous-time optimal pair-trade execution with cross-impacts and common risk factors[1]

Masamitsu OHNISHI[2]
Faculty of Informatics, Yamato University
Center for Mathematical Modeling and Data Science, Osaka University
Makoto SHIMOSHIMIZU[3]
Department of Industrial and Systems Engineering, Tokyo University of Science

Abstract

We show how to incorporate cross-impacts and common risk factors into an optimal trade execution strategy in a transient market impact model. Our framework, based on Fukasawa et al. (2025), enables us to take into account some common risk factors, the number of which can differ from that of financial assets that a large trader executes. The optimal trade execution strategy for a risk-averse large trader becomes a time-dependent affine function of the state variables. This result implys that the cross-impacts of buying/selling asset i on asset j affects the execution volume of both assets i and j. Additionally, the time-dependent coefficients are derived from a solution to a system of ordinary differential equations (ODEs) with terminal conditions.

Keywords: Optimal execution; Cross-impacts; Common risk factors; HJB equation
JEL Classification: C73; C02; C61; G11
AMS subject classification: 49J15; 93E20; 91G80

1 Introduction

A considerable number of empirical studies conducted show that the so-called *cross–impacts* of multiple assets is important when constructing an optimal execution strategy (e.g., Benzaquen et al. [2]). Several factors for the cross–impact are involved: correlation of different assets' returns, the commonality of liquidity across assets, the possibility of a (large) trade in an asset to adjust the bid–ask spread by a market maker. In addition, Schneider and Lillo [33] shows that a price manipulation may occur unless several conditions hold for the decay kernel that represents a transient price impact of a large trader. Along with the empirical facts, a large number of theoretical studies analyze an optimal execution problems of multiple assets for large traders (e.g., Gârleanu and Pedersen [17, 18]; Cartea et al. [7]; Ma et al. [25]; Tsoukalas et al. [34]; Bergault et al. [5]; Ohnishi and Shimoshimizu [30]).

This paper addresses a continuous-time optimal pair–trade execution problem for a single large trader. We focus on the following situations which the institutional trader (or large trader) may face in a real marketplace. Institutional traders manage their trading with multiple assets correlated with each other to mitigate the price risk. The market model we describe in Section 2 includes the effect of temporary, permanent, and transient price impacts caused by a large trader. For the detail on optimal execution problems, see, e.g., Cartea et al. [8], Guéant [19], Laruelle and Lehalle [23], and Shimoshimizu [32].

The 'pair–trading' has already been discussed in much empirical literature, clarifying the importance of trading co-integrated multiple assets (e.g., in Gatev et al. [15]). The pair-trading

[1]This research was supported in part by the Research Institute for Mathematical Sciences, a Joint Usage/Research Center, Kyoto University in 2025.
[2]Address: Suita, Osaka, 564–0082, Japan
E-mail address: onishi.masamitsu@yamato-u.ac.jp
[3]Address: Noda, Chiba, 278–0022, Japan
E-mail address: shimosshi1q94@gmail.com

is a trading method that makes a profit by taking advantage of changes in the difference between two financial instruments that are highly correlated. In the case of stocks, stock prices in the same industry tend to move in a similar manner, so two stocks with similar movements are held, with the undervalued stock 'bought' and the overvalued stock 'shorted.' This method is less susceptible to changes in the market. The situation we deal with in this paper differs from what is called the 'pair–trading' having been used in a real trading market. Our model deals with an execution problem of a large volume of trading with multiple risky assets in a short term (for example, an intraday trading). An example of the situation we assume here is as follows: An institutional investor managing, e.g., an index fund, replaces their stock holdings in line with the periodic replacement of stocks (e.g., in the S & P 500 or the Nikkei 225). The stocks that make up stock price indexes such as the S & P 500 and the Nikkei 225 may change according to some criteria, the judgment of the index constructors, or the actions of related listed companies. In such cases, institutional investors who manage index funds that track those particular stock price indexes must replace the stocks that make up the funds to keep up with those indexes. This is a typical case of pair-trade execution, in which one stock is sold and another is purchased.

Notation

$\mathbb{R}_+$ ($\mathbb{R}_{++}$, respectively) denotes the set of non-negative (strictly positive) real numbers. We represent by $\mathbb{C}$ the set of complex numbers. Let $\mathbb{K}$ be either $\mathbb{R}$ or $\mathbb{C}$. For any positive integers $n, m \in \{1, 2, \ldots\}$, we use the notation $\mathbb{K}^n$ to describe the set of all n-dimensional $\mathbb{K}$-valued column vectors and $\mathcal{M}^{n,m}(\mathbb{K})$ to describe the set of all $n \times m$ $\mathbb{K}$-valued matrices. In particular, $\mathcal{M}^n(\mathbb{K}) := \mathcal{M}^{n,n}(\mathbb{K})$ is the set of all n-order $\mathbb{K}$-valued square matrices. The set of n-order $\mathbb{K}$-valued symmetric matrices is denoted by $\mathcal{S}^n(\mathbb{K})$. We use the notation $\mathcal{S}^n_{++}(\mathbb{K})$ ($\mathcal{S}^n_{+}(\mathbb{K})$, respectively) to denote the set of n-order $\mathbb{K}$-valued positive definite (positive semidefinite) matrices. $\mathbf{I}_n \in \mathcal{S}^n_{++}(\mathbb{K})$ denotes the n-order identity matrix. For an $n \times m$ $\mathbb{K}$-valued matrix or vector $\mathbf{A}$, we let $\mathbf{A}^\top$ stand for the transpose of the matrix or vector. In addition, for any n-order square matrix $\mathbf{A} \in \mathcal{M}^n(\mathbb{R})$, $\mathrm{tr}(\mathbf{A})$ is the trace of $\mathbf{A}$. Also, for any positive integers $n, m, l, k \in \{1, 2, \ldots\}$ and two matrices $\mathbf{A} \in \mathcal{M}^{n,m}(\mathbb{K})$, $\mathbf{B} \in \mathcal{M}^{l,k}(\mathbb{K})$, $\mathbf{A} \otimes \mathbf{B} \in \mathcal{M}^{nl,mk}(\mathbb{K})$ represents the Kronecker product of the two matrices. Finally, for any $\mathbf{A}_1, \mathbf{A}_2 \in \mathcal{S}^n(\mathbb{R})$, we write $\mathbf{A}_1 \geq \mathbf{A}_2$ if and only if $\mathbf{A}_1 - \mathbf{A}_2 \in \mathcal{S}^n_+$. This inequality defines the natural order on symmetric matrices.

2 Market Model

We consider an execution problem of multiple financial assets, indexed by $i \in \{1, \ldots, n\}$. In a financial market, a risk-averse large trader must purchase $\mathfrak{Q}^i$ ($\in \mathbb{R}$) volume of risky asset $i \in \{1, \ldots, n\}$ in a time window $[0, T]$. Let Q^i_t ($\in \mathbb{R}$) for $i \in \{1, \ldots, n\}$ be the cumulative purchase up to time $t \in [0, T]$ of the large trader. Then, the number of shares that remained to purchase at time $t \in [0, T]$ is given by

$$\overline{Q}^i_t = \mathfrak{Q}^i - Q^i_t, \tag{1}$$

with the initial conditions that $\overline{Q}^i_0 = \mathfrak{Q}^i$. We consider a continuous trading strategy:

$$\mathrm{d}Q^i_t = \dot{Q}^i_t \mathrm{d}t. \tag{2}$$

It is assumed that Q^i_t is continuously differentiable in time $t \in [0, T]$ for all $i \in \{1, \ldots, n\}$. We denote by the positive and negative $\dot{Q}^i_t$ the acquisition and liquidation of risky asset $i \in \{1, \ldots, n\}$, respectively. This leads to a similar setup for a selling problem. We express the stacked vector of remaining execution volumes, cumulative purchased volumes up to time $t \in [0,T]$, and the trading speeds for all assets as follows:

$$\overline{\boldsymbol{Q}}_t \coloneqq \begin{pmatrix} \overline{Q}_t^1 \\ \vdots \\ \overline{Q}_t^n \end{pmatrix} \in \mathbb{R}^n; \quad \boldsymbol{Q}_t \coloneqq \begin{pmatrix} Q_t^1 \\ \vdots \\ Q_t^n \end{pmatrix} \in \mathbb{R}^n; \quad \dot{\boldsymbol{Q}}_t \coloneqq \begin{pmatrix} \dot{Q}_t^1 \\ \vdots \\ \dot{Q}_t^n \end{pmatrix} \in \mathbb{R}^n. \tag{3}$$

By Eqs. (1) and (2), the relationships that $\overline{\boldsymbol{Q}}_t = \boldsymbol{\mathfrak{Q}} - \boldsymbol{Q}_t$ and $\mathrm{d}\boldsymbol{Q}_t = \dot{\boldsymbol{Q}}_t \mathrm{d}t$ hold where $\boldsymbol{\mathfrak{Q}} = (\mathfrak{Q}^1, \dots, \mathfrak{Q}^n)^\top$.

The market price (or quoted price) of the risky asset $i \in \{1, \dots, n\}$ is denoted by P_t^i. The execution price of the asset $\widehat{P}$ is then assumed to follow a linear market impact model (e.g., Bertsimas and Lo [3]; Almgren and Chriss [1]; Gatheral et al. [16]; Obizhaeva and Wang [27]; Kuno and Ohnishi [21]; Cartea and Jaimungal [?, ?]; Kuno et al. [22]; Lehalle and Neuman [24]; Ohnishi and Shimoshimizu [28, 29, 30, 31]; Fukasawa et al. [12, 13]):

$$\widehat{\boldsymbol{P}}_t = \boldsymbol{P}_t + \boldsymbol{\Lambda}_t \dot{\boldsymbol{Q}}_t, \tag{4}$$

where P_t $(\in \mathbb{R})$ represents the stacked vector of risky asset prices at time $t \in [0,T]$, and

$$\boldsymbol{\Lambda}_t \coloneqq \begin{pmatrix} \lambda_t^{11} & \cdots & \lambda_t^{1n} \\ \vdots & \ddots & \vdots \\ \lambda_t^{n1} & \cdots & \lambda_t^{nn} \end{pmatrix} \in \mathcal{M}^n(\mathbb{R}), \tag{5}$$

is the matrix of market impact coefficients at time $t \in [0,T]$. The off-diagonal element λ_t^{ij} for $i, j \in \{1, \dots, n\}$ $(i \neq j)$ represents the *cross-impacts*: A large trader's liquidation (either buy or sell) of risky asset j impacts the execution price of risky asset i. In the rest of this paper, we call the matrix $\boldsymbol{\Lambda}_t$ as *cross-impact matrix* at time $t \in [0,T]$. For the cross-impact matrix, we impose the following assumption:

Assumption 2.1. The cross-impact matrix (5) is *symmetric*.

This assumption stems from the results in Schneider and Lillo [33] that the asymmetric cross-impact may lead to (dynamic) arbitrage.

In addition to Assumption 2.1, we assume the following:

Assumption 2.2 (Definiteness of cross-impact matrix)**.** The cross-impact matrix (5) is *positive definite*.

In the sequel of this paper, we assume that the buy- and sell-trade of the large trader induce the same (instantaneous) market impact, although it would be different in the real market. We can, however, justify this assumption from the statistical analysis of market data shown by, for instance, Cartea and Jaimungal [?, ?]. By Eq. (4), the large trader's wealth process at time $t \in [0,T]$, denoted by W_t, evolves as

$$\mathrm{d}W_t = -\widehat{\boldsymbol{P}}_t^\top \mathrm{d}\boldsymbol{Q}_t = -\widehat{\boldsymbol{P}}_t^\top \dot{\boldsymbol{Q}}_t \mathrm{d}t = -\left(\boldsymbol{P}_t + \boldsymbol{\Lambda}_t \dot{\boldsymbol{Q}}_t\right)^\top \dot{\boldsymbol{Q}}_t \mathrm{d}t. \tag{6}$$

Besides the above (instantaneous) market impact, we consider the permanent and transient parts of the market impact. The residual effect of past market impacts is defined as

$$\boldsymbol{R}_t = \boldsymbol{\mathfrak{G}}(t)\boldsymbol{R}_0 + \int_0^t \boldsymbol{\mathfrak{G}}(s)\boldsymbol{\alpha}_s \boldsymbol{\Lambda}_s \dot{\boldsymbol{Q}}_s \mathrm{d}s, \tag{7}$$

where $\mathfrak{G}\colon [0,T] \to \mathcal{M}^n(\mathbb{R})$ is the exponential decay kernel matrix defining the transient market impact and is defined as

$$\mathfrak{G}(t) := \mathrm{e}^{-\boldsymbol{\Upsilon} t}, \tag{8}$$

through some matrix $\boldsymbol{\Upsilon} \in \mathcal{M}^n(\mathbb{R}_+)$, and

$$\boldsymbol{\alpha}_t := \begin{pmatrix} \alpha_t^{11} & \cdots & \alpha_t^{1n} \\ \vdots & \ddots & \vdots \\ \alpha_t^{n1} & \cdots & \alpha_t^{nn} \end{pmatrix} \in \mathcal{M}^n(\mathbb{R}), \tag{9}$$

is the deterministic coefficient of linear temporary market impacts. In the differential form, we can rewrite Eq. (8) as

$$\mathrm{d}\boldsymbol{R}_t = -\boldsymbol{\Upsilon}\boldsymbol{R}_t \mathrm{d}t + \boldsymbol{\alpha}_t \boldsymbol{\Lambda}_t \dot{\boldsymbol{Q}}_t \mathrm{d}t. \tag{10}$$

Note that for the derivation of Eq. (10), we have used the fact that

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathfrak{G}(t) = -\boldsymbol{\Upsilon}\mathfrak{G}(t) = -\mathfrak{G}(t)\boldsymbol{\Upsilon}. \tag{11}$$

In this model, $\boldsymbol{\Upsilon} \in \mathcal{M}^n(\mathbb{R}_+)$ characterizes the deterministic resilience speed.[4] This residual effect indicates that the market impact decays gradually over the trading window $[0,T]$. R_0 is assumed to be zero in the following analysis. The assumption is quite plausible from the fact that the trader has no market impact on any risky asset before liquidating/acquiring the risky asset, and thereby there exist no residual effects caused by the trader on the price before the execution. Note that $\boldsymbol{\alpha}_t \boldsymbol{\Lambda}_t \dot{\boldsymbol{Q}}_t$ represents the temporary market impact caused by the large trader.

Remark 2.1. For example, we can consider

$$\mathfrak{G}(t) := \exp\left\{ t \begin{pmatrix} \rho_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \rho_n \end{pmatrix} \right\} = \begin{pmatrix} \mathrm{e}^{\rho_1 t} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \mathrm{e}^{\rho_n t} \end{pmatrix} \in \mathcal{M}^n(\mathbb{R}_+), \tag{12}$$

as Tsoukalas et al. (2019). In addition, for any $\mathbf{A} \in \mathcal{M}^n(\mathbb{R})$ with the following decomposition:

$$\mathbf{A} = \mathfrak{P}\boldsymbol{\mathcal{J}}\mathfrak{P}^{-1}, \tag{13}$$

where $\boldsymbol{\mathcal{J}}$ is the Jordan matrix and $\mathfrak{P}$ is a nonsingular matrix, we can consider

$$\mathfrak{G}(t) := \mathrm{e}^{-\mathbf{A}t} = \mathfrak{P} \exp\{-\boldsymbol{\mathcal{J}} t\} \mathfrak{P}, \tag{14}$$

Remark 2.2. For any matrix $\mathbf{A} \in \mathcal{M}^n(\mathbb{R})$, the exponential matrix function $\mathrm{e}^{-\mathbf{A}t}$ is nonsingular since we have $|\mathrm{e}^{-\mathbf{A}t}| = \mathrm{e}^{\mathrm{tr}(\mathbf{A}t)} > 0$.

[4] Much of theoretical analysis, such as in Obizhaeva and Wang [27], Tsoukalas et al. [34], deal with a (deterministic and) time-independent resilience speed. Many empirical kinds of research, however, demonstrate that liquidity is variable over time, suggesting that the resilience speed is time-dependent. Our analysis allows the time-dependence for the resilience speed, i.e., ρ_t for all $t \in [0,T]$, as considered in Fruth et al. [11]. Notwithstanding a meaningful extension from the viewpoint of real market analysis, we henceforth formulate the model without a time-dependent parameter (i.e., with ρ) since the dependence will not offer additional intriguing results in the following analysis.

The market price is assumed to consist of the sum of the following two components:

$$\boldsymbol{P}_t = \boldsymbol{P}_t^f + \boldsymbol{R}_t, \tag{15}$$

where

$$\boldsymbol{P}_t^f := \begin{pmatrix} P^{1,f} \\ \vdots \\ P^{n,f} \end{pmatrix} \in \mathbb{R}^n, \tag{16}$$

for $t \in [0, T]$ stands for what we call the *fundamental price* of the risky asset. This assumption stems from the definition of the transient market impact. The transient market impact is the discounted sum of the past temporary market impact. Thus, the transient market impact can be deemed to not influence the fundamental part of the market price. We define the dynamics of the fundamental price as follows:

$$\mathrm{d}\boldsymbol{P}_t^f = \left(\boldsymbol{\beta}_t \boldsymbol{\Lambda}_t \dot{\boldsymbol{Q}}_t + \boldsymbol{\kappa}_t \boldsymbol{\mathcal{I}}_t\right) \mathrm{d}t + \mathrm{d}\boldsymbol{Z}_t. \tag{17}$$

The first term in Eq. (17),

$$\boldsymbol{\beta}_t \boldsymbol{\Lambda}_t \dot{\boldsymbol{Q}}_t \in \mathbb{R}^n, \tag{18}$$

represents the permanent market impact (with cross-impacts) caused by the large trader. $\boldsymbol{\beta}$ denotes the matrix of coefficients and is given by

$$\boldsymbol{\beta}_t := \begin{pmatrix} \beta_t^{11} & \cdots & \beta_t^{1n} \\ \vdots & \ddots & \vdots \\ \beta_t^{n1} & \cdots & \beta_t^{nn} \end{pmatrix} \in \mathcal{M}^n(\mathbb{R}). \tag{19}$$

Contrary to the transient market impact, the permanent market impact influences the market price by definition, which in turn suggests that the permanent market impact directly affects the fundamental part of the market price. $\boldsymbol{Z}_t \in \mathbb{R}^n$ stands for the effect of some public news/information about the economic situation which may affect the market price (or quoted price). Adding to these two factors, we assume that what we define as the *common risk factors*, denoted by $\boldsymbol{\mathcal{I}}_t \in \mathbb{R}^m$, affects the fundamental price of the risky asset. $\boldsymbol{\kappa} \in \mathcal{M}^{n,m}(\mathbb{R})$ is the matrix of the coefficients of the common risk factors and si given by

$$\boldsymbol{\kappa}_t := \begin{pmatrix} \kappa_t^{11} & \cdots & \kappa_t^{1n} \\ \vdots & \ddots & \vdots \\ \kappa_t^{n1} & \cdots & \kappa_t^{nn} \end{pmatrix} \in \mathcal{M}^n(\mathbb{R}_+). \tag{20}$$

This notion is motivated by the so-called factor models of, e.g., Fama and French **??**, Carhart **??**, and Fama and French **??**, to mention only a few.

The more mathematically formal setting of the above model is as follows. Let $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{0 \le t \le T}, \mathbb{P})$ be a filtered probability space. The processes of the Markovian environment $\boldsymbol{\mathcal{I}}_t$ and the effect $\boldsymbol{Z}_t$ caused by public news or information on the (quoted) price are defined on the space as follows. The dynamics of the effect caused by public news or information on the (quoted) price, $\{\boldsymbol{Z}_t\}_{0 \le t \le T}$, is given by

$$\mathrm{d}\boldsymbol{Z}_t = \boldsymbol{\mu}^{\boldsymbol{Z}} \mathrm{d}t + \boldsymbol{\Sigma}^{\boldsymbol{Z}} \mathrm{d}\boldsymbol{B}_t^{\boldsymbol{Z}}, \tag{21}$$

where

$$\boldsymbol{\mu}^{\boldsymbol{Z}} := \begin{pmatrix} \mu^{Z,1} \\ \vdots \\ \mu^{Z,n} \end{pmatrix} \in \mathbb{R}^n; \tag{22}$$

$$\boldsymbol{\Sigma}^{\boldsymbol{Z}} := \begin{pmatrix} \sigma^{Z,11} & \dots & \sigma^{Z,1n} \\ \vdots & \ddots & \vdots \\ \sigma^{Z,n1} & \dots & \sigma^{Z,nn} \end{pmatrix} \in \mathcal{M}^n(\mathbb{R}_+), \tag{23}$$

and $\mathcal{B}^{\boldsymbol{Z}} := \{\boldsymbol{B}_t^{\boldsymbol{Z}}\}_{0\le t\le T}$ stands for n-dimensional Brownian motions with $B_0^{\boldsymbol{Z}} = \mathbf{0}$ a.s. The dynamics of common risk factors, $\{\boldsymbol{\mathcal{I}}_t\}_{0\le t\le T}$, is given by

$$\mathrm{d}\boldsymbol{\mathcal{I}}_t = \left(\boldsymbol{\mathfrak{J}}_t^{\boldsymbol{\mathcal{I}}} - \boldsymbol{\mathfrak{K}}_t^{\boldsymbol{\mathcal{I}}}\boldsymbol{\mathcal{I}}_t\right)\mathrm{d}t + \boldsymbol{\Sigma}^{\boldsymbol{\mathcal{I}}}\mathrm{d}\boldsymbol{B}_t^{\boldsymbol{\mathcal{I}}}, \tag{24}$$

where

$$\boldsymbol{\mathfrak{J}}^{\boldsymbol{\mathcal{I}}} := \begin{pmatrix} \mathfrak{J}^{\boldsymbol{\mathcal{I}},1} \\ \vdots \\ \mathfrak{J}^{\boldsymbol{\mathcal{I}},m} \end{pmatrix} \in \mathbb{R}^m; \tag{25}$$

$$\boldsymbol{\mathfrak{K}}^{\boldsymbol{\mathcal{I}}} := \begin{pmatrix} \mathfrak{K}^{\boldsymbol{\mathcal{I}},11} & \dots & \mathfrak{K}^{\boldsymbol{\mathcal{I}},1m} \\ \vdots & \ddots & \vdots \\ \mathfrak{K}^{\boldsymbol{\mathcal{I}},m1} & \dots & \mathfrak{K}^{\boldsymbol{\mathcal{I}},mm} \end{pmatrix} \in \mathcal{M}^m(\mathbb{R}); \tag{26}$$

$$\boldsymbol{\Sigma}^{\boldsymbol{\mathcal{I}}} := \begin{pmatrix} \sigma^{\boldsymbol{\mathcal{I}},11} & \dots & \sigma^{\boldsymbol{\mathcal{I}},1m} \\ \vdots & \ddots & \vdots \\ \sigma^{\boldsymbol{\mathcal{I}},m1} & \dots & \sigma^{\boldsymbol{\mathcal{I}},mm} \end{pmatrix} \in \mathcal{M}^m(\mathbb{R}_+), \tag{27}$$

and $\mathcal{B}^{\boldsymbol{\mathcal{I}}} := \{\boldsymbol{B}_t^{\boldsymbol{\mathcal{I}}}\}_{0\le t\le T}$ stands for m-dimensional Brownian motions with $\boldsymbol{B}_0^{\boldsymbol{\mathcal{I}}} = \mathbf{0}$ a.s., and $\boldsymbol{\mathcal{I}}_0 = \mathbf{0}$. Eq. (24) is a *generalized multi-dimensional Ornstein–Uhlenbeck (OU) process.*

For notational simplicity, we also assume that the filtration $\{\mathcal{F}_t\}_{0\le t\le T}$ is the natural filtration generated by $(\mathcal{B}^{\boldsymbol{\mathcal{I}}}, \mathcal{B}^{\boldsymbol{Z}})$, that is,

$$\mathcal{F}_t := \sigma\left\{(\boldsymbol{B}_s^{\boldsymbol{\mathcal{I}}}, \boldsymbol{B}_s^{\boldsymbol{Z}}), 0 \le s \le t\right\}, \tag{28}$$

and satisfies the usual conditions. We assume that the two Brownian motions are independent.[5]

If we assume that the information flow accessible for the large trader is carried by the filtration $\{\mathcal{F}_t\}_{0\le t\le T}$, then the executed volume Q_t of the large trader by time $t \in [0,T]$ is an $\mathcal{F}_t$-measurable (real-valued) random variable. Thus, the set of admissible execution strategies is defined as follows:

$$\mathcal{A} := \Big\{\{\boldsymbol{Q}_t\}_{0\le t\le T}\Big|\{\mathcal{F}_t\}_{0\le t\le T}\text{–adapted process with a continuously differentiable path,}$$
$$\boldsymbol{Q}_0 = \mathbf{0}\Big\}. \tag{30}$$

According to the dynamics of the market model, it turns out that the wealth process, price dynamics, remaining execution volume, and residual effect depend on the process of the cumulative

[5]The quadratic co-variation of $\boldsymbol{B}_t^{\boldsymbol{\mathcal{I}}}$ and $\boldsymbol{B}_t^{\boldsymbol{Z}}$ takes the following form:

$$\mathrm{d}\langle \boldsymbol{B}^{\boldsymbol{\mathcal{I}}}, \boldsymbol{B}^{\boldsymbol{Z}}\rangle_t = \rho^{\boldsymbol{\mathcal{I}},\boldsymbol{Z}}\mathrm{d}t. \tag{29}$$

This equation implies that these two processes are allowed to be correlated with each other.

purchase (or liquidation) denoted by $\mathcal{Q} = \{\boldsymbol{Q}_s\}_{0 \le s \le t}$:

$$
\begin{aligned}
\mathrm{d}W_t^{\mathcal{Q}} &= -(\widehat{\boldsymbol{P}}_t^{\boldsymbol{Q}})^\top \mathrm{d}\overline{\boldsymbol{Q}}_t = -(\widehat{\boldsymbol{P}}_t^{Q})^\top \dot{\boldsymbol{Q}}_t \mathrm{d}t = -\left(\boldsymbol{P}_t^{\boldsymbol{Q}} + \boldsymbol{\Lambda}_t \dot{\boldsymbol{Q}}_t\right)^\top \dot{\boldsymbol{Q}}_t \mathrm{d}t;\\
\mathrm{d}\boldsymbol{P}_t^{\mathcal{Q}} &= \boldsymbol{\beta}_t \boldsymbol{\Lambda}_t \dot{\boldsymbol{Q}}_t \mathrm{d}t + \boldsymbol{\kappa}_t \boldsymbol{\mathcal{I}}_t \mathrm{d}t + \mathrm{d}\boldsymbol{Z}_t + \mathrm{d}\boldsymbol{R}_t^{Q};\\
\mathrm{d}\overline{\boldsymbol{Q}}_t^{\mathcal{Q}} &= -\mathrm{d}\boldsymbol{Q}_t = -\dot{\boldsymbol{Q}}_t \mathrm{d}t;\\
\mathrm{d}\boldsymbol{R}_t^{\mathcal{Q}} &= -\rho \boldsymbol{R}_t^{\mathbf{Q}} \mathrm{d}t + \boldsymbol{\alpha}_t \boldsymbol{\Lambda}_t \dot{\boldsymbol{Q}}_t \mathrm{d}t.
\end{aligned}
$$

However, to simplify the notations, we suppress the superscript $\mathcal{Q}$ in the above expressions representing the dependence on $\mathcal{Q}$ to each state variable, and simply use the ones $(W_t, \boldsymbol{P}_t, \overline{\boldsymbol{Q}}_t, \boldsymbol{R}_t)$ defined in the previous description except the cases when we should emphasize the dependence explicitly.

3 Performance Criteria and HJB Equation

3.1 Performance Criteria of Large Trader

The state of the process at time $t \in [0, T]$, denoted by $\boldsymbol{s}_t$, is a 5-tuple and is defined as

$$
\boldsymbol{s}_t := \begin{pmatrix} W_t \\ \boldsymbol{P}_t \\ \overline{\boldsymbol{Q}}_t \\ \boldsymbol{R}_t \\ \boldsymbol{\mathcal{I}}_t \end{pmatrix} \in \mathbb{R} \times \mathbb{R}^n \times \mathbb{R}^n \times \mathbb{R}^n \times \mathbb{R}^m = \mathbb{R}^{1+3n+m} =: \mathbb{S}. \tag{31}
$$

As we have mentioned above, each component of the state is dependent on the process of the cumulative purchase/liquidation: $\{\boldsymbol{Q}_s\}_{0 \le s \le t}$.

The large trader's utility function takes the form of a *constant absolute risk-aversion* (CARA) *von Neumann–Morgenstern* (vN-M) *utility function*. The utility payoff (or reward) arises only from the terminal wealth at maturity:

$$
g_T(\boldsymbol{s}_T) := -\exp\left\{ -\gamma \Big[W_T - (\boldsymbol{P}_T + \boldsymbol{\chi}_T \overline{\boldsymbol{Q}}_T)^\top \overline{\boldsymbol{Q}}_T \Big] \right\}, \tag{32}
$$

where $\gamma \in \mathbb{R}_{++}$ denotes the risk-aversion parameter. In the formulation of this criteria, we assume that a large trader can execute her remaining execution volume at the terminal with closing price $\boldsymbol{P}_T$ and that the execution of all the remaining volume at time T imposes the large trader to pay the additive cost $\boldsymbol{\chi}_T \overline{\boldsymbol{Q}}_t$. In the following, we have the following assumption:

Assumption 3.1. The matrix:

$$
\boldsymbol{\chi}_T := \begin{pmatrix} \chi_T^{11} & \chi_T^{12} & \cdots & \chi_T^{1n} \\ \chi_T^{21} & \chi_T^{22} & \cdots & \chi_T^{2n} \\ \vdots & \vdots & \ddots & \vdots \\ \chi_T^{n1} & \chi_T^{n2} & \cdots & \chi_T^{nn} \end{pmatrix} \in \mathcal{M}^n(\mathbb{R}), \tag{33}
$$

is positive semidefinite.

The interpretation of this assumption is that the large trader cannot obtain some earnings by liquidating all the remaining volume at maturity.

Example 3.1 (Examples and its interpretation of terminal costs)**.** We can consider a variety of concrete forms of $\boldsymbol{\chi}_T \in \mathcal{M}^n(\mathbb{R})$ under Assumption 3.1. For example, one of a simple type of terminal costs is given by

$$\widetilde{\boldsymbol{\chi}}_T \coloneqq \begin{pmatrix} \chi_T^1 & 0 & \cdots & 0 \\ 0 & \chi_T^2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \chi_T^n \end{pmatrix} \in \mathcal{S}^n(\mathbb{R}_+). \tag{34}$$

This matrix implies that for each financial asset $i \in \{1, \dots, n\}$, liquidating the remaining volume $\overline{Q}_T^i$ at time T incurrs an additive costs $\chi^i \overline{Q}_T^i$. In addition, we can consider the following types of terminal costs:

$$\widehat{\boldsymbol{\chi}}_T \coloneqq \begin{pmatrix} \chi^{11} & \chi^{12} & \cdots & \chi^{1n} \\ \chi^{21} & \chi^{22} & \cdots & \chi^{2n} \\ \vdots & \vdots & \ddots & \vdots \\ \chi^{n1} & \chi^{22} & \cdots & \chi^{nn} \end{pmatrix} \in \mathcal{S}^n(\mathbb{R}), \tag{35}$$

where $\chi^{ij} < 0$ for $i, j \in \{1, \dots, n\}$ $(i \neq j)$. This corresponds to the case where by executing multiple assets, one can reduce the terminal cost to some degree.

We define the (conditional) expected utility of the large trader at time $t \in [0, T]$ on an execution strategy $\mathcal{Q} \coloneqq \{Q_t\}_{0 \le t \le T} \in \mathcal{A}$ as

$$V_t^{\mathcal{Q}} \coloneqq \mathbb{E}\left[-\exp\left\{ -\gamma \left[W_T - (\boldsymbol{P}_T + \boldsymbol{\chi}_T \overline{\boldsymbol{Q}}_T)^\top \overline{\boldsymbol{Q}}_T \right] \right\} \middle| \mathcal{F}_t \right]. \tag{36}$$

Let the optimal (expected utility) value from time $t \in [0, T]$ by

$$V_t \coloneqq \operatorname*{ess\,sup}_{\mathcal{Q} \in \mathcal{A}} V_t^{\mathcal{Q}}, \quad t \in [0, T]. \tag{37}$$

From the Markov property of the state dynamics, V_t depends on the history or information $\mathcal{F}_t$ only through the (controlled) state $\boldsymbol{s}_t \in \mathbb{S}$ and thus this functional dependence is represented by the optimal value function as

$$V[t, W_t, \boldsymbol{P}_t, \overline{\boldsymbol{Q}}_t, \boldsymbol{R}_t, \boldsymbol{\mathcal{I}}_t] \coloneqq V_t, \quad t \in [0, T]. \tag{38}$$

3.2 HJB Equation

From the dynamic programming principle, the optimal value function, denoted by

$$V[t, \boldsymbol{s}_t] \coloneqq V[t, W_t, \boldsymbol{P}_t, \overline{\boldsymbol{Q}}_t, \boldsymbol{R}_t, \boldsymbol{\mathcal{I}}_t] \tag{39}$$

with the terminal condition:

$$V[T, W_T, \boldsymbol{P}_T, \overline{\boldsymbol{Q}}_T, \boldsymbol{R}_T, \boldsymbol{\mathcal{I}}_T] = -\exp\left\{ -\gamma \left[W_T - (\boldsymbol{P}_T + \boldsymbol{\chi}_T \overline{\boldsymbol{Q}}_T)^\top \overline{\boldsymbol{Q}}_T \right] \right\}, \tag{40}$$

satisfies the following Hamilton–Jacobi–Bellman (HJB) equation (or dynamic programming equation) for the optimal execution speed (or optimal control) $\dot{Q}$:

$$\begin{aligned} \sup_{\dot{Q}_t \in \mathbb{R}} \Big[& \partial_t V - (\boldsymbol{P}_t + \boldsymbol{\Lambda}_t \dot{\boldsymbol{Q}}_t)^\top \dot{\boldsymbol{Q}}_t \partial_W V + \big\{ -\rho R_t + (\alpha_t + \beta_t)\lambda_t \dot{Q}_t + \boldsymbol{\mathcal{I}}_t + \mu_t^Z \big\} \partial_P V - \dot{\boldsymbol{Q}}_t^\top \partial_{\overline{\boldsymbol{Q}}} V \\ & + \big(-\rho R_t + \alpha_t \lambda_t \dot{Q}_t \big) \partial_R V + (\boldsymbol{a}_t^{\boldsymbol{\mathcal{I}}} - \mathbf{B}_t^{\boldsymbol{\mathcal{I}}} \boldsymbol{\mathcal{I}}_t)^\top \partial_{\boldsymbol{\mathcal{I}}} V \\ & + \frac{1}{2} \big\{ (\boldsymbol{\Sigma}_t^{\boldsymbol{Z}})^2 \partial_{\boldsymbol{PP}} V + 2\sigma_t^Z \sigma_t^{\boldsymbol{\mathcal{I}}} \rho^{Z,\boldsymbol{\mathcal{I}}} \partial_{P\boldsymbol{\mathcal{I}}} V + (\sigma_t^{\boldsymbol{\mathcal{I}}})^2 \partial_{\boldsymbol{\mathcal{I}\mathcal{I}}} V \big\} \Big] = 0, \quad 0 \le t < T, \end{aligned} \tag{41}$$

if we assume that the function $V\colon [0,T]\times\mathbb{S}\to\mathbb{R}$ is in $\mathcal{C}^{1,2}$.[6] Rewriting this results in

$$\sup_{\dot{\boldsymbol{Q}}_t\in\mathbb{R}^n}\left[-\left(\boldsymbol{P}_t+\boldsymbol{\Lambda}_t\dot{\boldsymbol{Q}}_t\right)^\top\dot{\boldsymbol{Q}}_t\partial_W V+(\alpha_t+\beta_t)\lambda_t\dot{Q}_t\partial_P V-\dot{\boldsymbol{Q}}_t\partial_{\overline{\boldsymbol{Q}}}V+\alpha_t\lambda_t\dot{\boldsymbol{Q}}_t\partial_{\boldsymbol{R}}V\right]$$
$$+\partial_t V+\left(-\rho R_t+\boldsymbol{\mathcal{I}}_t+\mu_t^Z\right)\partial_P V+(-\rho R_t)\partial_R V+(a_t^{\boldsymbol{\mathcal{I}}}-b_t^{\boldsymbol{\mathcal{I}}}\boldsymbol{\mathcal{I}}_t)\partial_{\boldsymbol{\mathcal{I}}}V$$
$$+\frac{1}{2}\left\{(\sigma_t^Z)^2\partial_{PP}V+2\sigma_t^{\boldsymbol{\mathcal{I}}}\sigma_t^Z\rho^{\boldsymbol{\mathcal{I}},Z}\partial_{P\boldsymbol{\mathcal{I}}}V+(\sigma_t^{\boldsymbol{\mathcal{I}}})^2\partial_{\boldsymbol{\mathcal{I}}\boldsymbol{\mathcal{I}}}V\right\}=0,\quad 0\le t<T. \tag{42}$$

We can derive the optimal execution strategy and its associated optimal value function of Eq. (38) explicitly by appropriately guessing the ansatz of the optimal value function and verifying the obtained solution.

3.3 Optimal Value Function and Optimal Execution Strategy

Theorem 3.1 (Optimal Execution Strategy and Optimal Value Function)**.** If the solution to the system of ordinary differential equations (ODEs) given by Eqs. (**??**) to (**??**) uniquely exists, the following statements hold:

1. The HJB equation (42) admits a solution $V[t,\boldsymbol{s}_t]:=V[t,W_t,P_t,\overline{Q}_t,R_t,\boldsymbol{\mathcal{I}}_t]$ at time $t\in[0,T]$ represented as

$$V[t,W_t,\boldsymbol{P}_t,\overline{\boldsymbol{Q}}_t,\boldsymbol{R}_t,\boldsymbol{\mathcal{I}}_t]$$
$$=-\exp\Big\{-\gamma\Big[W_t-\boldsymbol{P}_t^\top\overline{\boldsymbol{Q}}_t+\overline{\boldsymbol{Q}}_t^\top\mathbf{G}_t\overline{\boldsymbol{Q}}_t+\boldsymbol{H}_t^\top\overline{\boldsymbol{Q}}_t+\overline{\boldsymbol{Q}}_t^\top\mathbf{I}_t\boldsymbol{R}_t+\boldsymbol{R}_t^\top\mathbf{J}_t\boldsymbol{R}_t+\boldsymbol{L}_t^\top\boldsymbol{R}_t$$
$$+\overline{\boldsymbol{Q}}_t^\top\mathbf{M}_t\boldsymbol{\mathcal{I}}_t+\boldsymbol{R}_t^\top\mathbf{N}_t\boldsymbol{\mathcal{I}}_t+\boldsymbol{\mathcal{I}}_t^\top\mathbf{X}_t\boldsymbol{\mathcal{I}}_t+\boldsymbol{Y}_t^\top\boldsymbol{\mathcal{I}}_t+K_t\Big]\Big\}, \tag{43}$$

where

$$\dot{\mathbf{G}}_t=\frac{1}{2}\gamma\boldsymbol{\Sigma}^{\boldsymbol{Z}}(\boldsymbol{\Sigma}^{\boldsymbol{Z}})^\top+\frac{1}{2}\gamma\mathbf{M}_t\boldsymbol{\Sigma}_t^{\boldsymbol{\mathcal{I}}}(\boldsymbol{\Sigma}_t^{\boldsymbol{\mathcal{I}}})^\top\mathbf{M}_t^\top-\frac{\gamma}{4}\widetilde{\mathbf{B}}_t^\top\boldsymbol{\Lambda}^{-1}\widetilde{\mathbf{B}}_t; \tag{44}$$
$$\dot{\boldsymbol{H}}_t=\boldsymbol{\mu}_t^{\boldsymbol{Z}}-\mathbf{M}_t\boldsymbol{\mathfrak{J}}_t^{\boldsymbol{\mathcal{I}}}+\gamma\mathbf{M}_t\boldsymbol{\Sigma}_t^{\boldsymbol{\mathcal{I}}}(\boldsymbol{\Sigma}_t^{\boldsymbol{\mathcal{I}}})^\top\boldsymbol{Y}_t-\frac{\gamma}{2}\widetilde{\mathbf{B}}_t^\top\boldsymbol{\Lambda}^{-1}\widetilde{\boldsymbol{A}}_t; \tag{45}$$
$$\dot{\mathbf{I}}_t=-\boldsymbol{\Upsilon}+\mathbf{I}_t\boldsymbol{\Upsilon}+\gamma\mathbf{M}_t\boldsymbol{\Sigma}_t^{\boldsymbol{\mathcal{I}}}(\boldsymbol{\Sigma}_t^{\boldsymbol{\mathcal{I}}})^\top\mathbf{N}_t^\top-\frac{\gamma}{2}\widetilde{\mathbf{B}}_t^\top\boldsymbol{\Lambda}^{-1}\widetilde{\mathbf{C}}_t; \tag{46}$$
$$\dot{\mathbf{J}}_t=\boldsymbol{\Upsilon}(\mathbf{J}_t+\mathbf{J}_t^\top)+\frac{1}{2}\gamma\mathbf{N}_t\boldsymbol{\Sigma}_t^{\boldsymbol{\mathcal{I}}}(\boldsymbol{\Sigma}_t^{\boldsymbol{\mathcal{I}}})^\top\mathbf{N}_t-\frac{\gamma}{4}\widetilde{\mathbf{C}}_t^\top\boldsymbol{\Lambda}^{-1}\widetilde{\mathbf{C}}_t; \tag{47}$$
$$\dot{\boldsymbol{L}}_t=\boldsymbol{\Upsilon}^\top\boldsymbol{L}_t-\mathbf{N}_t\boldsymbol{\mathfrak{J}}_t^{\boldsymbol{\mathcal{I}}}+\gamma\mathbf{N}_t\boldsymbol{\Sigma}_t^{\boldsymbol{\mathcal{I}}}(\boldsymbol{\Sigma}_t^{\boldsymbol{\mathcal{I}}})^\top\boldsymbol{Y}_t-\frac{\gamma}{2}\widetilde{\mathbf{C}}_t^\top\boldsymbol{\Lambda}^{-1}\widetilde{\boldsymbol{A}}_t; \tag{48}$$
$$\dot{\mathbf{M}}_t=\boldsymbol{\kappa}_t+\mathbf{M}_t\boldsymbol{\mathfrak{K}}_t^{\boldsymbol{\mathcal{I}}}+\gamma\mathbf{M}_t\boldsymbol{\Sigma}_t^{\boldsymbol{\mathcal{I}}}(\boldsymbol{\Sigma}_t^{\boldsymbol{\mathcal{I}}})^\top(\mathbf{X}_t+\mathbf{X}_t^\top)-\frac{\gamma}{2}\widetilde{\mathbf{B}}_t^\top\boldsymbol{\Lambda}^{-1}\widetilde{\mathbf{D}}_t; \tag{49}$$
$$\dot{\mathbf{N}}_t=\boldsymbol{\Upsilon}^\top\mathbf{N}_t+\mathbf{N}_t\boldsymbol{\mathfrak{K}}_t^{\boldsymbol{\mathcal{I}}}+\gamma\mathbf{N}_t\boldsymbol{\Sigma}_t^{\boldsymbol{\mathcal{I}}}(\boldsymbol{\Sigma}_t^{\boldsymbol{\mathcal{I}}})^\top(\mathbf{X}_t+\mathbf{X}_t^\top)-\frac{\gamma}{2}\widetilde{\mathbf{C}}_t^\top\boldsymbol{\Lambda}^{-1}\widetilde{\mathbf{D}}_t; \tag{50}$$
$$\dot{\mathbf{X}}_t=(\boldsymbol{\mathfrak{K}}_t^{\boldsymbol{\mathcal{I}}})^\top(\mathbf{X}_t+\mathbf{X}_t^\top)+\frac{1}{2}\gamma(\mathbf{X}_t+\mathbf{X}_t^\top)\boldsymbol{\Sigma}_t^{\boldsymbol{\mathcal{I}}}(\boldsymbol{\Sigma}_t^{\boldsymbol{\mathcal{I}}})^\top(\mathbf{X}_t+\mathbf{X}_t^\top)-\frac{\gamma}{4}\widetilde{\mathbf{D}}_t^\top\boldsymbol{\Lambda}^{-1}\widetilde{\mathbf{D}}_t; \tag{51}$$
$$\dot{\boldsymbol{Y}}_t=-(\mathbf{X}_t+\mathbf{X}_t^\top)\boldsymbol{\mathfrak{J}}_t^{\boldsymbol{\mathcal{I}}}+(\boldsymbol{\mathfrak{K}}_t^{\boldsymbol{\mathcal{I}}})^\top\boldsymbol{Y}_t+\gamma(\mathbf{X}_t+\mathbf{X}_t^\top)\boldsymbol{\Sigma}_t^{\boldsymbol{\mathcal{I}}}(\boldsymbol{\Sigma}_t^{\boldsymbol{\mathcal{I}}})^\top\boldsymbol{Y}_t-\frac{\gamma}{2}\widetilde{\mathbf{D}}_t^\top\boldsymbol{\Lambda}^{-1}\widetilde{\boldsymbol{A}}_t; \tag{52}$$
$$\dot{K}_t=-(\boldsymbol{\mathfrak{J}}_t^{\boldsymbol{\mathcal{I}}})^\top\boldsymbol{Y}_t+\frac{1}{2}\gamma\boldsymbol{Y}_t^\top\boldsymbol{\Sigma}_t^{\boldsymbol{\mathcal{I}}}(\boldsymbol{\Sigma}_t^{\boldsymbol{\mathcal{I}}})^\top\boldsymbol{Y}_t-\frac{1}{2}\mathrm{tr}\Big((\boldsymbol{\Sigma}_t^{\boldsymbol{\mathcal{I}}})^\top(\mathbf{X}_t+\mathbf{X}_t^\top)\boldsymbol{\Sigma}_t^{\boldsymbol{\mathcal{I}}}\Big)-\frac{\gamma}{4}\widetilde{\boldsymbol{A}}_t^\top\boldsymbol{\Lambda}^{-1}\widetilde{\boldsymbol{A}}_t. \tag{53}$$

[6]That is, V is continuously differentiable with respect to time and continuously twice differentiable with respect to each state variable.

with the terminal conditions:

$$\begin{aligned}&\mathbf{G}_T = -\boldsymbol{\chi}_T \in \mathcal{M}^n(\mathbb{R}); \quad \boldsymbol{H}_T = \boldsymbol{L}_T = \mathbf{0} \in \mathbb{R}^n;\\ &\mathbf{I}_T = \mathbf{J}_T = \mathbf{0} \in \mathcal{M}^n(\mathbb{R}); \quad \mathbf{M}_T = \mathbf{N}_T = \mathbf{0} \in \mathcal{M}^{n,m}(\mathbb{R});\\ &\mathbf{X}_T = \mathbf{0} \in \mathcal{M}^m(\mathbb{R}); \quad \mathbf{Y}_T = \mathbf{0} \in \mathbb{R}^m; \quad K_T = 0 \in \mathbb{R}.\end{aligned} \tag{54}$$

Thus, if the optimal value function is of the form (62) and in $\mathcal{C}^{1,2}$, the unique solution is the optimal value function. Also, its coefficient functions $\mathbf{G}_t, \boldsymbol{H}_t, \mathbf{I}_t, \mathbf{J}_t, \boldsymbol{L}_t, \mathbf{M}_t, \mathbf{N}_t, \mathbf{X}_t, \boldsymbol{Y}_t, K_t$ are characterized (and obtained) by Eqs. (**??**) to (**??**).

2. Under the above assumption, the optimal execution speed at time $t \in [0, T]$, denoted as $\dot{\boldsymbol{Q}}_t^*$, becomes an *affine* function of the remaining execution volume $\overline{\boldsymbol{Q}}_t$, the cumulative residual effect $\boldsymbol{R}_t$, and the Markovian environment $\boldsymbol{\mathcal{I}}_t$ at time t:

$$\dot{\boldsymbol{Q}}_t^* = f_t(\boldsymbol{s}_t) = \boldsymbol{a}_t + \mathbf{b}_t^\top \overline{\boldsymbol{Q}}_t + \mathbf{c}_t^\top \boldsymbol{R}_t + \mathbf{d}_t^\top \boldsymbol{\mathcal{I}}_t, \quad 0 \le t \le T. \tag{55}$$

where

$$\boldsymbol{a}_t := -\frac{1}{2}\boldsymbol{\Lambda}_t^{-1}\left(\boldsymbol{H}_t^\top - \boldsymbol{L}_t^\top \boldsymbol{\alpha}_t \boldsymbol{\lambda}_t\right)^\top; \tag{56}$$

$$\mathbf{b}_t := -\frac{1}{2}\boldsymbol{\Lambda}_t^{-1}\left\{(\boldsymbol{\alpha}_t + \boldsymbol{\beta}_t)\boldsymbol{\Lambda}_t + (\mathbf{G}_t + \mathbf{G}_t^\top) - \boldsymbol{I}_t \boldsymbol{\alpha}_t \boldsymbol{\Lambda}_t\right\}^\top; \tag{57}$$

$$\mathbf{c}_t := -\frac{1}{2}\boldsymbol{\Lambda}_t^{-1}\left\{\mathbf{I}_t^\top - (\mathbf{J}_t + \mathbf{J}_t^\top)\boldsymbol{\alpha}_t \boldsymbol{\Lambda}_t\right\}^\top; \tag{58}$$

$$\mathbf{d}_t := -\frac{1}{2}\boldsymbol{\Lambda}_t^{-1}\left(\mathbf{M}_t^\top - \mathbf{N}_t^\top \boldsymbol{\alpha}_t \boldsymbol{\Lambda}_t\right)^\top. \tag{59}$$

Proof. See Appendix A. □

From Theorem 3.1, the optimal execution speed $\dot{\boldsymbol{Q}}_t^*$ depends on the state $\boldsymbol{s}_t = (W_t, \boldsymbol{P}_t, \overline{\boldsymbol{Q}}_t, \boldsymbol{R}_t, \boldsymbol{\mathcal{I}}_t)$ of the controlled process only through the remaining execution volume $\overline{\boldsymbol{Q}}_t$, the cumulative residual effect $\boldsymbol{R}_t$, and the Markovian environment $\boldsymbol{\mathcal{I}}_t$, not through the wealth W_t or market price $\boldsymbol{P}_t$. In addition, by definition of Markovian environment, the optimal execution volume $\dot{\boldsymbol{Q}}_t$ includes a nondeterministic term (random variable), the optimal execution strategy being a stochastic one. Then, the following corollary immediately holds.

Corollary 3.1. If the Markovian environment $\boldsymbol{\mathcal{I}}_t$ for all $t \in [0, T]$ are deterministic, the optimal execution speed $\dot{Q}_t^*$ at time $t \in [0, T]$ also becomes a deterministic function of time in a class of the static and non-randomized execution strategy.

3.4 Special case: No transient market impact

As a special case, we examine an optimal execution strategy for the model described so far without transient market impacts. If $\boldsymbol{\alpha}_t$ is zero, the residual effect of past market impacts becomes zero since we assume that $\boldsymbol{R}_0 = 0$. In this cases, the market price dynamics become

$$\mathrm{d}\boldsymbol{P}_t = \boldsymbol{\beta}_t \boldsymbol{\Lambda}_t \dot{\boldsymbol{Q}}_t \mathrm{d}t + \boldsymbol{\mathcal{I}}_t \mathrm{d}t + \mathrm{d}\boldsymbol{Z}_t, \tag{60}$$

that is, we have a permanent impact model with Markovian environment. In this case, our model partially includes the so-called "target zone models," in which the price of an asset traded by a large trader has one or more barriers and is reflected at the barriers (e.g., Krugman [20]; Neuman and Schied [26]; Belak et al. [4]). To be precise, if the large trader executes no orders, the price is capped by the mean-reverting process $\boldsymbol{\mathcal{I}}_t$. This model is motivated by the fact that monetary authorities (e.g., central banks) keep a currency exchange rate above some threshold.

In the case without transient market impacts, by setting the state variables as $\boldsymbol{s}_t := (W_t, P_t, \overline{Q}_t, \boldsymbol{\mathcal{I}}_t)$ we have the following results:

Corollary 3.2. Under the financial market considered in Section 2 without transient market impacts, the optimal execution speed becomes an *affine* function of the current remained execution volume and common risk factors:

$$\dot{\boldsymbol{Q}}_t^* = \widetilde{\boldsymbol{a}}_t + \widetilde{\mathbf{b}}_t \overline{\boldsymbol{Q}}_t + \widetilde{\mathbf{d}}_t \boldsymbol{\mathcal{I}}_t, \tag{61}$$

where $\widetilde{\boldsymbol{a}}_t$, $\widetilde{\mathbf{b}}_t$, and $\widetilde{\mathbf{d}}_t$ are all deterministic functions of time.

Proof. Omitted. □

4 Conclusion

We have examined how a large trader incorporates cross-impacts and common risk factors into an optimal pair-trade execution strategy in a finite continuous-time framework. The large trader maximizes the expected CARA utility arising from his/her wealth at the end of the trading epoch in a market. By formulating a generalized market impact model, the backward induction method of dynamic programming based on the dynamic programming principle permitted us to derive the optimal execution strategy. This kind of work concerned with an execution problem through the backward induction procedure of dynamic programming will be explored from a more in-depth and extensive perspective, which we can expect will also give us a more illuminating insight into all the other problems left in this field of research as follows.

Acknowledgements

The authors would like to thank Dr. Vaibhav Lalwani, Dr. Kohta Takehara, Dr. Takanori Adachi, Dr. Alexander Novikov, Dr. Nino Kordzakhia, Dr. Len Patrick Garces, and the participants at JAFEE-SOFINE-ISM International Symposium on Quantitative Finance and Actuarial Science and TMU Workshop on Finance 2025 for the comments and discussions, which improved the quality of this paper. We ALSO thank the mathematical research institute MATRIX in Australia where part of this research was performed.

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

This work was supported by Japan Society for the Promotion of Science under KAKENHI [Grant Numbers 24K16400, 25K05203, 25K00650].

References

[1] Almgren, R. & Chriss, N. (2000). Optimal execution of portfolio transactions. *Journal of Risk*, **3**, 5–40.

[2] Benzaquen, M., Mastromatteo, I., Eisler, Z., & Bouchaud, J. P. (2017). Dissecting cross-impact on stock markets: An empirical analysis. *Journal of Statistical Mechanics: Theory and Experiment*, **2017**(2), 023406.

[3] Bertsimas, D. & Lo, A. W. (1998). Optimal control of execution costs. *Journal of Financial Markets*, **1**(1), 1–50.

[4] Belak, C., Muhle-Karbe, J., & Ou, K. (2018). Liquidation in target zone models. *Market Microstructure and Liquidity*, **4**(03n04), 1950010.

[5] Bergault, P., Drissi, F., and Guéant, O. (2022). Multi-asset optimal execution and statistical arbitrage strategies under Ornstein–Uhlenbeck dynamics. *SIAM Journal on Financial Mathematics*, **13**(1), 353–390.

[6] Carhart, M. M. (1997). On persistence in mutual fund performance. *The Journal of Finance*, **52**(1), 57–82.

[7] Cartea, Á., Gan, L., & Jaimungal, S. (2019). Trading co-integrated assets with price impact. *Mathematical Finance*, **29**(2), 542–567.

[8] Cartea, Á, Jaimungal, S., & Penalva, J. (2015). *Algorithmic and High-Frequency Trading.* Cambridge University Press.

[9] Fama, E. F., & French, K. R. (1993). Common risk factors in the returns on stocks and bonds. *Journal of Financial Economics*, **33**(1), 3–56.

[10] Fama, E. F., & French, K. R. (2015). A five-factor asset pricing model. *Journal of Financial Economics*, **116**(1), 1–22.

[11] Fruth, A., Schöneborn, T., & Urusov, M. (2014). Optimal trade execution and price manipulation in order books with time-varying liquidity. *Mathematical Finance*, **24**(4), 651–695.

[12] Fukasawa, M., Ohnishi, M., & Shimoshimizu, M. (2021). Discrete–time optimal execution under a generalized price impact model with Markovian exogenous orders. *International Journal of Theoretical and Applied Finance*, **24**(05), 2150025.

[13] Fukasawa, M., Ohnishi, M., & Shimoshimizu, M. (2021). Optimal execution under a generalized price impact model with Markovian exogenous orders in a continuous–time setting. *RIMS Kokyuroku*, **2207**, 1–22.

[14] Fukasawa, M., Ohnishi, M., & Shimoshimizu, M. (2025). Continuous-time optimal execution and utility-based statistical arbitrage in a Markovian environment. Preprint.

[15] Gatev, E., Goetzmann, W. N., & Rouwenhorst, K. G. (2006). Pairs trading: Performance of a relative-value arbitrage rule. *The Review of Financial Studies*, **19**(3), 797–827.

[16] Gatheral, J., Schied, A., & Slynko, A. (2012). Transient linear price impact and Fredholm integral equations. *Mathematical Finance*, **22**(3), 445–474.

[17] Gârleanu, N. & Pedersen, L. H. (2013). Dynamic trading with predictable returns and transaction costs. *The Journal of Finance*, **68** (6), 2309–2340.

[18] Gârleanu, N. & Pedersen, L. H. (2016). Dynamic portfolio choice with frictions. *Journal of Economic Theory*, **165**, 487–516.

[19] Guéant, O. (2016). *The Financial Mathematics of Market Liquidity: From Optimal Execution to Market Making.* Boca Raton, Florida: CRC Press.

[20] Krugman, P. R. (1991). Target zones and exchange rate dynamics. *The Quarterly Journal of Economics*, **106**(3), 669–682.

[21] Kuno, S. & Ohnishi, M. (2015). Optimal execution in illiquid market with the absence of price manipulation. *Journal of Mathematical Finance*, **5**(01), 1–14.

[22] Kuno, S., Ohnishi, M., & Shimizu, P. (2017). Optimal off-exchange execution with closing price. *Journal of Mathematical Finance*, **7**(1), 54–64.

[23] Laruelle, S., & Lehalle, C. A. (2018). *Market Microstructure in Practice Second Edition.* World Scientific.

[24] Lehalle, C. A., & Neuman, E. (2019). Incorporating signals into optimal trading. *Finance and Stochastics*, **23**(2), 275–311.

[25] Ma, G., Siu, C. C. & Zhu, S.-P. (2019). Dynamic portfolio choice with return predictability and transaction costs. *European Journal of Operational Research* **278**(3), 976–988.

[26] Neuman, E., & Schied, A. (2016). Optimal portfolio liquidation in target zone models and catalytic superprocesses. *Finance and Stochastics*, **20**, 495–509.

[27] Obizhaeva, A. A., & Wang, J. (2013). Optimal trading strategy and supply/demand dynamics. *Journal of Financial Markets*, **16**(1), 1–32.

[28] Ohnishi, M., & Shimoshimizu, M. (2019). Equilibrium execution strategies with generalized price impacts. *RIMS Kokyuroku*, **2111**, 84–106.

[29] Ohnishi, M., & Shimoshimizu, M. (2020). Optimal and equilibrium execution strategies with generalized price impact. *Quantitative Finance*, **20**(10), 1625–1644.

[30] Ohnishi, M., & Shimoshimizu, M. (2022). Optimal pair–trade execution with generalized cross–impact. *Asia-Pacific Financial Markets*, **29**, 253–289.

[31] Ohnishi, M., and Shimoshimizu, M. (2024). Trade execution games in a Markovian environment. *Applied Mathematical Finance*, **31**(4), 239–277.

[32] Shimoshimizu, M. Introduction to optimal execution. In *Machine Learning Approaches in Financial Analytics*, pp.3–50. Springer, Cham. (Patnaik, S., Tripathy, N., and Maglaras, L. A. (eds)).

[33] Schneider, M. & Lillo, F. (2019). Cross-impact and no-dynamic-arbitrage. *Quantitative Finance*, **19**(1), 137–154.

[34] Tsoukalas, G., Wang, J., & Giesecke, K. (2019). Dynamic portfolio execution. *Management Science*, **65**(5), 2015–2040.

Appendix

A Proof of Theorem 3.1

From the single asset case of Fukasawa et al. [14], we guess the objective (or value) function as follows:

$$\begin{aligned} V[t, W_t, \boldsymbol{P}_t, \overline{\boldsymbol{Q}}_t, \boldsymbol{R}_t, \boldsymbol{\mathcal{I}}_t] = -\exp\Big\{-\gamma\Big[& W_t - \boldsymbol{P}_t^\top \overline{\boldsymbol{Q}}_t + \overline{\boldsymbol{Q}}_t^\top \mathbf{G}_t \overline{\boldsymbol{Q}}_t + \boldsymbol{H}_t^\top \overline{\boldsymbol{Q}}_t + \overline{\boldsymbol{Q}}_t^\top \mathbf{I}_t \boldsymbol{R}_t + \boldsymbol{R}_t^\top \mathbf{J}_t \boldsymbol{R}_t + \boldsymbol{L}_t^\top \boldsymbol{R}_t \\ & + \overline{\boldsymbol{Q}}_t^\top \mathbf{M}_t \boldsymbol{\mathcal{I}}_t + \boldsymbol{R}_t^\top \mathbf{N}_t \boldsymbol{\mathcal{I}}_t + \boldsymbol{\mathcal{I}}_t^\top \mathbf{X}_t \boldsymbol{\mathcal{I}}_t + \boldsymbol{Y}_t^\top \boldsymbol{\mathcal{I}}_t + K_t\Big]\Big\}, \end{aligned} \tag{62}$$

with the terminal condition:

$$V[T, W_T, \boldsymbol{P}_T, \overline{\boldsymbol{Q}}_T, \boldsymbol{R}_T, \boldsymbol{\mathcal{I}}_T] = -\exp\Big\{-\gamma\Big[W_T - (\boldsymbol{P}_T + \boldsymbol{\chi}_T\overline{\boldsymbol{Q}}_T)^\top\overline{\boldsymbol{Q}}_T\Big]\Big\}. \tag{63}$$

The partial differentiation of $V\left[t, W_t, \boldsymbol{P}_t, \overline{\boldsymbol{Q}}_t, \boldsymbol{R}_t, \boldsymbol{\mathcal{I}}_t\right]$ with respect to time and each state variable is calculated as follows:

$$\begin{aligned}
\partial_t V &= -\gamma\Big\{\overline{\boldsymbol{Q}}_t^\top \dot{\mathbf{G}}_t\overline{\boldsymbol{Q}}_t + \dot{\boldsymbol{H}}_t^\top\overline{\boldsymbol{Q}}_t + \overline{\boldsymbol{Q}}_t^\top\dot{\mathbf{I}}_t\boldsymbol{R}_t + \boldsymbol{R}_t^\top\dot{\mathbf{J}}_t\boldsymbol{R}_t + \dot{\boldsymbol{L}}_t^\top\boldsymbol{R}_t \\
&\qquad + \overline{\boldsymbol{Q}}_t^\top\dot{\mathbf{M}}_t\boldsymbol{\mathcal{I}}_t + \boldsymbol{R}_t^\top\dot{\mathbf{N}}_t\boldsymbol{\mathcal{I}}_t + \boldsymbol{\mathcal{I}}_t^\top\dot{\mathbf{X}}_t\boldsymbol{\mathcal{I}}_t + \dot{\boldsymbol{Y}}_t^\top\boldsymbol{\mathcal{I}}_t + \dot{K}_t\Big\}V; \\
\partial_W V &= -\gamma V; \\
\partial_{\boldsymbol{P}} V &= \gamma\overline{\boldsymbol{Q}}_t V; \\
\partial_{\overline{\boldsymbol{Q}}} V &= -\gamma\left(-\boldsymbol{P}_t + 2\mathbf{G}_t\overline{\boldsymbol{Q}}_t + \boldsymbol{H}_t + \mathbf{I}_t\boldsymbol{R}_t + \mathbf{M}_t\boldsymbol{\mathcal{I}}_t\right)V; \\
\partial_{\boldsymbol{R}} V &= -\gamma\left(\mathbf{I}_t\overline{\boldsymbol{Q}}_t + 2\mathbf{J}_t\boldsymbol{R}_t + \boldsymbol{L}_t + \mathbf{N}_t\boldsymbol{\mathcal{I}}_t\right)V; \\
\partial_{\boldsymbol{\mathcal{I}}} V &= -\gamma(\mathbf{M}_t\overline{\boldsymbol{Q}}_t + \mathbf{N}_t\boldsymbol{R}_t + 2\mathbf{X}_t\boldsymbol{\mathcal{I}}_t + \boldsymbol{Y}_t)V; \\
\partial_{\boldsymbol{P}\boldsymbol{P}^\top} V &= \gamma^2\overline{\boldsymbol{Q}}_t\overline{\boldsymbol{Q}}_t^\top V; \\
\partial_{\boldsymbol{P}\boldsymbol{\mathcal{I}}^\top} V &= -\gamma^2\overline{\boldsymbol{Q}}_t(\mathbf{M}_t\overline{\boldsymbol{Q}}_t + \mathbf{N}_t\boldsymbol{R}_t + 2\mathbf{X}_t\boldsymbol{\mathcal{I}}_t + \boldsymbol{Y}_t)^\top V; \\
\partial_{\boldsymbol{\mathcal{I}}\boldsymbol{\mathcal{I}}^\top} V &= -2\gamma\mathbf{X}_t V + \gamma^2(\mathbf{M}_t\overline{\boldsymbol{Q}}_t + \mathbf{N}_t\boldsymbol{R}_t + 2\mathbf{X}_t\boldsymbol{\mathcal{I}}_t + \boldsymbol{Y}_t)(\mathbf{M}_t\overline{\boldsymbol{Q}}_t + \mathbf{N}_t\boldsymbol{R}_t + 2\mathbf{X}_t\boldsymbol{\mathcal{I}}_t + \boldsymbol{Y}_t)^\top V.
\end{aligned}$$

Therefore, by substituting these into Eq. (42), we have

$$\begin{aligned}
&\sup_{\dot{\boldsymbol{Q}}_t\in\mathbb{R}^n} \gamma\Bigg[\dot{\boldsymbol{Q}}_t^\top\boldsymbol{\Lambda}_t\dot{\boldsymbol{Q}}_t + \Big[\overline{\boldsymbol{Q}}_t^\top\Big\{(\boldsymbol{\alpha}_t + \boldsymbol{\beta}_t)\boldsymbol{\Lambda}_t + (\mathbf{G}_t + \mathbf{G}_t^\top) - \boldsymbol{I}_t\boldsymbol{\alpha}_t\boldsymbol{\Lambda}_t\Big\} + \boldsymbol{R}_t^\top\Big\{\mathbf{I}_t^\top - (\mathbf{J}_t + \mathbf{J}_t^\top)\boldsymbol{\alpha}_t\boldsymbol{\Lambda}_t\Big\} \\
&\qquad + \boldsymbol{\mathcal{I}}_t^\top\Big(\mathbf{M}_t^\top - \mathbf{N}_t^\top\boldsymbol{\alpha}_t\boldsymbol{\Lambda}_t\Big) + \Big(\boldsymbol{H}_t^\top - \boldsymbol{L}_t^\top\boldsymbol{\alpha}_t\boldsymbol{\lambda}_t\Big)\Big]\dot{\boldsymbol{Q}}_t\Bigg]V \\
&\qquad + \gamma\overline{\boldsymbol{Q}}_t^\top\Big\{-\dot{\mathbf{G}}_t + \frac{1}{2}\gamma\boldsymbol{\Sigma}^{\boldsymbol{Z}}(\boldsymbol{\Sigma}^{\boldsymbol{Z}})^\top + \frac{1}{2}\gamma\mathbf{M}_t\boldsymbol{\Sigma}_t^{\boldsymbol{\mathcal{I}}}(\boldsymbol{\Sigma}_t^{\boldsymbol{\mathcal{I}}})^\top\mathbf{M}_t^\top\Big\}\overline{\boldsymbol{Q}}_t V \\
&\qquad + \gamma\left\{-\dot{\boldsymbol{H}}_t + \boldsymbol{\mu}_t^{\boldsymbol{Z}} - \mathbf{M}_t\boldsymbol{\mathfrak{J}}_t^{\boldsymbol{\mathcal{I}}} + \gamma\mathbf{M}_t\boldsymbol{\Sigma}_t^{\boldsymbol{\mathcal{I}}}(\boldsymbol{\Sigma}_t^{\boldsymbol{\mathcal{I}}})^\top\boldsymbol{Y}_t\right\}^\top\overline{\boldsymbol{Q}}_t V \\
&\qquad + \gamma\overline{\boldsymbol{Q}}_t^\top\left\{-\dot{\mathbf{I}}_t - \boldsymbol{\Upsilon} + \mathbf{I}_t\boldsymbol{\Upsilon} + \gamma\mathbf{M}_t\boldsymbol{\Sigma}_t^{\boldsymbol{\mathcal{I}}}(\boldsymbol{\Sigma}_t^{\boldsymbol{\mathcal{I}}})^\top\mathbf{N}_t^\top\right\}\boldsymbol{R}_t V \\
&\qquad + \gamma\boldsymbol{R}_t^\top\left\{-\dot{\mathbf{J}}_t + \boldsymbol{\Upsilon}(\mathbf{J}_t + \mathbf{J}_t^\top) + \frac{1}{2}\gamma\mathbf{N}_t\boldsymbol{\Sigma}_t^{\boldsymbol{\mathcal{I}}}(\boldsymbol{\Sigma}_t^{\boldsymbol{\mathcal{I}}})^\top\mathbf{N}_t\right\}\boldsymbol{R}_t V \\
&\qquad + \gamma\left\{-\dot{\boldsymbol{L}}_t + \boldsymbol{\Upsilon}^\top\boldsymbol{L}_t - \mathbf{N}_t\boldsymbol{\mathfrak{J}}_t^{\boldsymbol{\mathcal{I}}} + \gamma\mathbf{N}_t\boldsymbol{\Sigma}_t^{\boldsymbol{\mathcal{I}}}(\boldsymbol{\Sigma}_t^{\boldsymbol{\mathcal{I}}})^\top\boldsymbol{Y}_t\right\}^\top\boldsymbol{R}_t V \\
&\qquad + \gamma\overline{\boldsymbol{Q}}_t^\top\left\{-\dot{\mathbf{M}}_t + \boldsymbol{\kappa}_t + \mathbf{M}_t\boldsymbol{\mathfrak{K}}_t^{\boldsymbol{\mathcal{I}}} + \gamma\mathbf{M}_t\boldsymbol{\Sigma}_t^{\boldsymbol{\mathcal{I}}}(\boldsymbol{\Sigma}_t^{\boldsymbol{\mathcal{I}}})^\top(\mathbf{X}_t + \mathbf{X}_t^\top)\right\}\boldsymbol{\mathcal{I}}_t V \\
&\qquad + \gamma\boldsymbol{R}_t^\top\left\{-\dot{\mathbf{N}}_t + \boldsymbol{\Upsilon}^\top\mathbf{N}_t + \mathbf{N}_t\boldsymbol{\mathfrak{K}}_t^{\boldsymbol{\mathcal{I}}} + \gamma\mathbf{N}_t\boldsymbol{\Sigma}_t^{\boldsymbol{\mathcal{I}}}(\boldsymbol{\Sigma}_t^{\boldsymbol{\mathcal{I}}})^\top(\mathbf{X}_t + \mathbf{X}_t^\top)\right\}\boldsymbol{\mathcal{I}}_t V \\
&\qquad + \gamma\boldsymbol{\mathcal{I}}_t^\top\left\{-\dot{\mathbf{X}}_t + (\boldsymbol{\mathfrak{K}}_t^{\boldsymbol{\mathcal{I}}})^\top(\mathbf{X}_t + \mathbf{X}_t^\top) + \frac{1}{2}\gamma(\mathbf{X}_t + \mathbf{X}_t^\top)\boldsymbol{\Sigma}_t^{\boldsymbol{\mathcal{I}}}(\boldsymbol{\Sigma}_t^{\boldsymbol{\mathcal{I}}})^\top(\mathbf{X}_t + \mathbf{X}_t^\top)\right\}\boldsymbol{\mathcal{I}}_t V \\
&\qquad + \gamma\left\{-\dot{\boldsymbol{Y}}_t - (\mathbf{X}_t + \mathbf{X}_t^\top)\boldsymbol{\mathfrak{J}}_t^{\boldsymbol{\mathcal{I}}} + (\boldsymbol{\mathfrak{K}}_t^{\boldsymbol{\mathcal{I}}})^\top\boldsymbol{Y}_t + \gamma(\mathbf{X}_t + \mathbf{X}_t^\top)\boldsymbol{\Sigma}_t^{\boldsymbol{\mathcal{I}}}(\boldsymbol{\Sigma}_t^{\boldsymbol{\mathcal{I}}})^\top\boldsymbol{Y}_t\right\}^\top\boldsymbol{\mathcal{I}}_t V \\
&\qquad + \gamma\left\{-\dot{K}_t - (\boldsymbol{\mathfrak{J}}_t^{\boldsymbol{\mathcal{I}}})^\top\boldsymbol{Y}_t + \frac{1}{2}\gamma\boldsymbol{Y}_t^\top\boldsymbol{\Sigma}_t^{\boldsymbol{\mathcal{I}}}(\boldsymbol{\Sigma}_t^{\boldsymbol{\mathcal{I}}})^\top\boldsymbol{Y}_t - \frac{1}{2}\mathrm{tr}\Big((\boldsymbol{\Sigma}_t^{\boldsymbol{\mathcal{I}}})^\top(\mathbf{X}_t + \mathbf{X}_t^\top)\boldsymbol{\Sigma}_t^{\boldsymbol{\mathcal{I}}}\Big)\right\}V \\
&= 0.
\end{aligned} \tag{64}$$

Since we assume the negative exponential utility function above,

$$
\begin{aligned}
&\sup_{\dot{\boldsymbol{Q}}_t\in\mathbb{R}^n}\gamma\Big[\dot{\boldsymbol{Q}}_t^\top\boldsymbol{\Lambda}_t\dot{\boldsymbol{Q}}_t+\Big[\overline{\boldsymbol{Q}}_t^\top\Big\{(\boldsymbol{\alpha}_t+\boldsymbol{\beta}_t)\boldsymbol{\Lambda}_t+(\mathbf{G}_t+\mathbf{G}_t^\top)-\boldsymbol{I}_t\boldsymbol{\alpha}_t\boldsymbol{\Lambda}_t\Big\}+\boldsymbol{R}_t^\top\Big\{\mathbf{I}_t^\top-(\mathbf{J}_t+\mathbf{J}_t^\top)\boldsymbol{\alpha}_t\boldsymbol{\Lambda}_t\Big\}\\
&\qquad+\boldsymbol{\mathcal{I}}_t^\top\Big(\mathbf{M}_t^\top-\mathbf{N}_t^\top\boldsymbol{\alpha}_t\boldsymbol{\Lambda}_t\Big)+\Big(\boldsymbol{H}_t^\top-\boldsymbol{L}_t^\top\boldsymbol{\alpha}_t\boldsymbol{\lambda}_t\Big)\Big]\dot{\boldsymbol{Q}}_t\Big]V\\
&=V\inf_{\dot{\boldsymbol{Q}}_t\in\mathbb{R}^n}\gamma\Big[\dot{\boldsymbol{Q}}_t^\top\boldsymbol{\Lambda}_t\dot{\boldsymbol{Q}}_t+\Big[\overline{\boldsymbol{Q}}_t^\top\Big\{(\boldsymbol{\alpha}_t+\boldsymbol{\beta}_t)\boldsymbol{\Lambda}_t+(\mathbf{G}_t+\mathbf{G}_t^\top)-\boldsymbol{I}_t\boldsymbol{\alpha}_t\boldsymbol{\Lambda}_t\Big\}+\boldsymbol{R}_t^\top\Big\{\mathbf{I}_t^\top-(\mathbf{J}_t+\mathbf{J}_t^\top)\boldsymbol{\alpha}_t\boldsymbol{\Lambda}_t\Big\}\\
&\qquad+\boldsymbol{\mathcal{I}}_t^\top\Big(\mathbf{M}_t^\top-\mathbf{N}_t^\top\boldsymbol{\alpha}_t\boldsymbol{\Lambda}_t\Big)+\Big(\boldsymbol{H}_t^\top-\boldsymbol{L}_t^\top\boldsymbol{\alpha}_t\boldsymbol{\lambda}_t\Big)\Big]\dot{\boldsymbol{Q}}_t\Big]\\
&=V\inf_{\dot{Q}_t}\gamma\Big[\dot{\boldsymbol{Q}}_t^\top\boldsymbol{\Lambda}_t\dot{\boldsymbol{Q}}_t+\Big[\overline{\boldsymbol{Q}}_t^\top\mathbf{B}_t^\top+\boldsymbol{R}_t^\top\mathbf{C}_t^\top+\boldsymbol{\mathcal{I}}_t^\top\mathbf{D}_t+\boldsymbol{A}_t^\top\Big]\dot{\boldsymbol{Q}}_t\Big], \qquad (65)
\end{aligned}
$$

where

$$
\begin{aligned}
\boldsymbol{A}_t^\top &\coloneqq \boldsymbol{H}_t^\top-\boldsymbol{L}_t^\top\boldsymbol{\alpha}_t\boldsymbol{\lambda}_t; &(66)\\
\mathbf{B}_t^\top &\coloneqq (\boldsymbol{\alpha}_t+\boldsymbol{\beta}_t)\boldsymbol{\Lambda}_t+(\mathbf{G}_t+\mathbf{G}_t^\top)-\boldsymbol{I}_t\boldsymbol{\alpha}_t\boldsymbol{\Lambda}_t; &(67)\\
\mathbf{C}_t^\top &\coloneqq \mathbf{I}_t^\top-(\mathbf{J}_t+\mathbf{J}_t^\top)\boldsymbol{\alpha}_t\boldsymbol{\Lambda}_t; &(68)\\
\mathbf{D}_t^\top &\coloneqq \mathbf{M}_t^\top-\mathbf{N}_t^\top\boldsymbol{\alpha}_t\boldsymbol{\Lambda}_t. &(69)
\end{aligned}
$$

Therefore, Eq. (65) attains the infimum at the optimal execution speed:

$$
\dot{\boldsymbol{Q}}_t^*=-\frac{1}{2}\boldsymbol{\Lambda}_t^{-1}\Big(\widetilde{\mathbf{B}}_t\overline{\boldsymbol{Q}}_t+\widetilde{\mathbf{C}}_t\boldsymbol{R}_t+\widetilde{\mathbf{D}}_t\boldsymbol{\mathcal{I}}_t+\widetilde{\boldsymbol{A}}_t\Big)\eqqcolon \boldsymbol{a}_t+\mathbf{b}_t\overline{\boldsymbol{Q}}_t+\mathbf{c}_t\boldsymbol{R}_t+\mathbf{d}_t\boldsymbol{\mathcal{I}}_t, \qquad (70)
$$

where

$$
\boldsymbol{a}_t\coloneqq-\frac{1}{2}\boldsymbol{\Lambda}_t^{-1}\widetilde{\boldsymbol{A}}_t;\quad \mathbf{b}_t\coloneqq-\frac{1}{2}\boldsymbol{\Lambda}_t^{-1}\widetilde{\mathbf{B}}_t;\quad \mathbf{c}_t\coloneqq-\frac{1}{2}\boldsymbol{\Lambda}_t^{-1}\widetilde{\mathbf{C}}_t;\quad \mathbf{d}_t\coloneqq-\frac{1}{2}\boldsymbol{\Lambda}_t^{-1}\widetilde{\mathbf{D}}_t. \qquad (71)
$$

Substituting this into Eq. (64) yields

$$
\begin{aligned}
&\gamma\Big\{-\dot{\mathbf{G}}_t+\frac{1}{2}\gamma\boldsymbol{\Sigma}^{\boldsymbol{Z}}(\boldsymbol{\Sigma}^{\boldsymbol{Z}})^\top+\frac{1}{2}\gamma\mathbf{M}_t\boldsymbol{\Sigma}_t^{\boldsymbol{\mathcal{I}}}(\boldsymbol{\Sigma}_t^{\boldsymbol{\mathcal{I}}})^\top\mathbf{M}_t^\top-\frac{\gamma}{4}\widetilde{\mathbf{B}}_t^\top\boldsymbol{\Lambda}^{-1}\widetilde{\mathbf{B}}_t\Big\}\overline{Q}_t^2V\\
&+\gamma\Big\{-\dot{\boldsymbol{H}}_t+\boldsymbol{\mu}_t^{\boldsymbol{Z}}-\mathbf{M}_t\boldsymbol{\mathfrak{J}}_t^{\boldsymbol{\mathcal{I}}}+\gamma\mathbf{M}_t\boldsymbol{\Sigma}_t^{\boldsymbol{\mathcal{I}}}(\boldsymbol{\Sigma}_t^{\boldsymbol{\mathcal{I}}})^\top\boldsymbol{Y}_t-\frac{\gamma}{2}\widetilde{\mathbf{B}}_t^\top\boldsymbol{\Lambda}^{-1}\widetilde{\boldsymbol{A}}_t\Big\}\overline{Q}_tV\\
&+\gamma\Big\{-\dot{\mathbf{I}}_t-\boldsymbol{\Upsilon}+\mathbf{I}_t\boldsymbol{\Upsilon}+\gamma\mathbf{M}_t\boldsymbol{\Sigma}_t^{\boldsymbol{\mathcal{I}}}(\boldsymbol{\Sigma}_t^{\boldsymbol{\mathcal{I}}})^\top\mathbf{N}_t-\frac{\gamma}{2}\widetilde{\mathbf{B}}_t^\top\boldsymbol{\Lambda}^{-1}\widetilde{\mathbf{C}}_t\Big\}\overline{Q}_tR_tV\\
&+\gamma\Big\{-\dot{\mathbf{J}}_t+\boldsymbol{\Upsilon}(\mathbf{J}_t+\mathbf{J}_t^\top)+\frac{1}{2}\gamma\mathbf{N}_t\boldsymbol{\Sigma}_t^{\boldsymbol{\mathcal{I}}}(\boldsymbol{\Sigma}_t^{\boldsymbol{\mathcal{I}}})^\top\mathbf{N}_t-\frac{\gamma}{4}\widetilde{\mathbf{C}}_t^\top\boldsymbol{\Lambda}^{-1}\widetilde{\mathbf{C}}_t\Big\}R_t^2V\\
&+\gamma\Big\{-\dot{\boldsymbol{L}}_t+\boldsymbol{\Upsilon}^\top\boldsymbol{L}_t-\mathbf{N}_t\boldsymbol{\mathfrak{J}}_t^{\boldsymbol{\mathcal{I}}}+\gamma\mathbf{N}_t\boldsymbol{\Sigma}_t^{\boldsymbol{\mathcal{I}}}(\boldsymbol{\Sigma}_t^{\boldsymbol{\mathcal{I}}})^\top\boldsymbol{Y}_t-\frac{\gamma}{2}\widetilde{\mathbf{C}}_t^\top\boldsymbol{\Lambda}^{-1}\widetilde{\boldsymbol{A}}_t\Big\}R_tV\\
&+\gamma\Big\{-\dot{\mathbf{M}}_t+\boldsymbol{\kappa}_t+\mathbf{M}_t\boldsymbol{\mathfrak{K}}_t^{\boldsymbol{\mathcal{I}}}+\gamma\mathbf{M}_t\boldsymbol{\Sigma}_t^{\boldsymbol{\mathcal{I}}}(\boldsymbol{\Sigma}_t^{\boldsymbol{\mathcal{I}}})^\top(\mathbf{X}_t+\mathbf{X}_t^\top)-\frac{\gamma}{2}\widetilde{\mathbf{B}}_t^\top\boldsymbol{\Lambda}^{-1}\widetilde{\mathbf{D}}_t\Big\}\overline{Q}_t\boldsymbol{\mathcal{I}}_tV\\
&+\gamma\Big\{-\dot{\mathbf{N}}_t+\boldsymbol{\Upsilon}^\top\mathbf{N}_t+\mathbf{N}_t\boldsymbol{\mathfrak{K}}_t^{\boldsymbol{\mathcal{I}}}+\gamma\mathbf{N}_t\boldsymbol{\Sigma}_t^{\boldsymbol{\mathcal{I}}}(\boldsymbol{\Sigma}_t^{\boldsymbol{\mathcal{I}}})^\top(\mathbf{X}_t+\mathbf{X}_t^\top)-\frac{\gamma}{2}\widetilde{\mathbf{C}}_t^\top\boldsymbol{\Lambda}^{-1}\widetilde{\mathbf{D}}_t\Big\}R_t\boldsymbol{\mathcal{I}}_tV\\
&+\gamma\Big\{-\dot{\mathbf{X}}_t+(\boldsymbol{\mathfrak{K}}_t^{\boldsymbol{\mathcal{I}}})^\top(\mathbf{X}_t+\mathbf{X}_t^\top)+\frac{1}{2}\gamma(\mathbf{X}_t+\mathbf{X}_t^\top)\boldsymbol{\Sigma}_t^{\boldsymbol{\mathcal{I}}}(\boldsymbol{\Sigma}_t^{\boldsymbol{\mathcal{I}}})^\top(\mathbf{X}_t+\mathbf{X}_t^\top)-\frac{\gamma}{4}\widetilde{\mathbf{D}}_t^\top\boldsymbol{\Lambda}^{-1}\widetilde{\mathbf{D}}_t\Big\}\boldsymbol{\mathcal{I}}_t^2V\\
&+\gamma\Big\{-\dot{\boldsymbol{Y}}_t-(\mathbf{X}_t+\mathbf{X}_t^\top)\boldsymbol{\mathfrak{J}}_t^{\boldsymbol{\mathcal{I}}}+(\boldsymbol{\mathfrak{K}}_t^{\boldsymbol{\mathcal{I}}})^\top\boldsymbol{Y}_t+\gamma(\mathbf{X}_t+\mathbf{X}_t^\top)\boldsymbol{\Sigma}_t^{\boldsymbol{\mathcal{I}}}(\boldsymbol{\Sigma}_t^{\boldsymbol{\mathcal{I}}})^\top\boldsymbol{Y}_t-\frac{\gamma}{2}\widetilde{\mathbf{D}}_t^\top\boldsymbol{\Lambda}^{-1}\widetilde{\boldsymbol{A}}_t\Big\}\boldsymbol{\mathcal{I}}_tV\\
&+\gamma\Big\{-\dot{K}_t-(\boldsymbol{\mathfrak{J}}_t^{\boldsymbol{\mathcal{I}}})^\top\boldsymbol{Y}_t+\frac{1}{2}\gamma\boldsymbol{Y}_t^\top\boldsymbol{\Sigma}_t^{\boldsymbol{\mathcal{I}}}(\boldsymbol{\Sigma}_t^{\boldsymbol{\mathcal{I}}})^\top\boldsymbol{Y}_t-\frac{1}{2}\mathrm{tr}\Big((\boldsymbol{\Sigma}_t^{\boldsymbol{\mathcal{I}}})^\top(\mathbf{X}_t+\mathbf{X}_t^\top)\boldsymbol{\Sigma}_t^{\boldsymbol{\mathcal{I}}}\Big)-\frac{\gamma}{4}\widetilde{\boldsymbol{A}}_t^\top\boldsymbol{\Lambda}^{-1}\widetilde{\boldsymbol{A}}_t\Big\}V\\
&=0. \qquad (72)
\end{aligned}
$$

Since this equation holds for all states, the following conditions must hold:

$$\dot{\mathbf{G}}_t = \frac{1}{2}\gamma \boldsymbol{\Sigma}^{\boldsymbol{Z}}(\boldsymbol{\Sigma}^{\boldsymbol{Z}})^\top + \frac{1}{2}\gamma \mathbf{M}_t \boldsymbol{\Sigma}_t^{\boldsymbol{\mathcal{I}}}(\boldsymbol{\Sigma}_t^{\boldsymbol{\mathcal{I}}})^\top \mathbf{M}_t^\top - \frac{\gamma}{4}\widetilde{\mathbf{B}}_t^\top \boldsymbol{\Lambda}^{-1}\widetilde{\mathbf{B}}_t; \tag{73}$$

$$\dot{\boldsymbol{H}}_t = \boldsymbol{\mu}_t^{\boldsymbol{Z}} - \mathbf{M}_t \boldsymbol{\mathfrak{J}}_t^{\boldsymbol{\mathcal{I}}} + \gamma \mathbf{M}_t \boldsymbol{\Sigma}_t^{\boldsymbol{\mathcal{I}}}(\boldsymbol{\Sigma}_t^{\boldsymbol{\mathcal{I}}})^\top \boldsymbol{Y}_t - \frac{\gamma}{2}\widetilde{\mathbf{B}}_t^\top \boldsymbol{\Lambda}^{-1}\widetilde{\boldsymbol{A}}_t; \tag{74}$$

$$\dot{\mathbf{I}}_t = -\boldsymbol{\Upsilon} + \mathbf{I}_t\boldsymbol{\Upsilon} + \gamma \mathbf{M}_t \boldsymbol{\Sigma}_t^{\boldsymbol{\mathcal{I}}}(\boldsymbol{\Sigma}_t^{\boldsymbol{\mathcal{I}}})^\top \mathbf{N}_t^\top - \frac{\gamma}{2}\widetilde{\mathbf{B}}_t^\top \boldsymbol{\Lambda}^{-1}\widetilde{\mathbf{C}}_t; \tag{75}$$

$$\dot{\mathbf{J}}_t = \boldsymbol{\Upsilon}(\mathbf{J}_t + \mathbf{J}_t^\top) + \frac{1}{2}\gamma \mathbf{N}_t \boldsymbol{\Sigma}_t^{\boldsymbol{\mathcal{I}}}(\boldsymbol{\Sigma}_t^{\boldsymbol{\mathcal{I}}})^\top \mathbf{N}_t - \frac{\gamma}{4}\widetilde{\mathbf{C}}_t^\top \boldsymbol{\Lambda}^{-1}\widetilde{\mathbf{C}}_t; \tag{76}$$

$$\dot{\boldsymbol{L}}_t = \boldsymbol{\Upsilon}^\top \boldsymbol{L}_t - \mathbf{N}_t \boldsymbol{\mathfrak{J}}_t^{\boldsymbol{\mathcal{I}}} + \gamma \mathbf{N}_t \boldsymbol{\Sigma}_t^{\boldsymbol{\mathcal{I}}}(\boldsymbol{\Sigma}_t^{\boldsymbol{\mathcal{I}}})^\top \boldsymbol{Y}_t - \frac{\gamma}{2}\widetilde{\mathbf{C}}_t^\top \boldsymbol{\Lambda}^{-1}\widetilde{\boldsymbol{A}}_t; \tag{77}$$

$$\dot{\mathbf{M}}_t = \boldsymbol{\kappa}_t + \mathbf{M}_t \boldsymbol{\mathfrak{K}}_t^{\boldsymbol{\mathcal{I}}} + \gamma \mathbf{M}_t \boldsymbol{\Sigma}_t^{\boldsymbol{\mathcal{I}}}(\boldsymbol{\Sigma}_t^{\boldsymbol{\mathcal{I}}})^\top (\mathbf{X}_t + \mathbf{X}_t^\top) - \frac{\gamma}{2}\widetilde{\mathbf{B}}_t^\top \boldsymbol{\Lambda}^{-1}\widetilde{\mathbf{D}}_t; \tag{78}$$

$$\dot{\mathbf{N}}_t = \boldsymbol{\Upsilon}^\top \mathbf{N}_t + \mathbf{N}_t \boldsymbol{\mathfrak{K}}_t^{\boldsymbol{\mathcal{I}}} + \gamma \mathbf{N}_t \boldsymbol{\Sigma}_t^{\boldsymbol{\mathcal{I}}}(\boldsymbol{\Sigma}_t^{\boldsymbol{\mathcal{I}}})^\top (\mathbf{X}_t + \mathbf{X}_t^\top) - \frac{\gamma}{2}\widetilde{\mathbf{C}}_t^\top \boldsymbol{\Lambda}^{-1}\widetilde{\mathbf{D}}_t; \tag{79}$$

$$\dot{\mathbf{X}}_t = (\boldsymbol{\mathfrak{K}}_t^{\boldsymbol{\mathcal{I}}})^\top (\mathbf{X}_t + \mathbf{X}_t^\top) + \frac{1}{2}\gamma (\mathbf{X}_t + \mathbf{X}_t^\top)\boldsymbol{\Sigma}_t^{\boldsymbol{\mathcal{I}}}(\boldsymbol{\Sigma}_t^{\boldsymbol{\mathcal{I}}})^\top (\mathbf{X}_t + \mathbf{X}_t^\top) - \frac{\gamma}{4}\widetilde{\mathbf{D}}_t^\top \boldsymbol{\Lambda}^{-1}\widetilde{\mathbf{D}}_t; \tag{80}$$

$$\dot{\boldsymbol{Y}}_t = -(\mathbf{X}_t + \mathbf{X}_t^\top)\boldsymbol{\mathfrak{J}}_t^{\boldsymbol{\mathcal{I}}} + (\boldsymbol{\mathfrak{K}}_t^{\boldsymbol{\mathcal{I}}})^\top \boldsymbol{Y}_t + \gamma (\mathbf{X}_t + \mathbf{X}_t^\top)\boldsymbol{\Sigma}_t^{\boldsymbol{\mathcal{I}}}(\boldsymbol{\Sigma}_t^{\boldsymbol{\mathcal{I}}})^\top \boldsymbol{Y}_t - \frac{\gamma}{2}\widetilde{\mathbf{D}}_t^\top \boldsymbol{\Lambda}^{-1}\widetilde{\boldsymbol{A}}_t; \tag{81}$$

$$\dot{K}_t = -(\boldsymbol{\mathfrak{J}}_t^{\boldsymbol{\mathcal{I}}})^\top \boldsymbol{Y}_t + \frac{1}{2}\gamma \boldsymbol{Y}_t^\top \boldsymbol{\Sigma}_t^{\boldsymbol{\mathcal{I}}}(\boldsymbol{\Sigma}_t^{\boldsymbol{\mathcal{I}}})^\top \boldsymbol{Y}_t - \frac{1}{2}\mathrm{tr}\Big((\boldsymbol{\Sigma}_t^{\boldsymbol{\mathcal{I}}})^\top (\mathbf{X}_t + \mathbf{X}_t^\top)\boldsymbol{\Sigma}_t^{\boldsymbol{\mathcal{I}}}\Big) - \frac{\gamma}{4}\widetilde{\boldsymbol{A}}_t^\top \boldsymbol{\Lambda}^{-1}\widetilde{\boldsymbol{A}}_t, \tag{82}$$

with the terminal conditions:

$$\begin{aligned} &\mathbf{G}_T = -\boldsymbol{\chi}_T \in \mathcal{M}^n(\mathbb{R}); \quad \boldsymbol{H}_T = \boldsymbol{L}_T = \mathbf{0} \in \mathbb{R}^n; \\ &\mathbf{I}_T = \mathbf{J}_T = \mathbf{0} \in \mathcal{M}^n(\mathbb{R}); \quad \mathbf{M}_T = \mathbf{N}_T = \mathbf{0} \in \mathcal{M}^{n,m}(\mathbb{R}); \\ &\mathbf{X}_T = \mathbf{0} \in \mathcal{M}^m(\mathbb{R}); \quad \boldsymbol{Y}_T = \mathbf{0} \in \mathbb{R}^m; \quad K_T = 0 \in \mathbb{R}. \end{aligned} \tag{83}$$

By substituting the dynamics of $\boldsymbol{a}_t, \mathbf{b}_t, \mathbf{c}_t, \mathbf{d}_t$ into the condition derived above and rearranging, we obtain a system of ordinary differential equations consisting of $\mathbf{G}_t, \boldsymbol{H}_t, \mathbf{I}_t, \mathbf{J}_t, \boldsymbol{L}_t, \mathbf{M}_t, \mathbf{N}_t, \mathbf{X}_t, \boldsymbol{Y}_t, K_t$.
$\square$