Quantitative Finance

ISSN: (Print) (Online) Journal homepage: www.tandfonline.com/journals/rquf20

Relative entropy-regularized robust optimal order execution

Meng Wang & Tai-Ho Wang

To cite this article: Meng Wang & Tai-Ho Wang (2025) Relative entropy-regularized robust optimal order execution, Quantitative Finance, 25:3, 383-401, DOI: 10.1080/14697688.2025.2465704

To link to this article: https://doi.org/10.1080/14697688.2025.2465704

Published online: 10 Mar 2025.

Submit your article to this journal

Article views: 76

View related articles

View Crossmark data

Full Terms & Conditions of access and use can be found at
https://www.tandfonline.com/action/journalInformation?journalCode=rquf20

Quantitative Finance, 2025
Vol. 25, No. 3, 383–401, https://doi.org/10.1080/14697688.2025.2465704

Relative entropy-regularized robust optimal order execution

MENG WANG and TAI-HO WANG*

Baruch College, City University of New York, 1 Bernard Baruch Way, New York, NY 10010, USA

(*Received 7 December 2023; accepted 1 February 2025*)

The problem of order execution is cast as a relative entropy-regularized robust optimal control problem in this article. The order execution agent's goal is to maximize an objective functional associated with his profit-and-loss of trading and simultaneously minimize the inventory risk associated with market's liquidity and uncertainty. We model the market's liquidity and uncertainty by the principle of least relative entropy associated with the market trading rate. The problem of order execution is made into a relative entropy-regularized stochastic differential game. Standard argument of dynamic programming yields that the value function of the differential game satisfies a relative entropy-regularized Hamilton–Jacobi–Isaacs (rHJI) equation. Under the assumptions of linear-quadratic model with Gaussian prior, the rHJI equation reduces to a system of Riccati and linear differential equations. Further imposing constancy of the corresponding coefficients, the system of differential equations can be solved in closed form, resulting in analytical expressions for optimal strategy and trajectory as well as the posterior distribution of market trading rate. Numerical examples illustrating the optimal strategies and the comparisons with conventional trading strategies are conducted.

Keywords: Optimal execution; Price impact; Inventory cost; Entropy-regularized stochastic control

1. Introduction

Order execution, a mission that algorithmic trading departments and execution brokerage agencies embark on regularly, is cast as a relative entropy-regularized robust optimal control problem. During the course of executing a large order of significant amount, the agent faces with not only the risk of price impact that his own execution would incur towards the transaction price but also the liquidity and uncertainty of the market. The agent's goal is to maximize an objective functional associated with his profit-and-loss (P&L) of trading and simultaneously minimize the execution risk.

A quote by Kyle in Kyle (1985) states that

> *Roughly speaking, Black defines a liquid market as one which is almost infinitely tight, which is not infinitely deep, and which is resilient enough so that prices eventually tend to their underlying value.*

As such, we model the market's liquidity and uncertainty by the *principle of least relative entropy* associated with the market trading rate. In other words, the market is resilient enough so that the probability distribution of trading rate of the market, though impacted due to the presence of the order execution agent's trading, would stay as close, in terms of the Kullback–Leibler divergence, as possible to the distribution prior to the presence of the agent's trades. Such consideration also provides a framework for the order execution agent to assess the market liquidity and uncertainty risk incurred from his own trade. Henceforth, the agent will be able to dynamically adjust his trading strategy accordingly. It is worth to mention that, in literature on optimal order execution, market resilience was usually modeled as the resilience of a limit order book. For instance, as a pioneer work in this direction, Obizhaeva and Wang (2013) considered the case of flat order book density and Alfonsi *et al.* (2010) generalized it to general order book shape. In these models, the resilience of limit order book is depicted, after the limit order book is hit or lifted by market orders, as the recovery in an exponential rate back to its original stationary form of order book density. In the current article, we take a different route, we model market resilience by the principle of least relative entropy as mentioned earlier. This formulation also has the advantage of applying reinforcement learning techniques, similar to Kim and Yang (2020) and Wang *et al.* (2020), to the optimal execution framework. The regularization by Kullback–Leibler divergence prevents drastic change to market trading rate, the market is thus considered resilient in this sense.

*Corresponding author. Email: tai-ho.wang@baruch.cuny.edu

© 2025 Informa UK Limited, trading as Taylor & Francis Group

The order execution agent is considered as playing a stochastic differential game against the market in that he attempts to minimize the entropy-regularized risk while at the same time maximizing his own P&L at the liquidation horizon, with possible penalty at a terminal time should there be a final block trade. The standard argument of dynamic programming remains applicable in this setting, it follows that the value function of the differential game satisfies a relative entropy-regularized Hamilton–Jacobi–Isaacs (rHJI) equation. Under the assumptions of linear-quadratic model with Gaussian prior, the rHJI equation reduces to a system of Riccati and linear differential equations. Further imposing constancy of the corresponding coefficients, the system of differential equations can be solved in closed form, resulting in analytical expressions for the optimal trading strategy, the optimal expected trading trajectory as well as the posterior distribution of the market trading rate. The optimal strategies obtained in this case resemble the renowned Almgren–Chriss strategy in Almgren and Chriss (2001). However, parts of the parameters involved in the determination of the strategies are from the prior distribution of market trading rate, the liquidity and uncertainty risk, as well as relative entropy penalty factor, which are naturally embedded into the optimal strategies.

The rest of the paper is organized as follows. We formulate and model the order execution problem in section 2. Section 3 recasts the optimal order execution problem as a relative-entropy-regularized stochastic differential game and derives the associated Hamilton–Jacobi–Isaacs equation. Section 4 proposes two model assumptions that allow us to solve the Hamilton–Jacobi–Isaacs equation in closed form and presents the associated optimal trading strategies and trajectories. Verification theorems and duality between the primal and the dual problems are shown in sections 5. Section 6 briefly discusses the extension to non-constant price volatility. Numerical illustrations and discussions on the results are reported in section 7. Section 8 concludes the paper with discussions. Finally, we briefly review the relative entropy-regularized control problem and its corresponding Hamilton–Jacobi–Bellman equation, called the *relative entropy-regularized HJB equation*, as an appendix in section.

Throughout the paper, $(\Omega, \mathcal{F}, \mathbb{P})$ denotes a complete probability space equipped with a filtration describing the information structure $\mathbb{F} := \{\mathcal{F}_t\}_{t\in[0,T]}$, where t is the time variable and $T > 0$ the fixed finite liquidation horizon. Let $\{W_t^S, W_t^X\}_{t\in[0,T]}$ be a two-dimensional Brownian motion with constant correlation ρ defined on $(\Omega, \mathcal{F}, \mathbb{P})$. The filtration $\mathbb{F}$ is generated by the trajectories of the above Brownian motion, completed with all $\mathbb{P}$-null measure sets of $\mathcal{F}$.

2. Model setup

In this section, we layout the problem and set up the model in a continuous time setting for an order execution agent who is missioned to liquidate a given amount of shares of certain stock under price impact within the time interval $[0, T]$. In addition to execution risk, the model takes into account market's liquidity and uncertainty risk. In the following, X_t denotes the agent's holdings at time t, S_t the efficient or mid price of the stock, and $\tilde{S}_t$ the agent's transacted price at time t.

2.1. Price impact model

To take into account the liquidity and uncertainty risk from market trading rate when executing orders, we propose a price impact model as follows. Let a_t be the random market trading rate at time t, following distribution π_t. The efficient price S_t is assumed governed by the stochastic differential equation (SDE)

$$dS_t = \gamma\, dX_t + \gamma_M \langle a\rangle_t\, dt + \sigma_t^S\, dW_t^S, \tag{1}$$

where $\langle\cdot\rangle_t$ denotes averaging over π_t, i.e. $\langle a\rangle_t = \int a\pi_t(a)\, da$ and $\sigma_t^S = \sqrt{\int \tilde{\sigma}^2(a)\pi_t(a)\, da}$ for some $\tilde{\sigma}$. Though it is well-documented that the volatility σ_t^S is highly correlated with market trading rate, we shall for simplicity assume σ^S is constant throughout the article. The traded price $\tilde{S}_t$ at time t is given by

$$\tilde{S}_t = S_t + \eta v_t \tag{2}$$

for $\eta > 0$, where v_t is the agent's intended trading rate at time t. The $\gamma\, dX_t$ terms in (1) is usually referred to as the *permanent impact* and the ηv_t term in $\tilde{S}_t$ as the *temporary impact*. Note that in (1) we assume the averaged market trading rate $\langle a\rangle$ also has an impact permanently to the efficient price; however, it carries no temporary impact contributing to the transacted price. This assumption is consistent with regular activities in the market in the sense that if in average there are more buy orders than sell orders in the market, i.e. $\langle a\rangle > 0$, the efficient price will increase; otherwise, the price decreases. Also if buy and sell orders even out in average, apart from the possible permanent impact incurred from the agent's trading, the efficient price is a martingale.

We remark that, to our knowledge, most literature on order execution do not particularly single out the market trading rate $\langle a\rangle$ but rather include it in the diffusion part. However, there exist literature considering impact from other market participants by imposing stochastic price impact. For example, Mariani and Fatone (2022) introduce an additional noise to the execution price resulting from other investors' trading activities. Ma *et al.* (2020) model the net permanent impact from other participants in the market by a mean-reverting process, and both permanent and temporary price impact coefficients of the investor are assumed to follow a jump process on a finite state space. In this paper, we factor it out from the diffusion term so as to assess the market's liquidity and uncertainty risk in order execution. In order to mitigate the risk of model misspecification and be ambiguity aversion, we cast the agent's order execution problem as a robust control problem, i.e. optimizing the objective function in the worst-case scenario as will be described in section 3. We refer to the seminal papers Anderson *et al.* (2003), Hansen and Miao (2018) and Maccheroni *et al.* (2006) for more detailed studies and discussions on model misspecification, robust control as well as ambiguity aversion utilizing relative-entropy.

2.2. Agent's P&L

A recent paper by Carmona and Leal in Carmona and Leal (2023) showed pretty convincingly that the presence of a Brownian component in the agent's inventory is statistically significant. To incorporate, we model the agent's holdings X_t at time t as

$$
\begin{aligned}
&\mathrm{d}X_t = v_t\,\mathrm{d}t + \sigma^X\,\mathrm{d}W_t^X, \quad t \in [0, T], \\
&X_0 = x_0 > 0
\end{aligned} \tag{3}
$$

should the agent's intended trading rate is v_t at time t. The Brownian motion component W_t^X and the volatility σ^X proxy the uncertainty of the agent's holdings. The two Brownian motions W_t^S in (1) and W_t^X in (3) are assumed correlated with correlation ρ. Concerning the correlation ρ between the two Brownian motions, it is argued in Carmona and Webster (2019) that the correlation ρ is negative when trading with market orders and positive when trading with limit orders. Also, optimal order execution in this situation has been studied. For example, Cheng *et al.* (2017) focus on the order execution problem with different utility functions, Cheng *et al.* (2019) in addition consider the evolution of the riskiness of the penalty due to the final block trade, Di Giacinto *et al.* (2024) introduces the price pressure driven by market makers' inventories' risk, and Mariani and Fatone (2022) add additional noise in the execution price due to the activities of other investors.

The agent's profit-and-loss (P&L hereafter) Π_t^0 during the execution course up to time t is given by

$$
\Pi_t^0 := X_t(S_t - S_0) + \int_0^t (S_0 - \tilde{S}_u)\,\mathrm{d}X_u. \tag{4}
$$

Note that the first term on the right-hand side of (4) captures the change in fair (marked-to-market) value of the remaining untransacted shares, while the second term measures the transaction gain/loss resulting from selling shares due to spread, volatility, and price impact. Moreover, since the agent is mandated to liquidate all his position at the liquidation horizon T, if he is left in his holdings with X_T shares at the terminal time T, his final P&L Π_T at time T is to be penalized by $g(X_T)$, where $g(0) = 0$ and $g < 0$ if $X_T \neq 0$. Thus, the agent's P&L Π_T after taking a final block trade at time T is given by

$$
\Pi_T = \Pi_T^0 + g(X_T).
$$

We note that, by applying Ito's formula and straightforward calculations, the P&L Π_T^0 may be rewritten, without explicit dependence on price S, as

$$
\begin{aligned}
\Pi_T^0 = {} & \frac{\gamma}{2}(X_T^2 - X_0^2) + \gamma_M \int_0^T X_t \langle a \rangle_t\,\mathrm{d}t + \int_0^T X_t \sigma^S\,\mathrm{d}W_t^S \\
& + \frac{\gamma}{2}\int_0^T (\sigma^X)^2\,\mathrm{d}t \\
& + \int_0^T \rho\sigma^S\sigma^X\,\mathrm{d}t - \eta\int_0^T v_t^2\,\mathrm{d}t - \eta\int_0^T v_t\sigma^X\,\mathrm{d}W_t^X.
\end{aligned} \tag{5}
$$

Hence, the expected P&L Π^0 prior to a final block trade is given by

$$
\begin{aligned}
\mathbb{E}\left[\Pi_T^0\right] = {} & -\frac{\gamma}{2}X_0^2 + \mathbb{E}\left[\frac{\gamma}{2}X_T^2 + \int_0^T \left\{\gamma_M X_t \langle a \rangle_t + \frac{\gamma}{2}(\sigma^X)^2 \right.\right. \\
& \left.\left. + \rho\sigma^S\sigma^X - \eta v_t^2\right\}\mathrm{d}t\right].
\end{aligned}
$$

We shall ignore the constant term $-\frac{\gamma}{2}X_0^2$ in the equation above when time comes to formulate and solve the agent's order execution as a control problem.

2.3. Execution and market risk

The agent's risk, which consists of the execution risk and the market's liquidity and uncertainty risk within the time interval $[t, T]$, is modeled as

$$
g_M(S_T) + \int_t^T R(u, X_u, S_u, v_u, a_u)\,\mathrm{d}u
$$

for some given terminal risk g_M on the stock price and running risk R. For example, the running risk R may include the quadratic variation of P&L Π_0 over liquidation horizon as one of its risk components. We refer to the assumptions imposed in section 4 for more detailed discussions on imposing the running risk R.

In summary, the performance criterion for the agent's execution is given by the execution P&L penalized by the execution risk and the market liquidity and uncertainty risk. Namely,

$$
\begin{aligned}
& g(X_T) + X_t(S_t - S_0) + \int_0^t (S_0 - \tilde{S}_u)\,\mathrm{d}X_u + g_M(S_T) \\
& + \int_t^T R(u, X_u, S_u, v_u, a_u)\,\mathrm{d}u.
\end{aligned}
$$

We recall that the first three terms in the equation above correspond to the execution P&L penalized by a final block trade while the last two terms represent risks.

3. Order execution as regularized stochastic differential game

With the model set up in section 2, we recast the agent's order execution as a *relative entropy-regularized stochastic differential game* in this section. The agent is regarded as playing a game against the market in a conservative manner in the sense that, rather than rigorously minimizing the risks incurred from the market as in the methodology of robust optimal control, he only attempts to minimize entropy-regularized risk while at the same time maximizing his P&L at the liquidation horizon.

Alternatively, inspired by Fisher Black's notion on liquid market as stated in Kyle (1985), which we reiterate in the following

> *'Roughly speaking, Black defines a liquid market as one which is almost infinitely tight, which is not infinitely deep, and which is resilient enough so that prices eventually tend to their underlying value'*,

this concept of minimizing the relative entropy-regularized risk can be thought of as representing the market resilience via the *principle of least relative entropy* for market activity. In other words, the distribution π of market's trading rate a is also impacted due to the presence of the agent's trading. However, it is resilient enough so as to stay as close, in terms of the Kullback–Leibler divergence, as possible to the prior distribution π^0.

3.1. Order execution problem

We regard the differential game played by the market is to minimize the following relative entropy-regularized control problem among admissible distributions $\pi \in \mathcal{A}$ for the market trading rate a

$$\min_{\pi\in\mathcal{A}} \mathbb{E}\left[g_M(S_T) + \int_0^T \int \{R(u, X_u, S_u, v_u, a) + \frac{1}{\beta}\log\frac{\pi_u(a)}{\pi_u^0(a)}\right\} \pi_u(a)\,\mathrm{d}a\,\mathrm{d}u \right]$$

for some $\beta > 0$. The set of admissible distributions $\mathcal{A}$ consists of the distributions π_t that are adapted to the filtration $\mathcal{F}_t$ and absolutely continuous to a given prior distribution π_t^0 for $t \in [0, T]$. We now formulate the agent's order execution problem as the following relative entropy-regularized robust optimal control problem.

$$\max_{v\in\mathcal{V}} \min_{\pi\in\mathcal{A}} \mathbb{E}\left[g(X_T) + X_T S_T - \int_0^T \tilde{S}_t\,\mathrm{d}X_t + g_M(S_T) + \int_0^T \int \left\{ R(t, X_t, S_t, v_t, a) + \frac{1}{\beta}\log\frac{\pi_t(a)}{\pi_t^0(a)} \right\} \pi_t(a)\,\mathrm{d}a\,\mathrm{d}t \right], \tag{6}$$

where $\mathcal{V}$ consists of adapted, square integrable processes in $[0, T]$.

We remark that (a) the collection of admissible distributions in (6) represents a Knightian uncertainty† ; (b) the inner minimization problem can be further subject to certain constraints on the posterior distribution such as first and second order statistics. Finally, by substituting (5) for P&L, the problem (6) is transformed into

$$\max_{v\in\mathcal{V}} \min_{\pi\in\mathcal{A}} \mathbb{E}\left[G(X_T, S_T) + \iint_0^T \left\{ \tilde{R}(t, X_t, S_t, v_t, a) + \frac{1}{\beta}\log\frac{\pi_t(a)}{\pi_t^0(a)} \right\} \pi_t(a)\,\mathrm{d}a\,\mathrm{d}t \right], \tag{7}$$

where $G(X_T, S_T) := g(X_T) + \frac{\gamma}{2}X_T^2 + g_M(S_T)$ and

$$\tilde{R}(t, x, s, v, a) = \gamma_M x a + \frac{\gamma}{2}(\sigma^X)^2 + \rho\sigma^S\sigma^X - \eta v^2 + R(t, x, s, v, a).$$

† Knightian uncertainty is named after the renowned economist Frank Knight who formalized a distinction between risk and uncertainty in his 1921 book entitled *Risk, Uncertainty, and Profit*. An easy-to-access explanation on Knightian uncertainty can be found at the webpage news.mit.edu/2010/explained-knightian-0602.

We shall be mainly dealing with the problem (7) in the following sections.

3.2. The relative entropy-regularized Hamilton–Jacobi–Isaacs equation

To solve the execution problem (7) which can be regarded as a stochastic differential game, define the value function V by

$$V(t, x, s) = \max_{v\in\mathcal{V}_t} \min_{\pi\in\mathcal{A}_t} \mathbb{E}_t\left[G(X_T, S_T) + \iint_t^T \left\{ \tilde{R}(\tau, X_\tau, S_\tau, v_\tau, a_\tau) + \frac{1}{\beta}\log\frac{\pi_\tau(a)}{\pi_\tau^0(a)} \right\} \pi_\tau(a)\,\mathrm{d}a\,\mathrm{d}\tau \right]. \tag{8}$$

where $\mathbb{E}_t[\cdot]$ denotes the conditional expectation $\mathbb{E}[\cdot \mid \mathcal{F}_t]$ and assume $(X_t, S_t) = (x, s)$. $\mathcal{V}_t$ and $\mathcal{A}_t$ denote the admissible strategies and distributions in $\mathcal{V}$ and $\mathcal{A}$, respectively, that are restricted to the time interval $[t, T]$. By applying the standard dynamical programming principle argument, one can show that the value function V in (8) satisfies a *relative entropy-regularized Hamilton–Jacobi–Isaacs* (rHJI hereafter) equation which we summarize in the following theorem whose proof is omitted.

THEOREM 3.1 *The value function V in* (8) *satisfies the following rHJI equation*

$$\max_{v} \min_{\pi\ll\pi_t^0} \int \left\{ V_t + \mathcal{L}V + \tilde{R}(t, x, s, v, a) + \frac{1}{\beta}\log\frac{\pi(a)}{\pi_t^0(a)} \right\} \pi(a)\,\mathrm{d}a = 0, \tag{9}$$

with terminal condition $V(T, x, s) = G(x, s)$, *assuming enough regularity for the value function V.* $\mathcal{L}$ *denotes the infinitesimal generator for the SDEs*

$$\mathrm{d}S_t = \gamma\,\mathrm{d}X_t + \gamma_M a_t\,\mathrm{d}t + \sigma^S\,\mathrm{d}W_t^S,$$
$$\mathrm{d}X_t = v_t\,\mathrm{d}t + \sigma^X\,\mathrm{d}W_t^X,$$

where the Brownian motions W_t^S *and* W_t^X *are correlated with correlation* ρ.

4. Solutions in closed form

In this section, we show under further model assumptions that the rHJI equation (9) can be reduced into a Hamilton–Jacobi equation and present, in the linear-quadratic (LQ) framework, solutions in closed form to the resulting effective Hamilton–Jacobi equation. We remark that in the following when describing a Gaussian distribution, we shall mostly use the term *precision*, defined by the reciprocal of variance, rather than the commonly used term of variance since it simplifies the notations a little.

The following assumptions are imposed in order for the problem (7) to remain in a Gaussian and LQ structure through other choices are by all means possible.

(A1) The '*prior*' distribution π_t^0 is Gaussian with mean m_t and precision s_t. Precisely,

$$\pi_t^0(a) = \sqrt{\frac{s_t}{2\pi}} \mathrm{e}^{-\frac{s_t}{2}(a-m_t)^2}.$$

(A2) $g_M(s) = 0$, $g(x) = -\delta x^2$. R is independent of s and is quadratic in x, a. Specifically,

$$R = R(t, x, a) = \frac{R_{xx}}{2} x^2 + R_{xa} xa + \frac{R_{aa}}{2} a^2$$

where the coefficients may be time-dependent. Moreover, the coefficient R_{aa} is further assumed satisfying $s_t + \beta R_{aa} > 0$ for all t, where recall that s_t is the precision for the prior Gaussian distribution π_t^0 given in Assumption (A1).s

(A3) $g_M(s) = 0$, $g(x) = -\delta x^2$. R is independent of s and is quadratic in v, a. Specifically,

$$R = R(t, v, a) = \frac{R_{vv}}{2} v^2 + R_{va} va + \frac{R_{aa}}{2} a^2$$

where the coefficients may be time-dependent. Moreover, the coefficients R_{vv}, R_{va}, and R_{aa} are further assumed to satisfy

$$s_t + \beta R_{aa} > 0 \quad \text{and} \quad \eta - \frac{R_{vv}}{2} + \frac{\beta R_{va}^2}{2(s_t + \beta R_{aa})} > 0$$

for all t.

We briefly explain the financial rationale for the assumptions (A2) and (A3) on R being quadratic. Concerning (A2), as one observes in (5) that the x^2 term may result from the quadratic variation of P&L Π_t^0. For the xa term, consider the following *volume-weighted averaged efficient price* for the trading trajectory X_t.

$$\begin{aligned}
\frac{1}{X_0}\int_0^T S_t\,\mathrm{d}X_t &= \frac{1}{X_0}\left\{S_T X_T - X_0 S_0 - \int_0^T X_t\,\mathrm{d}S_t - [X,S]_T\right\} \\
&\approx -S_0 + \frac{1}{X_0}\left\{-\gamma\int_0^T X_t\,\mathrm{d}X_t - \gamma_M\int_0^T X_t a_t\,\mathrm{d}t \right. \\
&\quad \left. - \int_0^T X_t\sigma^S\,\mathrm{d}W_t^S - \int_0^T \rho\sigma^S\sigma^X\,\mathrm{d}t\right\} \\
&\approx -S_0 + \frac{1}{X_0}\left\{-\gamma\int_0^T X_t\,\mathrm{d}X_t - \gamma_M\int_0^T X_t a_t\,\mathrm{d}t \right. \\
&\quad \left. - \int_0^T X_t\sigma^S\,\mathrm{d}W_t^S\right\} \\
&\approx -S_0 + \frac{\gamma}{2}X_0 - \frac{\gamma_M}{X_0}\int_0^T X_t a_t\,\mathrm{d}t - \frac{1}{X_0}\int_0^T X_t\sigma^S\,\mathrm{d}W_t^S.
\end{aligned} \tag{10}$$

Hence, apart from the constants and the martingale term in (10), if the agent is subject to minimize the volume-weighted averaged efficient price of his strategy, he may consider penalizing his final P&L by a factor of the third term in (10), which is a special case of (A2). As for (A3), consider the situation where the order execution agent is required to track a percentage-of-volume (POV) strategy as a benchmark. That is, with a given market participant rate $\nu > 0$, at any point in time he would like his trading v_t to stay close to νa_t, should the market trading rate be a_t. In this case, he may penalize his P&L by the quantity $\lambda|v - \nu a|^2$, where $\lambda > 0$ is a risk aversion parameter, which indeed is a special case for R in (A3).

In both (A2) and (A3), R is assumed independent of price S_t. The rationale is that risk, be it of price or of execution, is considered closely related with the relative price movements, which originate from the agent's and the market's trading activities, rather than the absolute price level. We note that, since under these assumptions the problem (7) and hence its associated value function do not depend on S_t any more, the SDE for S_t can be discharged from the problem (7) and thus the infinitesimal generator $\mathcal{L}$ in the rHJI equation (9) reduces to $\mathcal{L} = \frac{(\sigma^X)^2}{2}\partial_x^2 + v\partial_x$.

4.1. Model 1

In this subsection, we show that under Assumptions (A1) and (A2), which we refer to as *Model 1*, the rHJI equation (9) can be reduced to an effective Hamilton–Jacobi equation that admits a solution in semi-analytical form subject to solving a system of ODEs. Further assuming time-independent coefficients, the solution to the aforementioned system of ODEs can be obtained in closed form, yielding closed-form expressions for the optimal trading rate and the optimal posterior distribution for market trading rate.

The following lemma is crucial in the calculations involved in reducing the rHJI equations into effective Hamilton–Jacobi equations and for the proof of duality in section 5.2.

LEMMA 4.1 *For a distribution π satisfying $\pi \ll \pi_t^0$, consider the functional*

$$F[\pi] := \int \left\{\frac{C_2}{2}a^2 + C_1 a + \frac{1}{\beta}\log\frac{\pi(a)}{\pi_t^0(a)}\right\}\pi(a)\,\mathrm{d}a.$$

where C_1 and C_2 are constants such that $s_t + \beta C_2 > 0$. Then, F is convex in π and its minimum is achieved uniquely at the distribution π^* whose probability density function is given by

$$\pi_t^*(a) = \sqrt{\frac{s_t + \beta C_2}{2\pi}}\mathrm{e}^{-\frac{(s_t+\beta C_2)(a-w^*)^2}{2}}, \quad w^* = \frac{s_t m_t - \beta C_1}{s_t + \beta C_2}, \tag{11}$$

i.e. π^* is a normal density with mean w^* in (11) *and precision $s_t + \beta C_2$. The minimum value is given by*

$$F[\pi^*] = \frac{1}{2\beta}\log\frac{s_t + \beta C_2}{s_t} + \frac{s_t m_t^2}{2\beta} - \frac{1}{2\beta}\frac{(s_t m_t - \beta C_1)^2}{s_t + \beta C_2}. \tag{12}$$

Proof We first show that F is convex. For any given distributions $\pi_1, \pi_2 \ll \pi_t^0$ and $\lambda \in [0, 1]$, let $\pi = \lambda\pi_1 + (1-\lambda)\pi_2$. We have

$$
\begin{aligned}
F[\pi] &= F[\lambda\pi_1 + (1-\lambda)\pi_2] \\
&= \int \left\{ \frac{C_2}{2}a^2 + C_1 a + \frac{1}{\beta}\log\frac{\pi(a)}{\pi_t^0(a)} \right\} \pi(a)\,\mathrm{d}a \\
&= \int \left\{ \frac{C_2}{2}a^2 + C_1 a \right\} \{\lambda\pi_1(a) + (1-\lambda)\pi_2(a)\}\,\mathrm{d}a \\
&\quad + \int \frac{1}{\beta}\frac{\pi(a)}{\pi_t^0(a)}\log\frac{\pi(a)}{\pi_t^0(a)}\pi_t^0(a)\,\mathrm{d}a \\
&\le \lambda \int \left\{ \frac{C_2}{2}a^2 + C_1 a \right\} \pi_1(a)\,\mathrm{d}a \\
&\quad + (1-\lambda)\int \left\{ \frac{C_2}{2}a^2 + C_1 a \right\} \pi_2(a)\,\mathrm{d}a \\
&\quad + \lambda \int \frac{1}{\beta}\frac{\pi_1(a)}{\pi_t^0(a)}\log\frac{\pi_1(a)}{\pi_t^0(a)}\pi_t^0(a)\,\mathrm{d}a \\
&\quad + (1-\lambda)\int \frac{1}{\beta}\frac{\pi_2(a)}{\pi_t^0(a)}\log\frac{\pi_2(a)}{\pi_t^0(a)}\pi_t^0(a)\,\mathrm{d}a \\
&= \lambda F[\pi_1] + (1-\lambda)F[\pi_2],
\end{aligned}
$$

where the inequality results from the fact that the function $x\log x$ is convex. We thus conclude that F is convex.

As for the minimum value, note that for any $\pi \ll \pi_t^0$ we have

$$
\begin{aligned}
&\int \left\{ \frac{C_2}{2}a^2 + C_1 a + \frac{1}{\beta}\log\frac{\pi(a)}{\pi_t^0(a)} \right\} \pi(a)\,\mathrm{d}a \\
&\quad = \int \left\{ \frac{C_2}{2}a^2 + C_1 a + \frac{1}{\beta}\log\frac{\pi^*(a)}{\pi_t^0(a)} \right\} \pi(a)\,\mathrm{d}a \\
&\quad\quad + \frac{1}{\beta}\int \log\frac{\pi(a)}{\pi^*(a)}\pi(a)\,\mathrm{d}a \\
&\quad \ge \int \left\{ \frac{C_2}{2}a^2 + C_1 a + \frac{1}{\beta}\log\frac{\pi^*(a)}{\pi_t^0(a)} \right\} \pi(a)\,\mathrm{d}a \\
&\quad = \left[C_1 + \frac{s_t + \beta C_2}{\beta}w^* - \frac{s_t m_t}{\beta} \right] \langle a\rangle_t \\
&\quad\quad + \frac{1}{2\beta}\log\frac{s_t+\beta C_2}{s_t} + \frac{s_t m_t^2}{2\beta} \\
&\quad\quad - \frac{s_t+\beta C_2}{2\beta}(w^*)^2 \\
&\quad = \frac{1}{2\beta}\log\frac{s_t+\beta C_2}{s_t} + \frac{s_t m_t^2}{2\beta} - \frac{1}{2\beta}\frac{(s_t m_t - \beta C_1)^2}{s_t+\beta C_2} \\
&\quad = \int \left\{ \frac{C_2}{2}a^2 + C_1 a + \frac{1}{\beta}\log\frac{\pi_t^*(a)}{\pi_t^0(a)} \right\} \pi_t^*(a)\,\mathrm{d}a,
\end{aligned}
$$

where the inequality results from the positivity of relative entropy. Thus, the minimum value of is achieved at π^* in (11) and the minimum value is given by (12). ■

LEMMA 4.2 *In Model* 1, *the minimal value for the inner minimization in* (9) *is equal to*

$$
\begin{aligned}
&V_t + \frac{(\sigma^X)^2}{2}V_{xx} + vV_x + \frac{\gamma}{2}(\sigma^X)^2 + \rho\sigma^S\sigma^X - \eta v^2 \\
&\quad - \frac{1}{2\beta}\log\left(\frac{s_t}{s_t+\beta R_{aa}}\right) \\
&\quad + \frac{1}{2}\left(R_{xx} - \frac{\beta(\gamma_M + R_{xa})^2}{s_t+\beta R_{aa}}\right)x^2 + \frac{(\gamma_M + R_{xa})s_t m_t}{s_t+\beta R_{aa}}x \\
&\quad + \frac{s_t m_t^2 R_{aa}}{2(s_t+\beta R_{aa})}.
\end{aligned} \tag{13}
$$

Proof For any given trading rate v_t and function V, the first-order criterion via variational calculus applied to the inner minimization in (9), see also (A3) and (A5) in the appendix, yields that the candidate minimizer π_t^* for the minimization in the rHJI equation (9) is given by

$$
\begin{aligned}
\pi_t^*(a) &= \sqrt{\frac{s_t+\beta R_{aa}}{2\pi}}\,\mathrm{e}^{-\frac{(s_t+\beta R_{aa})(a-w^*)^2}{2}}, \\
w^* &= \frac{-\beta(\gamma_M + R_{xa})x + s_t m_t}{s_t+\beta R_{aa}},
\end{aligned} \tag{14}
$$

since $s_t + \beta R_{aa} > 0$, and its corresponding value is equal to

$$
-\frac{1}{\beta}\log\int \pi_t^0(a\,|\,x)\,\mathrm{e}^{-\beta[V_t+\mathcal{L}V+\tilde{R}(t,x,a)]}\,\mathrm{d}a. \tag{15}
$$

We evaluate the integral in (15) as follows.

$$
\begin{aligned}
&\int \pi_t^0(a\,|\,x)\,\mathrm{e}^{-\beta\{V_t+\mathcal{L}V+\tilde{R}(t,x,a)\}}\,\mathrm{d}a \\
&\quad = \int \mathrm{e}^{-\beta\left\{V_t+\mathcal{L}V+\frac{\gamma}{2}(\sigma^X)^2+\rho\sigma^S\sigma^X-\eta v^2+\frac{1}{2}R_{xx}x^2+(\gamma_M+R_{xa})xa+\frac{1}{2}R_{aa}a^2\right\}} \\
&\quad\quad \times \sqrt{\frac{s_t}{2\pi}}\mathrm{e}^{-\frac{s_t}{2}(a-m_t)^2}\,\mathrm{d}a \\
&\quad = \mathrm{e}^{-\beta\left\{V_t+\frac{(\sigma^X)^2}{2}V_{xx}+vV_x+\frac{\gamma}{2}(\sigma^X)^2+\rho\sigma^S\sigma^X-\eta v^2+\frac{1}{2}R_{xx}x^2\right\}} \\
&\quad\quad \times \int \mathrm{e}^{-\beta\left\{(\gamma_M+R_{xa})xa+\frac{1}{2}R_{aa}a^2\right\}}\sqrt{\frac{s_t}{2\pi}}\mathrm{e}^{-\frac{s_t}{2}(a-m_t)^2}\,\mathrm{d}a \\
&\quad = \mathrm{e}^{-\beta\left\{V_t+\frac{(\sigma^X)^2}{2}V_{xx}+vV_x+\frac{\gamma}{2}(\sigma^X)^2+\rho\sigma^S\sigma^X-\eta v^2\right\}} \\
&\quad\quad \times \sqrt{\frac{s_t}{s_t+\beta R_{aa}}} \\
&\quad\quad \times \mathrm{e}^{-\frac{\beta}{2}\left\{R_{xx}+\frac{-\beta(\gamma_M+R_{xa})^2}{s_t+\beta R_{aa}}\right\}x^2+\frac{-\beta(\gamma_M+R_{xa})s_t m_t}{s_t+\beta R_{aa}}x-\frac{\beta s_t m_t^2 R_{aa}}{2(s_t+\beta R_{aa})}}.
\end{aligned}
$$

Hence, by taking logarithm of the last equation and multiplying the resulting expression by the factor $-\frac{1}{\beta}$, we get the desired quantity.

The π_t^* in (14) is indeed a minimizer for the minimization in (9) since, for any admissible distribution π_t, the objective function, after omitting all the terms independent of a, equals

$$
\int \left\{ (\gamma_M + R_{xa})xa + \frac{R_{aa}}{2}a^2 + \frac{1}{\beta}\log\frac{\pi_t(a)}{\pi_t^0(a)} \right\} \pi_t(a)\,\mathrm{d}a.
$$

Hence, with $C_2 = R_{aa}$ and $C_1 = (\gamma_M + R_{xa})x$, lemma 4.1 implies that π_t^* is the unique minimizer. ■

Next, we move on to deal with the outer maximization in (9) and show that it can be further reduced to the Hamilton–Jacobi equation (16).

LEMMA 4.3 *The rHJI equation* (9) *is reduced to the following Hamilton–Jacobi equation*

$$V_t + \frac{(\sigma^X)^2}{2}V_{xx} + \frac{V_x^2}{4\eta} - \frac{1}{2\beta}\log\left(\frac{s_t}{s_t+\beta R_{aa}}\right) + \frac{\gamma}{2}(\sigma^X)^2 + \rho\sigma^S\sigma^X + \frac{1}{2}\left\{R_{xx} - \frac{\beta(\gamma_M+R_{xa})^2}{s_t+\beta R_{aa}}\right\}x^2 + \frac{(\gamma_M+R_{xa})s_t m_t}{s_t+\beta R_{aa}}x + \frac{s_t m_t^2 R_{aa}}{2(s_t+\beta R_{aa})} = 0 \tag{16}$$

with terminal condition $V(T,x) = -gx^2$, *where* $g = \delta - \frac{\gamma}{2}$.

Proof Lemma 4.2 showed that the value of the inner minimization in (9) is equal to (13). Thus, (9) reduces to

$$\max_v \left\{V_t + \frac{(\sigma^X)^2}{2}V_{xx} + vV_x + \frac{\gamma}{2}(\sigma^X)^2 + \rho\sigma^S\sigma^X - \eta v^2 - \frac{1}{2\beta}\log\left(\frac{s_t}{s_t+\beta R_{aa}}\right) + \frac{1}{2}\left(R_{xx} - \frac{\beta(\gamma_M+R_{xa})^2}{s_t+\beta R_{aa}}\right)x^2 + \frac{(\gamma_M+R_{xa})s_t m_t}{s_t+\beta R_{aa}}x + \frac{s_t m_t^2 R_{aa}}{2(s_t+\beta R_{aa})}\right\} = 0. \tag{17}$$

The expression between the brackets is quadratic and concave in v, it admits a unique maximal value at the vertex. The first-order criterion thus implies that the maximizer v^* is given by $v^* = \frac{V_x}{2\eta}$. The proof is then completed by plugging the maximizer v^* into (17). ■

THEOREM 4.1 *In Model* 1, *the value function V for the control problem* (7) *is quadratic in x, sa,y* $V(t,x) = \frac{1}{2}H_2x^2 + H_1x + H_0$, *where the time-dependent coefficients* H_2, H_1, *and* H_0 *satisfy the following system of equations*

$$\dot{H}_2 + \frac{H_2^2}{2\eta} + A_1 = 0,$$

$$\dot{H}_1 + \frac{H_1H_2}{2\eta} + B_1 = 0,$$

$$\dot{H}_0 + \frac{(\sigma^X)^2}{2}H_2 + \frac{H_1^2}{4\eta} + \frac{\gamma}{2}(\sigma^X)^2 + \rho\sigma^S\sigma^X - \frac{1}{2\beta}\log c - \frac{s_t}{2\beta}m_t^2(c-1) = 0$$

with terminal conditions $H_2(T) = -2g$ *and* $H_1(T) = H_0(T) = 0$. A_1, B_1, *and c are defined by*

$$A_1 = R_{xx} - \frac{\beta(\gamma_M+R_{xa})^2}{s_t+\beta R_{aa}}, \quad B_1 = \frac{(\gamma_M+R_{xa})s_t m_t}{s_t+\beta R_{aa}}, \quad c = \frac{s_t}{s_t+\beta R_{aa}}.$$

Proof Assume the ansatz for value function

$$V(t,x) = \tfrac{1}{2}H_2x^2 + H_1x + H_0,$$

where the H_i's are functions of t. Substitute the ansatz for value function to the Hamilton–Jacobi equation (16) then comparing the coefficients yield the following system of ODEs satisfied by the H_i's

$$x^2 : \dot{H}_2 + \frac{H_2^2}{2\eta} + A_1 = 0,$$

$$x : \dot{H}_1 + \frac{H_1H_2}{2\eta} + B_1 = 0,$$

$$1 : \dot{H}_0 + \frac{(\sigma^X)^2}{2}H_2 + \frac{H_1^2}{4\eta} + \frac{\gamma}{2}(\sigma^X)^2 + \rho\sigma^S\sigma^X - \frac{1}{2\beta}\log c - \frac{s_t}{2\beta}m_t^2(c-1) = 0.$$

■

We note that, since the coefficients in the system of ODEs in theorem 4.1 may be time dependent, in general it does not admit solution in closed form. However, if we further assume the coefficients being constant, the system of ODEs does admit a solution in closed form. We summarize the result in the following corollary.

COROLLARY 4.1 *Further assume that all the coefficients in the system of ODEs in theorem* 4.1 *are constant and let*

$$\hat{A}_1 = \sqrt{\frac{-A_1}{2\eta}},$$

assuming $A_1 < 0$. *The functions* H_2 *and* H_1 *have the following closed-form expressions*

$$H_2(t) = -2\eta\hat{A}_1\coth(\hat{A}_1\{T-t\}+\alpha_1),$$

$$H_1(t) = -\frac{B_1}{\hat{A}_1}\frac{\cosh\alpha_1}{\sinh(\hat{A}_1\{T-t\}+\alpha_1)} + \frac{B_1}{\hat{A}_1}\coth(\hat{A}_1\{T-t\}+\alpha_1),$$

where $\alpha_1 = \coth^{-1}\left(\frac{g}{\eta\hat{A}_1}\right)$. *It follows that the optimal trading strategy* v_t^* *is obtained by*

$$v_t^* = \frac{1}{2\eta}(H_2x_t + H_1) = -\hat{A}_1\coth(\hat{A}_1\{T-t\}+\alpha_1)x_t - \frac{B_1}{2\eta\hat{A}_1}\frac{\cosh(\alpha_1) - \cosh(\hat{A}_1\{T-t\}+\alpha_1)}{\sinh(\hat{A}_1\{T-t\}+\alpha_1)} \tag{18}$$

and the expected optimal trading trajectory x_t^* *by*

$$x_t^* = \frac{\sinh(\hat{A}_1\{T-t\}+\alpha_1)}{\sinh(\hat{A}_1T+\alpha_1)}x_0 + \sinh(\hat{A}_1\{T-t\}+\alpha_1)\int_0^t \frac{H_1(s)}{2\eta\sinh(\hat{A}_1\{T-s\}+\alpha_1)}\,\mathrm{d}s. \tag{19}$$

We remark that, if the prior mean m_t is identically zero, then $B_1 = 0$. It follows that the function H_1 is also identically zero. In this case, the expected optimal trajectory in (19) reduces to

$$x_t^* = \frac{\sinh(\hat{A}_1\{T-t\}+\alpha_1)}{\sinh(\hat{A}_1 T+\alpha_1)} X \tag{20}$$

which resembles the Almgren–Chriss strategy except that the proxy of market trading rate precision s_t is naturally embedded in the parameter $\hat{A}_1$. Moreover, the limiting behavior of v_t^* as $g \to \infty$ and $A_1 \to 0$ is obtained in the following corollary.

COROLLARY 4.2 *Assume $m_t = 0$. As $g \to \infty$ and $A_1 \to 0$, the optimal execution strategy in feedback from* (18) *converges to the adapted TWAP strategy defined as*

$$v_t^* = -\frac{x_t}{T-t}. \tag{21}$$

A sufficient condition for $g \to \infty$ is $\delta \to \infty$, which means the final penalty on the inventory is extremely high. A sufficient condition for $A_1 \to 0$ is $R_{xx} \to 0$ and $s_t \to \infty$, which corresponds, respectively, to zero quadratic term in x for risk and the prior distribution being a Dirac measure concentrated at m_t.

Proof If $m_t = 0$, then $B_1 = 0$. It thus suffices to show that

$$\lim_{g\to\infty,\hat{A}_1\to 0} \hat{A}_1 \coth(\hat{A}_1\{T-t\}+\alpha_1) = \frac{1}{T-t}.$$

Indeed,

$$\begin{aligned}
&\lim_{g\to\infty,\hat{A}_1\to 0} \hat{A}_1 \coth(\hat{A}_1\{T-t\}+\alpha_1)\\
&= \lim_{g\to\infty,\hat{A}_1\to 0} \hat{A}_1 \frac{1+\coth(\hat{A}_1\{T-t\})\frac{g}{\eta\hat{A}_1}}{\coth(\hat{A}_1\{T-t\})+\frac{g}{\eta\hat{A}_1}}\\
&= \lim_{g\to\infty,\hat{A}_1\to 0} \frac{\frac{\eta\hat{A}_1^2}{g}+\hat{A}_1\coth(\hat{A}_1\{T-t\})}{\frac{\eta}{g}\hat{A}_1\coth(\hat{A}_1\{T-t\})+1}.
\end{aligned}$$

Since

$$\lim_{\hat{A}_1\to 0} \hat{A}_1 \coth(\hat{A}_1\{T-t\}) = \frac{1}{T-t},$$

we reach the conclusion. ■

4.2. Model 2

In this subsection, as for model 1 in section 4.1, we show that under Assumptions (A1) and (A3), which we shall refer to as *model 2* hereafter, the rHJI equation (9) is reduced to the effective Hamilton–Jacobi equation (22) that admits a solution in semi-analytical form subject to solving a system of ODEs. We also obtain closed-form expressions for the solution in the case of time-independent coefficients, yielding closed-form expressions for the optimal trading rate and the optimal posterior distribution for market trading rate.

The following lemma shows how to reduce the rHJI equation (9) to the effective Hamilton–Jacobi equation (22).

LEMMA 4.4 *In Model 2, the rHJI equation* (9) *is reduced to the following effective Hamilton–Jacobi equation.*

$$\begin{aligned}
V_t &+ \frac{(\sigma^X)^2}{2}V_{xx} + \frac{1}{4}\frac{(V_x + \frac{R_{va}s_t m_t}{s_t+\beta R_{aa}} - \frac{\beta\gamma_M R_{va}}{s_t+\beta R_{aa}}x)^2}{\eta - \frac{R_{vv}}{2} + \frac{\beta R_{va}^2}{2(s_t+\beta R_{aa})}}\\
&- \frac{\beta\gamma_M^2}{2(s_t+\beta R_{aa})}x^2 + \frac{s_t}{2\beta}m_t^2\left(1 - \frac{s_t}{s_t+\beta R_{aa}}\right)\\
&- \frac{1}{2\beta}\log\left(\frac{s_t}{s_t+\beta R_{aa}}\right)\\
&+ \frac{\gamma_M s_t m_t}{s_t+\beta R_{aa}}x + \frac{\gamma}{2}(\sigma^X)^2 + \rho\sigma^S\sigma^X = 0.
\end{aligned} \tag{22}$$

with terminal condition $V(T,x) = -gx^2$, where $g = \delta - \frac{\gamma}{2}$.

Proof Since the proof of the lemma is almost parallel to that of lemmas 4.2 and 4.3 for model 1, we show only the essential calculations for deriving the effective Hamilton–Jacobi equation (22). The first order criterion via variational calculus applied to the inner minimization in (9) yields that the minimizer π_t^* for the inner minimization in the rHJI equation (9) is given by

$$\begin{aligned}
\pi_t^*(a) &= \sqrt{\frac{s_t+\beta R_{aa}}{2\pi}}\,\mathrm{e}^{-\frac{s_t+\beta R_{aa}}{2}(a-w^*)^2},\\
w^* &= \frac{-\beta R_{va}v - \beta\gamma_M x + s_t m_t}{s_t+\beta R_{aa}},
\end{aligned}$$

since $s_t + \beta R_{aa} > 0$, and its corresponding value by

$$-\frac{1}{\beta}\log\int \pi_t^0(a\,|\,x)\,\mathrm{e}^{-\beta[V_t+\mathcal{L}V+\tilde{R}(t,x,s,v,a)]}\,\mathrm{d}a. \tag{23}$$

We now calculate the integral in (23) as follows. We have

$$\begin{aligned}
&\int \pi_t^0(a\,|\,x)\,\mathrm{e}^{-\beta\{V_t+\mathcal{L}V+\tilde{R}(t,x,s,v,a)\}}\,\mathrm{d}a\\
&= \int \mathrm{e}^{-\beta\left\{V_t+\mathcal{L}V+\frac{\gamma}{2}(\sigma^X)^2+\rho\sigma^S\sigma^X-\eta v^2+\gamma_M xa+\frac{1}{2}R_{vv}v^2+R_{va}va+\frac{1}{2}R_{aa}a^2\right\}}\\
&\quad\times\sqrt{\frac{s_t}{2\pi}}\,\mathrm{e}^{-\frac{s_t}{2}(a-m_t)^2}\,\mathrm{d}a\\
&= \mathrm{e}^{-\beta\left\{V_t+\frac{(\sigma^X)^2}{2}V_{xx}+vV_x+\frac{\gamma}{2}(\sigma^X)^2+\rho\sigma^S\sigma^X-\eta v^2+\frac{1}{2}R_{vv}v^2\right\}}\\
&\quad\times\int \mathrm{e}^{-\beta\left\{(\gamma_M x+R_{va}v)a+\frac{1}{2}R_{aa}a^2\right\}}\sqrt{\frac{s_t}{2\pi}}\,\mathrm{e}^{-\frac{s_t}{2}(a-m_t)^2}\,\mathrm{d}a\\
&= \mathrm{e}^{-\beta\left\{V_t+\frac{(\sigma^X)^2}{2}V_{xx}+vV_x+\frac{\gamma}{2}(\sigma^X)^2+\rho\sigma^S\sigma^X-\eta v^2+\frac{1}{2}R_{vv}v^2\right\}}\\
&\quad\times\sqrt{\frac{s_t}{s_t+\beta R_{aa}}}\,\mathrm{e}^{\frac{(s_t m_t-\beta[\gamma_M x+R_{va}v])^2}{2(s_t+\beta R_{aa})}-\frac{s_t}{2}m_t^2}.
\end{aligned}$$

Next, by taking logarithm of the last equation and multiplying the resulting expression by the factor $-\frac{1}{\beta}$, then substitute into (23), the rHJI equation reduces to the following HJB equation

$$\max_v \Bigg\{ V_t + \frac{(\sigma^X)^2}{2} V_{xx} + vV_x + \frac{\gamma}{2}(\sigma^X)^2 + \rho\sigma^S\sigma^X - \left(\eta - \frac{R_{vv}}{2}\right)v^2 - \frac{1}{2\beta}\log\left(\frac{s_t}{s_t+\beta R_{aa}}\right) - \frac{(s_t m_t - \beta\{\gamma_M x + R_{va} v\})^2}{2\beta(s_t+\beta R_{aa})} + \frac{s_t}{2\beta}m_t^2 \Bigg\} = 0. \quad (24)$$

The maximization problem in (24), after omitting the terms that are independent of v, reads

$$\max_{v\in\mathbb{R}} \left\{ -\left(\eta - \frac{R_{vv}}{2}\right)v^2 + vV_x - \frac{[\beta R_{va} v + \beta\gamma_M x - s_t m_t]^2}{2\beta(s_t+\beta R_{aa})} \right\}$$

Since the objective function in the maximization problem above is quadratic in v, its unique maximum is attended at

$$v = \frac{1}{2}\frac{V_x + \frac{R_{va}s_t m_t}{s_t+\beta R_{aa}} - \frac{\beta\gamma_M R_{va}}{s_t+\beta R_{aa}}x}{\eta - \frac{R_{vv}}{2} + \frac{\beta R_{va}^2}{2(s_t+\beta R_{aa})}} \quad (25)$$

since $\eta - \frac{R_{vv}}{2} + \frac{\beta R_{va}^2}{2(s_t+\beta R_{aa})} > 0$ by assumption.

Finally, by substituting the maximizer (25) into the HJB equation (24), we obtain the effective Hamilton–Jacobi equation given in (22), which completes the proof. ■

Following almost the same procedure as for model 1 in section 4.1, one can show that the value function in this case is also quadratic in x and the time-dependent coefficients satisfy similar Riccati equations as in theorem 4.1. We summarize the result in the following theorem but omit its proof.

THEOREM 4.2 *In Model* 2, *the value function* V *for the control problem* (7) *is quadratic in x, say,* $V(t,x) = \frac{1}{2}H_2x^2 + H_1x + H_0$, *where the time-dependent functions* H_2 *and* H_1 *satisfy the following system of ODEs*

$$\dot{H}_2 + \frac{1}{2\tilde{\eta}}\left(H_2 - \frac{\beta\gamma_M R_{va}}{s_t+\beta R_{aa}}\right)^2 - A_2 = 0,$$
$$\dot{H}_1 + \frac{1}{2\tilde{\eta}}\left(H_2 - \frac{\beta\gamma_M R_{va}}{s_t+\beta R_{aa}}\right)\left(H_1 + \frac{R_{va}s_t m_t}{s_t+\beta R_{aa}}\right) + B_2 = 0.$$

with terminal condition $H_2(T) = -2g$ *and* $H_1(T) = 0$. A_2, B_2, *and* $\tilde{\eta}$ *are given by*

$$A_2 = \frac{\beta\gamma_M^2}{s_t+\beta R_{aa}}, \quad B_2 = \frac{\gamma_M s_t m_t}{s_t+\beta R_{aa}},$$
$$\tilde{\eta} = \eta - \frac{R_{vv}}{2} + \frac{\beta R_{va}^2}{2(s_t+\beta R_{aa})},$$

where $s_t + \beta R_{aa} > 0$ *and* $\tilde{\eta} > 0$ *by assumption.*

Finally, as in model 1, if we further assume the coefficients are constant, we obtain closed-form expressions for the functions H_2 and H_1, thus closed-form expressions for optimal trading rate and expected optimal trading strategy as well. We summarize the result in the following corollary without proof.

COROLLARY 4.3 *Assume* $\tilde{\eta} > 0$ *and that all the coefficients are constant. Let*

$$\hat{A}_2 = \sqrt{\frac{A_2}{2\tilde{\eta}}}.$$

The functions H_2 *and* H_1 *admit the following closed-form expressions*

$$H_2(t) = -2\tilde{\eta}\hat{A}_2\coth(\hat{A}_2\{T-t\}+\alpha_2) + C,$$
$$H_1(t) = -\frac{D\hat{A}_2\sinh\alpha_2 + B_2\cosh\alpha_2}{\hat{A}_2\sinh(\hat{A}_2\{T-t\}+\alpha_2)} + \frac{B_2}{\hat{A}_2}\coth(\hat{A}_2\{T-t\}+\alpha_2) + D,$$

where $\alpha_2 = \coth^{-1}\left(\frac{2g+C}{2\tilde{\eta}\hat{A}_2}\right)$, $C = \frac{\beta\gamma_M R_{va}}{s_t+\beta R_{aa}}$, *and* $D = -\frac{R_{va}s_t m_t}{s_t+\beta R_{aa}}$. *It follows that the optimal trading strategy* v_t^* *is obtained by*

$$v_t^* = \frac{1}{2\tilde{\eta}}(H_2x_t + H_1 - Cx_t - D)$$
$$= -\hat{A}_2\coth(\hat{A}_2\{T-t\}+\alpha_2)x_t - \frac{D\hat{A}_2\sinh\alpha_2 + B_2\cosh\alpha_2}{2\tilde{\eta}\hat{A}_2\sinh(\hat{A}_2\{T-t\}+\alpha_2)} + \frac{B_2}{2\tilde{\eta}\hat{A}_2}\coth(\hat{A}_2\{T-t\}+\alpha_2).$$

We remark that, if the prior mean m_t is zero, then $B_2 = D = 0$, the optimal trading rate v_t^* reduces to

$$v_t^* = -\hat{A}_2\coth(\hat{A}_2\{T-t\}+\alpha_2)x_t. \quad (26)$$

In this case, the expected optimal trading trajectory x_t^* reads

$$x_t^* = \frac{\sinh(\hat{A}_2\{T-t\}+\alpha_2)}{\sinh(\hat{A}_2T+\alpha_2)}X, \quad (27)$$

which again resembles the Almgren–Chriss strategy but the proxy of market trading rate precision s_t is naturally embedded in the parameter $\hat{A}_2$. Finally, as in model 1, we state the limiting behavior of v_t^* as as $g \to \infty$ and $A_2 \to 0$ but omit its proof.

COROLLARY 4.4 *Assume* $m_t = 0$, *the optimal execution strategy converges to the adapted TWAP strategy* (21) *as* $g \to \infty$ *and* $A_2 \to 0$. *A sufficient condition for* $g \to \infty$ *is* $\delta \to \infty$, *which means a final block trade at terminal time is strictly prohibited. A sufficient condition for* $A_2 \to 0$ *is* $s_t \to \infty$, *which corresponds to the prior distribution being a Dirac measure concentrated at* m_t.

5. Verification theorem and duality

5.1. Verification theorem

We prove the verification theorems for the relative entropy-regularized stochastic differential game (7) in this subsection.

The argument is standard, which relies mainly on the saddle point property for a max-min problem. In fact, the argument applies not only to our models but to general problems, subject to certain regularity conditions.

THEOREM 5.1 *Let $V(t,x,s)$ be the unique smooth solution to the following rHJI equation*

$$\max_{v} \min_{\pi \ll \pi_t^0} \int \left\{ V_t + \mathcal{L}V + \tilde{R}(t,x,s,v,a) + \frac{1}{\beta} \log \frac{\pi(a)}{\pi_t^0(a)} \right\} \times \pi(a)\,\mathrm{d}a = 0 \tag{28}$$

with terminal condition $V(T,x,s) = G(x,s)$. Then V is equal to the value function U of the following maximization-minimization problem

$$U(t,x,s) = \max_{v\in\mathcal{V}_t} \min_{\pi\in\mathcal{A}_t} \mathbb{E}_t \Bigg[G(X_T,S_T) + \iint_t^T \left\{ \tilde{R}(\tau,X_\tau,S_\tau,v_\tau,a_\tau) + \frac{1}{\beta}\log\frac{\pi_\tau(a)}{\pi_\tau^0(a)} \right\} \times \pi_\tau(a)\,\mathrm{d}a\,\mathrm{d}\tau \Bigg],$$

where recall that $\mathbb{E}_t[\cdot]$ denotes the conditional expectation $\mathbb{E}[\cdot|\mathcal{F}_t]$ and $(X_t,S_t) = (x,s)$.

Proof For a given $t \in (0,T)$, consider the time interval (t,T). Since V satisfies the max-min equation (28), we have that, for any given admissible control $v \in \mathcal{V}_t$ in the interval $[t,T]$, there exists an admissible $\pi \in \mathcal{A}_t$ such that

$$\int \left\{ V_t + \mathcal{L}V + \tilde{R}(\tau,x,s,v_\tau,a) + \frac{1}{\beta}\log\frac{\pi_\tau(a)}{\pi_\tau^0(a)} \right\} \times \pi_\tau(a)\,\mathrm{d}a \le 0$$

for every $\tau \in (t,T)$. Applying Ito's formula to V then taking conditional expectation $\mathbb{E}_t[\cdot]$ yields

$$\begin{aligned} &\mathbb{E}_t[G(X_T,S_T) - V(t,X_t,S_t)] = \mathbb{E}_t[V(T,X_T,S_T)] - V(t,x,s) \\ &= \int_t^T \mathbb{E}_t[V_\tau + \mathcal{L}V]\,\mathrm{d}\tau \\ &\le -\int_t^T \mathbb{E}_t\Bigg[\int \left\{ \tilde{R}(\tau,X_\tau,S_\tau,v_\tau,a_\tau) + \frac{1}{\beta}\log\frac{\pi_\tau(a)}{\pi_\tau^0(a)} \right\} \times \pi_\tau(a)\,\mathrm{d}a\Bigg]\mathrm{d}\tau. \end{aligned}$$

It follows that

$$\begin{aligned} &V(t,x,s) \\ &\ge \mathbb{E}_t\Bigg[G(X_T,S_T) + \iint_t^T \Big\{ \tilde{R}(\tau,X_\tau,S_\tau,v_\tau,a_\tau) + \frac{1}{\beta}\log\frac{\pi_\tau(a)}{\pi_\tau^0(a)} \Big\} \pi_\tau(a)\,\mathrm{d}a\,\mathrm{d}\tau\Bigg] \\ &\ge \min_{\pi\in\mathcal{A}_t} \mathbb{E}_t\Bigg[G(X_T,S_T) + \iint_t^T \Big\{ \tilde{R}(\tau,X_\tau,S_\tau,v_\tau,a_\tau) + \frac{1}{\beta}\log\frac{\pi_\tau(a)}{\pi_\tau^0(a)} \Big\} \pi_\tau(a)\,\mathrm{d}a\,\mathrm{d}\tau\Bigg] \end{aligned}$$

Since v is arbitrary, we end up

$$\begin{aligned} &V(t,x,s) \\ &\ge \max_{v\in\mathcal{V}_t} \min_{\pi\in\mathcal{A}_t} \mathbb{E}_t\Bigg[G(X_T,S_T) + \iint_t^T \Big\{ \tilde{R}(\tau,X_\tau,S_\tau,v_\tau,a_\tau) + \frac{1}{\beta}\log\frac{\pi_\tau(a)}{\pi_\tau^0(a)} \Big\} \pi_\tau(a)\,\mathrm{d}a\,\mathrm{d}\tau\Bigg] \\ &= U(t,x,s). \end{aligned}$$

On the other hand, from the max-min equation (28), we also have that, for any $\epsilon > 0$, there exists a $v \in \mathcal{V}_t$ such that

$$\int \left\{ V_t + \mathcal{L}V + \tilde{R}(\tau,x,s,v_\tau,a) + \frac{1}{\beta}\log\frac{\pi_\tau(a)}{\pi_\tau^0(a)} \right\} \times \pi_\tau(a)\,\mathrm{d}a > -\epsilon.$$

for $\tau \in (t,T)$ and all $\pi \in \mathcal{A}_t$. Applying Ito's formula to V and taking conditional expectation $\mathbb{E}_t[\cdot]$ yields

$$\begin{aligned} &\mathbb{E}_t[G(X_T,S_T) - V(t,X_t,S_t)] = \mathbb{E}_t[V(T,X_T,S_T)] - V(t,x,s) \\ &= \int_t^T \mathbb{E}_t[V_\tau + \mathcal{L}V]\,\mathrm{d}\tau \\ &\ge -\int_t^T \mathbb{E}_t\Bigg[\int \left\{ \tilde{R}(\tau,X_\tau,S_\tau,v_\tau,a_\tau) + \frac{1}{\beta}\log\frac{\pi_\tau(a)}{\pi_\tau^0(a)} \right\} \times \pi_\tau(a)\,\mathrm{d}a\Bigg]\mathrm{d}\tau - \epsilon(T-t) \end{aligned}$$

It follows that

$$\begin{aligned} &V(t,x,s) \\ &\le \mathbb{E}_t\Bigg[G(X_T,S_T) + \iint_t^T \Big\{ \tilde{R}(\tau,X_\tau,S_\tau,v_\tau,a_\tau) + \frac{1}{\beta}\log\frac{\pi_\tau(a)}{\pi_\tau^0(a)} \Big\} \pi_\tau(a)\,\mathrm{d}a\,\mathrm{d}\tau\Bigg] + \epsilon(T-t) \end{aligned}$$

for all $\pi \in \mathcal{A}_t$. We end up

$$\begin{aligned} &V(t,x,s) \\ &\le \min_{\pi\in\mathcal{A}_t} \mathbb{E}_t\Bigg[G(X_T,S_T) + \iint_t^T \left\{ \tilde{R}(\tau,X_\tau,S_\tau,v_\tau,a_\tau) + \frac{1}{\beta}\log\frac{\pi_\tau(a)}{\pi_\tau^0(a)} \right\} \times \pi_\tau(a)\,\mathrm{d}a\,\mathrm{d}\tau\Bigg] + \epsilon(T-t) \\ &\le \max_{v\in\mathcal{V}_t} \min_{\pi\in\mathcal{A}_t} \mathbb{E}_t\Bigg[G(X_T,S_T) + \iint_t^T \left\{ \tilde{R}(\tau,X_\tau,S_\tau,v_\tau,a_\tau) + \frac{1}{\beta}\log\frac{\pi_\tau(a)}{\pi_\tau^0(a)} \right\} \times \pi_\tau(a)\,\mathrm{d}a\,\mathrm{d}\tau\Bigg] + \epsilon(T-t) \end{aligned}$$

$$= U(t,x,s) + \epsilon(T-t).$$

We conclude that, since ϵ is arbitrary, $V(t,x,s) \le U(t,x,s)$. This completes the proof. ■

Finally, we remark that, by the same token, one can also prove the verification theorem for the 'dual' problem, which we summarize in the following theorem without proof.

THEOREM 5.2 *Let $V(t,x,s)$ be the unique smooth solution to the following rHJI equation*

$$\min_{\pi \ll \pi_t^0} \max_{v} \int \left\{ V_t + \mathcal{L}V + \tilde{R}(t,x,s,v,a) + \frac{1}{\beta} \log \frac{\pi(a)}{\pi_t^0(a)} \right\} \times \pi(a)\, \mathrm{d}a = 0 \tag{29}$$

with terminal condition $V(T,x,s) = G(x,s)$. Then V is equal to the value function U of the following minimization-maximization problem

$$U(t,x,s) = \min_{\pi \in \mathcal{A}_t} \max_{v \in \mathcal{V}_t} \mathbb{E}_t \left[G(X_T, S_T) + \int_t^T \int \left\{ \tilde{R}(\tau, X_\tau, S_\tau, v_\tau, a_\tau) + \frac{1}{\beta} \log \frac{\pi_\tau(a)}{\pi_\tau^0(a)} \right\} \times \pi_\tau(a)\, \mathrm{d}a\, \mathrm{d}\tau \right]. \tag{30}$$

5.2. Duality

The problem (7) is referred to as the primal problem and the problem (30) as the dual problem. In this section, we prove that, for model 1 and model 2 considered, respectively, in sections 4.1 and 4.2, with an additional assumption for model 2, the duality between the primal problem (7) and the dual problem (30) holds. We hence conclude that, in the terminology of game theory, the stochastic differential game (7) is said to admit a value function as shown in Fleming and Souganidis (1989) for the classical case.

THEOREM 5.3 *Under Model 1 and Model 2, for Model 2, further assume that $\eta > \frac{R_{vv}}{2}$, the primal problem (7) and its dual (30) satisfy the strong duality, i.e. the two problems have the same value.*

Proof It suffices to show that the Hamiltonian in the rHJI equations for the primal (7) and the dual (30) satisfies the saddle point property. Recall that the Hamiltonian H_1 in model 1 for the primal and the dual, after disregarding all the terms that are not dependent on v or a and π, is given by

$$H_1(t,x,v,\pi) := \int \left\{ -\eta v^2 + vV_x + (R_{xa} + \gamma_M) xa + \frac{R_{aa}}{2} a^2 + \frac{1}{\beta} \log \frac{\pi_t(a)}{\pi_t^0(a)} \right\} \pi_t(a)\, \mathrm{d}a$$

for $v \in \mathbb{R}$ and $\pi \ll \pi_t^0$. Notice that the v dependence and the (a,π) dependence in H_1 are separated. It follows that the Hamiltonian in model 1 satisfies automatically the saddle point property

$$\max_{v} \min_{\pi \ll \pi_t^0} H_1(t,x,v,\pi) = \min_{\pi \ll \pi_t^0} \max_{v} H_1(t,x,v,\pi)$$

for fixed t and x.

As for model 2. Note that, again by disregarding all the terms that are not dependent on v or on a and π, the Hamiltonian H_2 in model 2 for the primal and the dual is given, for $v \in \mathbb{R}$ and $\pi \ll \pi_t^0$, by

$$H_2(t,x,v,\pi) := \int \left\{ -\hat{\eta} v^2 + vV_x + R_{va} va + \frac{R_{aa}}{2} a^2 + \gamma_M xa + \frac{1}{\beta} \log \frac{\pi_t(a)}{\pi_t^0(a)} \right\} \pi_t(a)\, \mathrm{d}a \tag{31}$$

where $\hat{\eta} := \eta - \frac{R_{vv}}{2}$ and $\hat{\eta} > 0$ by assumption. Apparently, for any fixed t, x, π, H_2 is concave in v and convex in π (by lemma 4.1) for any fixed t, x, v, thus H_2 also satisfies the saddle point property

$$\max_{v} \min_{\pi \ll \pi_t^0} H_2(t,x,v,\pi) = \min_{\pi \ll \pi_t^0} \max_{v} H_2(t,x,v,\pi)$$

for fixed t and x. This completes the proof. ■

The duality theorem asserts that, under certain assumptions on model parameters, the relative entropy-regularized robust order execution problem, when regarded as a stochastic differential game, admits an equilibrium between the agent's optimal trading strategy and the market's optimal reaction per the principle of least relative entropy.

6. Extensions to non-constant volatility

We consider a possible extension of the model to include non-constant, but only time-dependent, volatility of the price dynamics in this section. Rather than being a constant, let the price volatility σ^S be a function of time defined by

$$\sigma_t^S = \sqrt{\int \tilde{\sigma}^2(a) \pi_t(a)\, \mathrm{d}a.} \tag{32}$$

In other words, the price volatility σ^S is given by the square root of the average, in density π_t, of the variance of market trading rate $\tilde{\sigma}^2$. We summarize the findings in the following theorem without proof.

THEOREM 6.1 *The value function V satisfies the following rHJI equation*

$$\max_{v} \min_{\pi \ll \pi_t^0} \left\{ V_t + \mathcal{L}V + \int \left\{ \tilde{R}(t,x,s,v,a) + \frac{1}{\beta} \log \frac{\pi(a)}{\pi_t^0(a)} \right\} \times \pi(a)\, \mathrm{d}a + \rho \sigma^X \sigma_t^S \right\} = 0,$$

with terminal condition $V(T,x,s)=G(x,s)$*, assuming enough regularity for the value function V, where*

$$\hat{R}(t,x,s,v,a)=\gamma_M xa+\frac{\gamma}{2}(\sigma^X)^2-\eta v^2+R(t,x,s,v,a).$$

$\mathcal{L}$ *denotes the infinitesimal generator for the SDEs*

$$\begin{aligned}\mathrm{d}S_t&=\gamma\,\mathrm{d}X_t+\gamma_M\langle a\rangle_t\,\mathrm{d}t+\sigma_t^S\,\mathrm{d}W_t^S,\\ \mathrm{d}X_t&=v_t\,\mathrm{d}t+\sigma^X\,\mathrm{d}W_t^X,\end{aligned}$$

where the Brownian motions W_t^S *and* W_t^X *are correlated with correlation* ρ*.*

When $\rho=0$ *and under the assumptions of Model 1 or Model 2, the optimal solution can be solved explicitly and it coincides with the optimal solution with constant volatility.*

7. Numerical examples

In this section, we conduct numerical experiments on the implementation of optimal strategies obtained in section 4 and stress testing the strategies against various parameters. For Models 1 and 2, Monte Carlo simulations are conducted to illustrate the performances of the optimal strategies versus those of related adapted TWAP strategies as in (21). We note that a TWAP strategy is meant to dice a meta order into child orders and gradually release them to the market using evenly divided time slots between the initial and terminal time. It follows that, in the continuous time limit, a TWAP strategy converges to trading in a constant rate. Its goal is to make execution price stay close to the time-averaged price between the initial and terminal time so as to minimize market impact of execution. Similarly, a VWAP (volume-weighted average price) strategy is a TWAP strategy except the evolution of time is in terms of cumulative total traded volume. However, as opposed to TWAP strategies which basically require zero modeling for market environment, implementation of a VWAP strategy requires the modeling of market volume process. In practice, TWAP and VWAP strategies are benchmarks for assessing the performance of trading strategies or algorithms.

The performance criterion V in each experiment is decomposed into three additive components V_{PnL}, V_{risk}, and $V_{entropy}$. These components represent, respectively, contributions to the performance criterion V from the expected P&L, the risk, and the relative entropy. The performance criterion, as well as its three components, are then stress tested against certain extreme parameters. Monte Carlo simulations in each experiment are conducted with 4096 sample paths in 1000 time steps. The parameters chosen in the following numerical examples, mainly referenced to the parameters considered in Almgren and Chriss (2001), Cheng *et al.* (2017), and Di Giacinto *et al.* (2024), are for convenience only. In reality, the parameters need to be calibrated to the market data.

7.1. Model 1

In model 1, the three components for performance criterion V are specified as

$$V=V_{PnL}+V_{risk}+V_{entropy},$$

where

$$\begin{aligned}V_{PnL}&=-\delta X_T^2+X_TS_T-X_0S_0-\int_0^T\tilde{S}_t\,\mathrm{d}X_t,\\ V_{risk}&=\int_0^T\left[\frac{1}{2}R_{xx}x_t^2+R_{xa}x_t\langle a\rangle_t\right.\\ &\quad\left.+\frac{1}{2}R_{aa}\left(\langle a\rangle_t^2+\frac{1}{s_t+\beta R_{aa}}\right)\right]\mathrm{d}t,\\ V_{entropy}&=\int_0^T\left[\frac{1}{2\beta}\log\frac{s_t+\beta R_{aa}}{s_t}\right.\\ &\quad-\frac{R_{aa}}{2}\left(\langle a\rangle_t^2+\frac{1}{s_t+\beta R_{aa}}\right)\\ &\quad\left.+\frac{s_t+\beta R_{aa}}{2\beta}\langle a\rangle_t^2\right]\mathrm{d}t.\end{aligned}$$

The strategies, with their corresponding optimal posterior mean of market trading rate, implemented are

(1) Optimal strategy

$$\begin{aligned}v_t&=-\hat{A}_1\coth\{\hat{A}_1(T-t)+\alpha_1\}x_t,\\ \langle a\rangle_t&=-\frac{\beta(\gamma_M+R_{xa})}{s_t+\beta R_{aa}}x_t;\end{aligned}$$

(2) TWAP strategy

$$v_t=-\frac{x_t}{T-t},\quad\langle a\rangle_t=-\frac{\beta(\gamma_M+R_{xa})}{s_t+\beta R_{aa}}x_t.$$

In the simulations that follow, parameters below are chosen and fixed across simulations.

$$\begin{aligned}&\gamma=2.5\times10^{-7},\quad R_{xa}=-5\times10^{-6},\quad R_{aa}=9\times10^{-7},\\ &\beta=1,m_t=0,\\ &s_t=10^{-8},\quad X=10^6,\quad S=100,\quad\sigma^X=10^5,\quad\sigma^S=10,\\ &\rho=0.3,\quad T=1.\end{aligned}$$

The simulation results are shown in figures 2–4. In each figure, histograms for performance criterion and its three components are exhibited for the optimal and its related TWAP strategies; on top of the histograms show the Box plots generated by the data. The parameters as benchmark are set by

$$\gamma_M=2.5\times10^{-6},\quad\eta=2.5\times10^{-6},\quad\delta=1.25\times10^{-4},\quad R_{xx}=-10^{-6}$$

whose simulation results are shown in figure 1. Figure 2 shows the result of stress testing the optimal and TWAP strategies against large permanent impact from market $\gamma_M=10^{-5}$. Figure 3 shows the result of stress testing against relatively

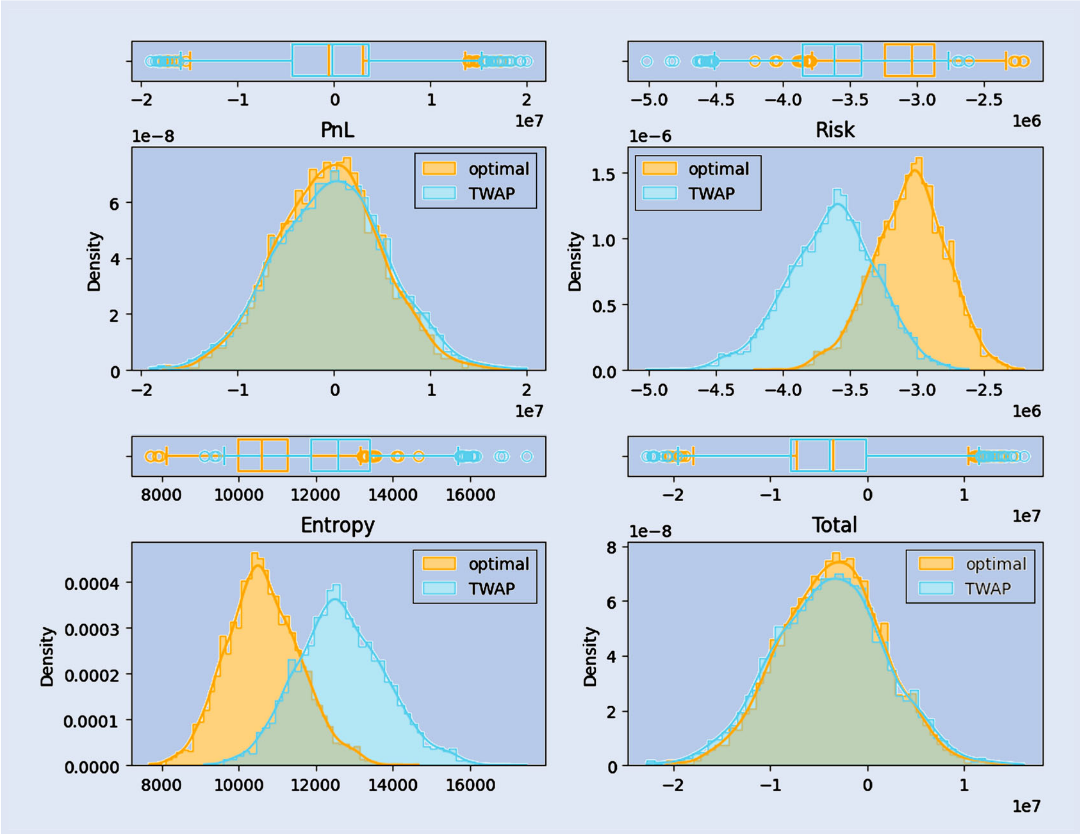

Figure 1. Performance under benchmark parameters in model 1.

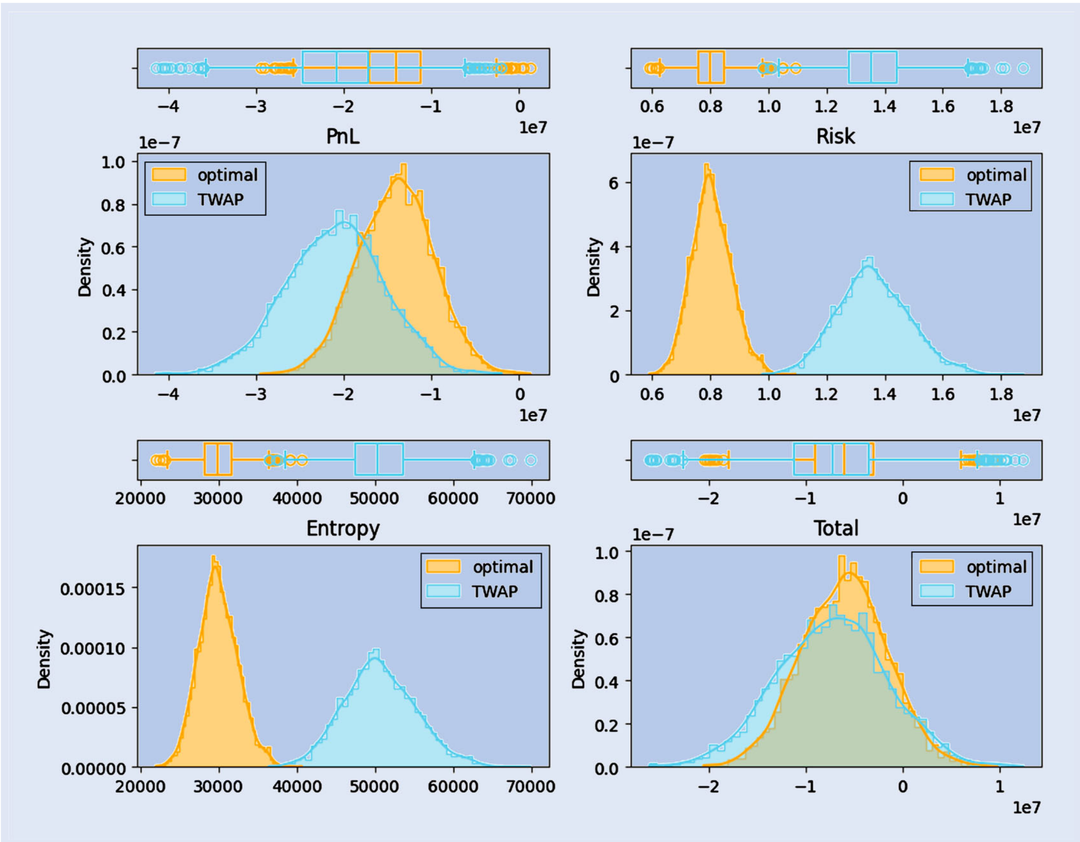

Figure 2. Performance under relatively large $\gamma_M = 10^{-5}$ in model 1.

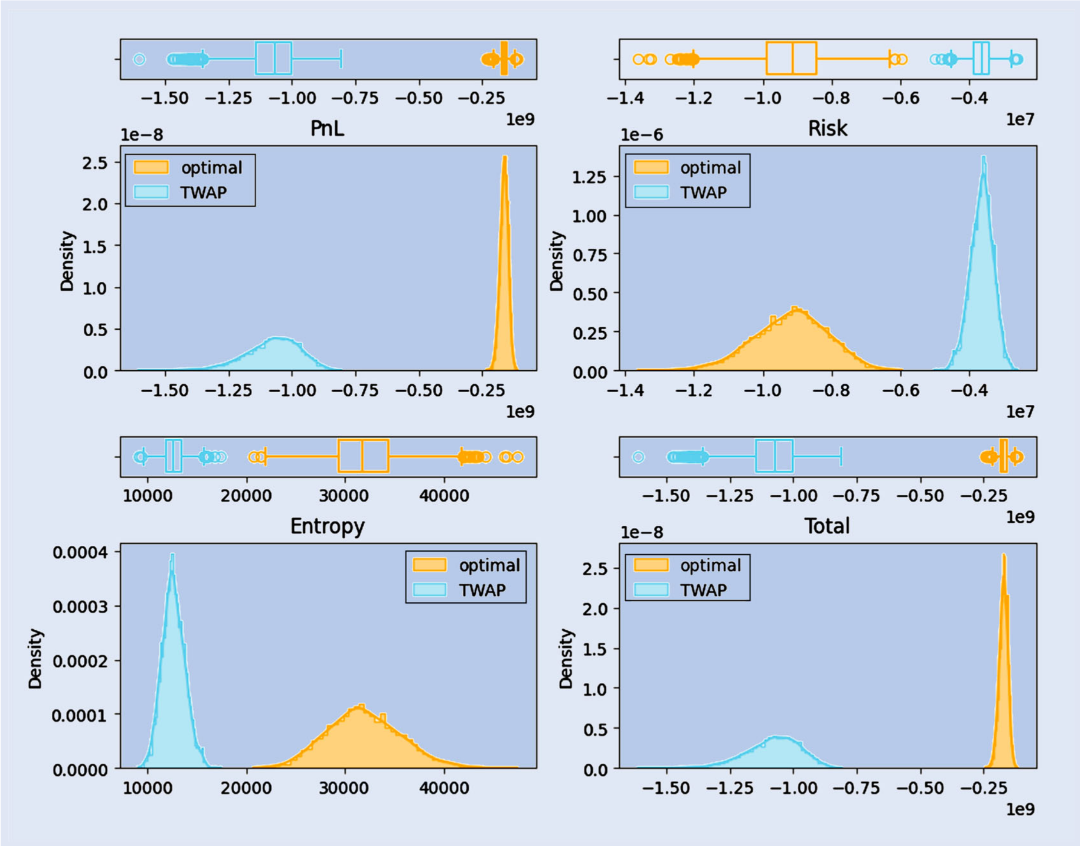

Figure 3. Performance under relatively large $\eta = 10^{-3}$ and small $\delta = 2 \times 10^{-4}$ in model 1.

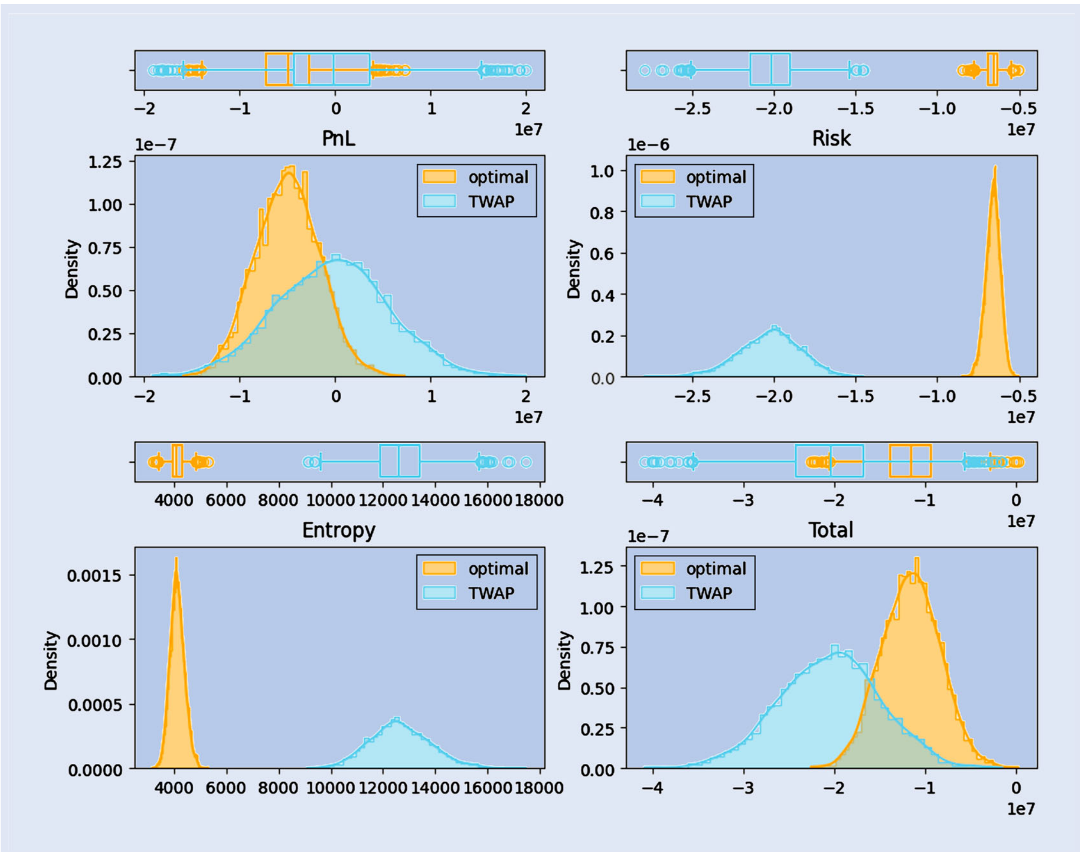

Figure 4. Performance under relatively large negative $R_{xx} = -10^{-4}$ in model 1.

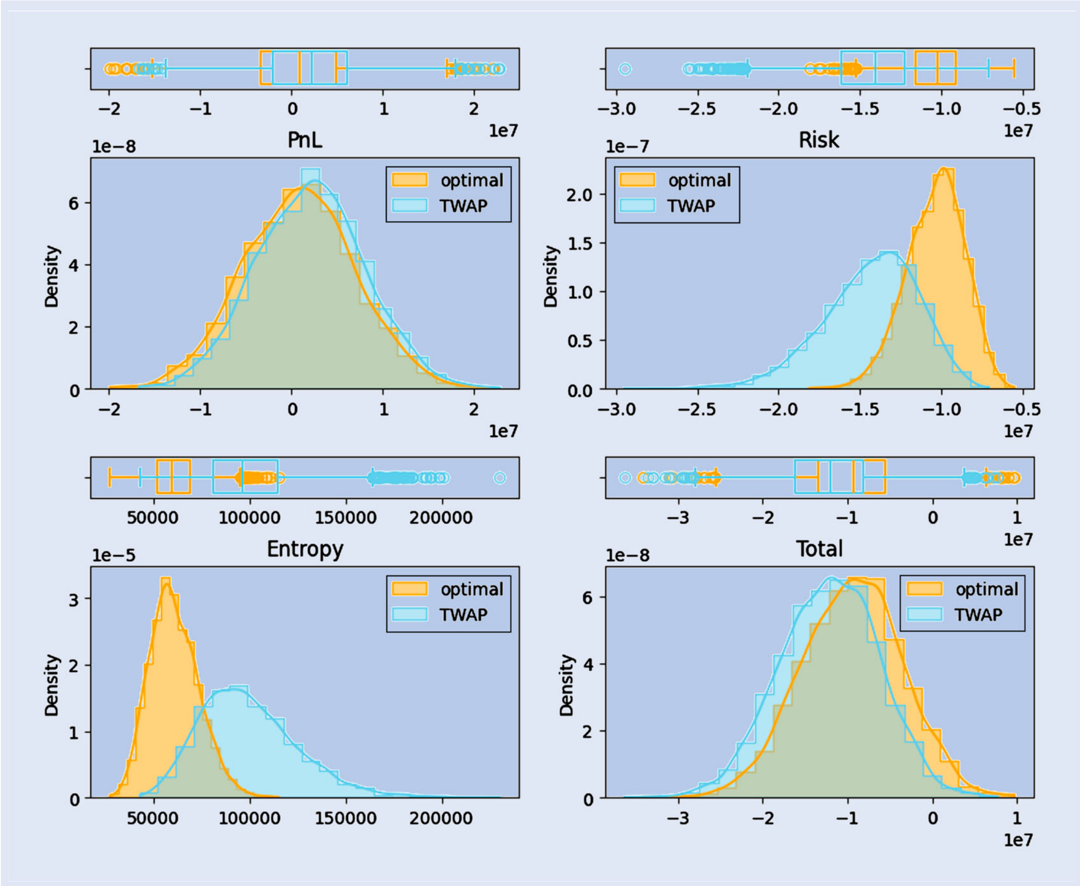

Figure 5. Performance under benchmark parameters in model 2.

Table 1. Parameters and figures for model 1.

	γ_M	η	δ	R_{xx}
Benchmark	2.5×10^{-6}	2.5×10^{-6}	1.25×10^{-4}	-10^{-6}
Figure 2	10^{-5}	2.5×10^{-6}	1.25×10^{-4}	-10^{-6}
Figure 3	2.5×10^{-6}	10^{-3}	2×10^{-4}	-10^{-6}
Figure 4	2.5×10^{-6}	2.5×10^{-6}	1.25×10^{-4}	-10^{-4}

large $\eta = 10^{-3}$ and small $\delta = 2\times10^{-4}$. Finally, figure 4 is for stress testing against relatively large $R_{xx} = 10^{-4}$. Table 1 summarizes the parameters in each case and their corresponding figures.

We observe from the simulation results that the total performances under TWAP and under optimal strategy in the benchmark case are almost identical while TWAP performs better in entropy but worse in risk. The variabilities of the parameters γ_M, γ, η, and δ mainly contribute to the variance in P&L, while R_{xx} mainly to the variance of risk. We remark that in all the experiments, the magnitude of the entropy term V_{entrop} are insignificant compared to those of the risk V_{risk} and of the P&L V_{PnL}. Also, it seems in figure 3, TWAP is doing better than the optimal strategy in V_{risk}. However, while the magnitude between V_{PnL} and V_{risk} in other cases are comparable, the magnitude of V_{PnL} is almost a hundred times higher than that of V_{risk} in this case. It turns out that in this case, in comparison with TWAP, the optimal strategy chooses to sacrifice risk for higher P&L. Overall, apart from the benchmark case, the optimal strategies outperform and incurring lower variance as opposed to their TWAP counterparties as shown in the figures.

7.2. Model 2

For model 2, the two components V_{PnL} and $V_{entropy}$ of the performance criterion V remain the same as in model 1, whereas the component V_{risk} in this case becomes

$$V_{risk} = \int_0^T \left[\frac{1}{2}R_{vv}v_t^2 + R_{va}v_t\langle a\rangle_t + \frac{1}{2}R_{aa}\left(\langle a\rangle_t^2 + \frac{1}{s_t+\beta R_{aa}}\right)\right]\mathrm{d}t.$$

The strategies implemented in model 2 are

(1) Optimal strategy

$$v_t = -\hat{A}_2\coth\{\hat{A}_2(T-t)+\alpha_2\}x_t,$$
$$\langle a\rangle_t = [C\hat{A}_2\coth\{\hat{A}_2(T-t)+\alpha_2\} - A_2]\frac{x_t}{\gamma_M};$$

(2) TWAP strategy

$$v_t = -\frac{x_t}{T-t}, \quad \langle a\rangle_t = \left[\frac{C}{T-t} - A_2\right]\frac{x_t}{\gamma_M}.$$

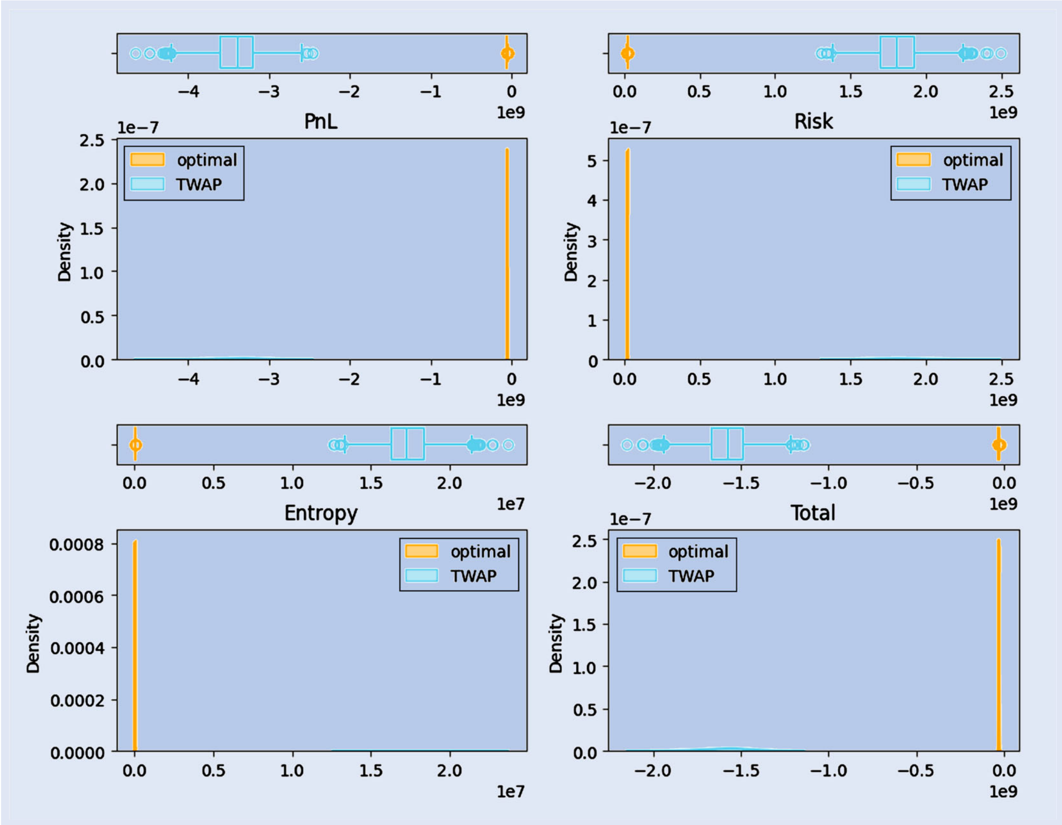

Figure 6. Performance under relatively large $\gamma_M = 10^{-5}$ in model 2.

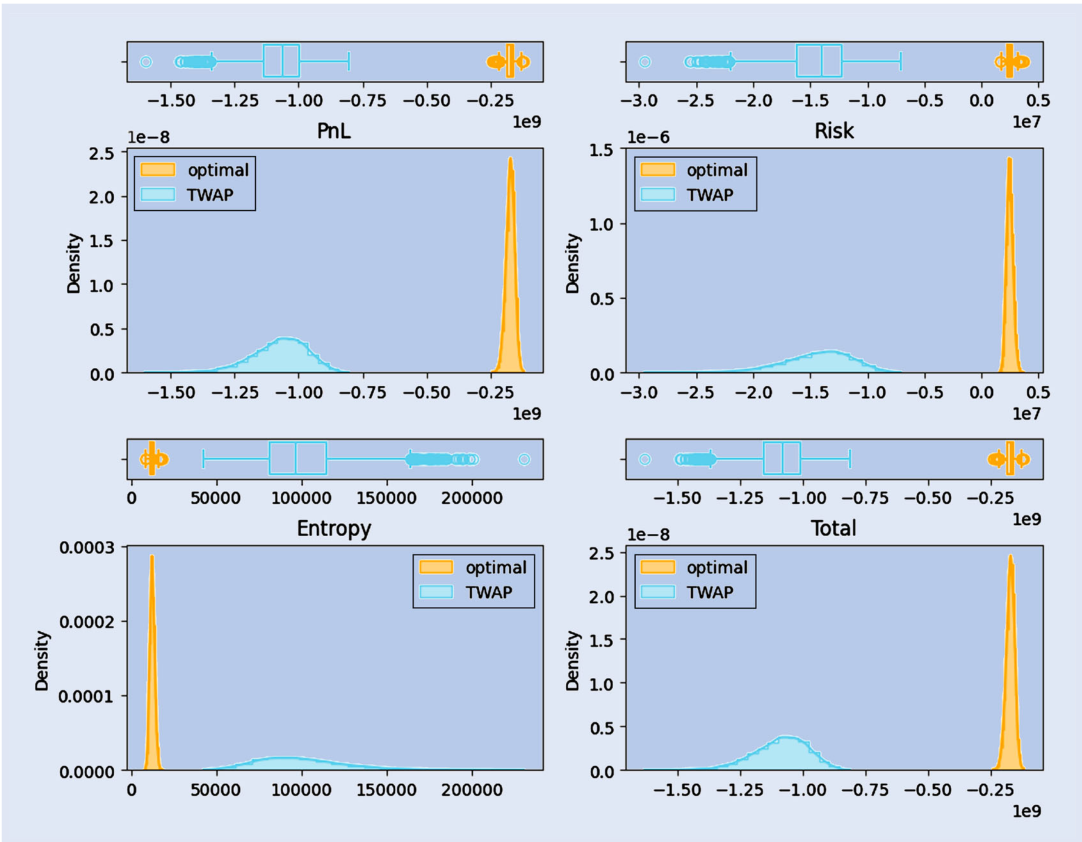

Figure 7. Performance under relatively large $\eta = 10^{-3}$ and small $\delta = 2 \times 10^{-4}$ in model 2.

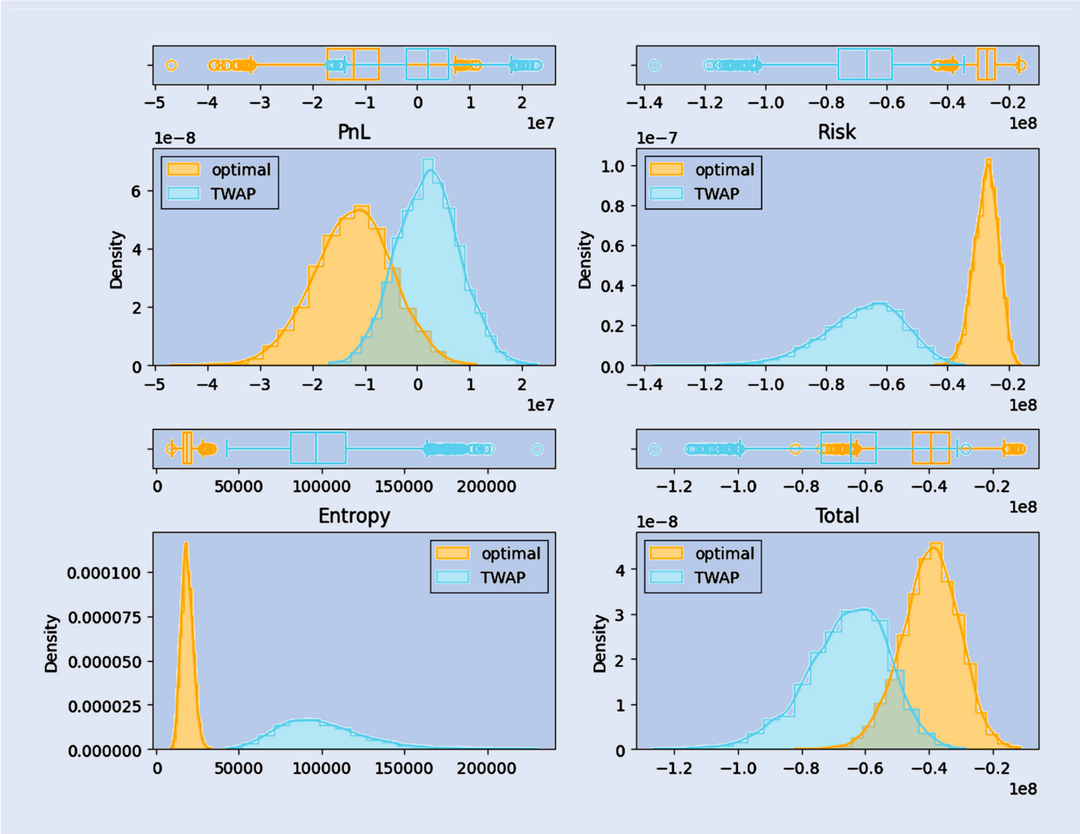

Figure 8. Performance under relatively large negative $R_{vv} = -10^{-4}$ in model 2.

As for model 1, in the simulations that follow the parameters below are chosen and fixed across simulations.

$$\gamma = 2.5 \times 10^{-7}, \quad R_{va} = 5 \times 10^{-6}, \quad R_{aa} = 9 \times 10^{-7},$$
$$\beta = 1, m_t = 0,$$
$$s_t = 10^{-8}, \quad X = 10^6, \quad S = 100, \quad \sigma^X = 10^5, \quad \sigma^S = 10,$$
$$\rho = 0.3, \quad T = 1.$$

We demonstrate the results in figures 6–8. Each figure exhibits histograms for the performance criterion and its three components of the optimal and its related TWAP strategies, topped with Box plots. The parameters as benchmark are set by

$$\gamma_M = 2.5 \times 10^{-6}, \quad \eta = 2.5 \times 10^{-6}, \quad \delta = 1.25 \times 10^{-4},$$
$$R_{vv} = -10^{-6}$$

whose simulation results are shown in figure 5. Figure 6 shows the result of stress testing the optimal and TWAP strategies against large permanent impact from market $\gamma_M = 10^{-5}$. Figure 7 shows the result of stress testing against relatively large $\eta = 10^{-3}$ and small $\delta = 2 \times 10^{-4}$. Finally, figure 8 is for stress testing against relatively large $R_{vv} = 10^{-4}$. Table 2 summarizes the parameters in each case and their corresponding figures.

We again summarize that, from the simulation results, the variabilities of the parameters γ_M, γ, η, and δ mainly contribute to the variance in P&L, while R_{vv} mainly to the variance of risk. We remark that, among the histograms shown, histograms under optimal strategies shown in figures 6 and 7 are much more concentrated than those of TWAP, when compared with the other two cases. Notice that these two cases are for stress tests with large impact coefficients: figure 6 for large market permanent impact coefficient γ_M and figure 7 for large temporary impact η. Heuristics can be that, as mentioned earlier that TWAP implementation does not take dynamic of market environment into account, when exposed to extreme market conditions, it performs much more wildly compared to the optimal. Overall, the optimal strategies outperform, in some cases much better, and incurring lower variance as opposed to their TWAP counterparties as shown in the figures.

Table 2. Parameters and figures for model 2.

	γ_M	η	δ	R_{vv}
Benchmark	2.5×10^{-6}	2.5×10^{-6}	1.25×10^{-4}	-10^{-6}
Figure 6	10^{-5}	2.5×10^{-6}	1.25×10^{-4}	-10^{-6}
Figure 7	2.5×10^{-6}	10^{-3}	2×10^{-4}	-10^{-6}
Figure 8	2.5×10^{-6}	2.5×10^{-6}	1.25×10^{-4}	-10^{-4}

8. Conclusion and discussion

In this article, we cast the order execution problem as a relative entropy-regularized robust optimal control problem based on the principle of least relative entropy for market liquidity and uncertainty. The market impact model proposed in the article added a permanent impact component from averaged market trading rate for the purpose of assessing the market's liquidity and uncertainty risk in order execution. Under the assumptions of Gaussian distribution and linear-quadratic framework, the value function, the optimal strategy as well as the posterior distribution for market trading rate for the resulting regularized stochastic differential game were obtainable subject to solving a system of Riccati and linear differential equations. Two models assumptions were proposed and the resulting Riccati equations and their corresponding solutions were shown. Calibration of the model to market data and significance of the market trading rate component worth following up in future studies. Extensions to the current model include (a) agent's impact being transient while market's impact remains permanent; (b) the use of f-divergence for regularization.

Acknowledgments

We are grateful for comments from the Mathematical Finance Seminar participants at Ritsumeikan University and the participants of the XXIV Quantitative Finance Workshop at University of Cassino in Gaeta, Italy. We thank Jiro Akahori, Arturo Kohatsu-Higa, and Claudio Tebaldi for helpful discussions and suggestions.

Funding

THW is partially supported by the National Natural Science Foundation of China [grant number 11971040].

Disclosure statement

No potential conflict of interest was reported by the author(s).

References

Alfonsi, A., Fruth, A. and Schied, A., Optimal execution strategies in limit order books with general shape functions. *Quant. Finance*, 2010, **10**(2), 143–157.

Almgren, R. and Chriss, N., Optimal execution of portfolio transactions. *J. Risk*, 2001, **3**(2), 5–39.

Anderson, E.W., Hansen, L.P. and Sargent, T.J., A quartet of semigroups for model specification, robustness, prices of risk, and model detection. *J. Eur. Econ. Assoc.*, 2003, **1**(1), 68–123.

Carmona, R. and Leal, L., Optimal execution with quadratic variation inventories. *SIAM J. Financ. Math.*, 2023, **14**(3), 751–776. https://doi.org/10.1137/21M1416564.

Carmona, R. and Webster, K., The self-financing equation in limit order book markets. *Finance Stochastics*, 2019, **23**(3), 729–759.

Cheng, X., Di Giacinto, M. and Wang, T.-H., Optimal execution with uncertain order fills in Almgren-Chriss framework. *Quant. Finance*, 2017, **17**(1), 55–69.

Cheng, X., Di Giacinto, M. and Wang, T.-H., Optimal execution with dynamic risk adjustment. *J. Oper. Res. Soc.*, 2019, **70**(10), 1662–1677.

Di Giacinto, M., Tebaldi, C. and Wang, T.-H., Optimal order execution under price impact: A hybrid model. *Ann. Oper. Res.*, 2024, **336**, 605–636.

Fleming, W.H. and Souganidis, P.E., On the existence of value functions of two-player, zero-sum stochastic differential games. *Indiana Univ. Math. J.*, 1989, **38**(2), 293–314.

Flemming, W.H. and Nisio, M., On stochastic relaxed control for partially observed diffusions. *Nagoya Math. J.*, 1984, **93**, 71–108.

Hansen, P. and Miao, J., Aversion to ambiguity and model misspecification in dynamic stochastic environments. *Proc. Natl. Acad. Sci.*, 2018, **115**(37), 9163–9168.

Kim, J. and Yang, I., *Hamilton-Jacobi-Bellman equations for maximum entropy optimal control*, preprint (2020). Available at https://arxiv.org/abs/2009.13097.

Kyle, A.S., Continuous auctions and insider trading. *Econometrica*, 1985, **53**(6), 1315–1335.

Ma, G., Siu, C.C., Zhu, S.-P. and Elliott, R.J., Optimal portfolio execution problem with stochastic price impact. *Automatica*, 2020, **112**, 108739.

Maccheroni, F., Marinacci, M. and Rustichini, A., Ambiguity aversion, robustness, and the variational representation of preferences. *Econometrica*, 2006, **74**(6), 1447–1498.

Mariani, F. and Fatone, L., Optimal solution of the liquidation problem under execution and price impact risks. *Quant. Finance*, 2022, **22**(6), 1037–1049.

Obizhaeva, A.A. and Wang, J., Optimal trading strategy and supply/demand dynamics. *J. Financ. Mark.*, 2013, **16**, 1–32.

Wang, H., Zariphopoulou, T. and Zhou, X.Y., Reinforcement learning in continuous time and space: A stochastic control approach. *J. Mach. Learn. Res.*, 2020, **21**(198), 1–34.

Appendix. Relative entropy-regularized control problem

In this appendix, we briefly review the relative entropy-regularized control problem for reader's convenience. For more detailed discussions on entropy-regularized stochastic control problem we refer to Flemming and Nisio (1984) and more recently Kim and Yang (2020), Wang *et al.* (2020), and the references therein.

The goal of relative entropy-regularized control problem is to determine an optimal distribution of controls that maximizes or minimizes the objective functional. Specifically, let $\pi_t^0(a\,|\,x)$ be the 'prior' distribution of controls at time t and state x. For any given distribution $\pi_t(a)$ of control a at time t, consider the controlled SDE

$$\mathrm{d}X_t = \bar{\mu}(X_t, \pi_t)\,\mathrm{d}t + \bar{\sigma}(X_t, \pi_t)\,\mathrm{d}W_t,$$

where

$$\bar{\mu}(x,\pi) = \int \mu(x,a)\,\mathrm{d}\pi(a), \quad \bar{\sigma}(x,\pi) = \sqrt{\int \sigma^2(x,a)\,\mathrm{d}\pi(a)}$$

for some drift μ and volatility σ. The relative entropy-regularized control problem seeks to determine an optimal 'posterior' distribution $\pi_t(a)$ of controls at time t amongst certain admissible distributions $\mathcal{A}$ that optimizes the following expected relative entropy-regularized, or Kullback–Leibler divergence regularized, objective functional

$$\min_{\pi\in\mathcal{A}} \mathbb{E}\left[g(X_T) + \int_0^T \int \left\{ R(t, X_t, a) + \frac{1}{\beta}\log\frac{\pi_t(a)}{\pi_t^0(a\,|\,X_t)} \right\} \times \pi_t(a)\,\mathrm{d}a\,\mathrm{d}t \right] \tag{A1}$$

subject to π_t being a probability measure for all t, i.e. $\pi_t \geq 0$ and $\int \pi_t(a)\,\mathrm{d}a = 1$ for all t. $\beta > 0$ is the parameter that enforces the closeness in the sense of Kullback–Leibler divergence between the distributions π_t and π_t^0; the larger the β, the less tightness between the two distributions is allowed.

To solve the problem, as in the classical control theory, define the value function V for the relative entropy-regularized control problem (A1) as

$$V(t,x) = \min_{\pi \in \mathcal{A}_t} \mathbb{E}\left[g(X_T) + \int_t^T \int \left\{ R(s,X_s,a) + \frac{1}{\beta} \log \frac{\pi_s(a)}{\pi_t^0(a \mid X_s)} \right\} \times \pi_s(a)\, \mathrm{d}a\, \mathrm{d}s \middle| X_t = x \right].$$

One can show, by applying the standard argument as in the classical control theory, that V satisfies the following *relative entropy-regularized HJB (rHJB hereafter) equation*.

$$\min_{\pi_t} \int \left\{ V_t + \frac{\sigma^2}{2} V_{xx} + \mu V_x + R(t,x,a) + \frac{1}{\beta} \log \frac{\pi_t(a)}{\pi_t^0(a \mid x)} \right\} \times \pi_t(a)\, \mathrm{d}a = 0, \tag{A2}$$

where $\pi_t \geq 0$ and $\int \pi_t(a \mid x)\, \mathrm{d}a = 1$ for all x, with terminal condition $V(T,x) = g(x)$.

First order criterion by variational calculus implies that the optimal 'posterior' distribution π_t is given in terms of feedback control as

$$\pi_t(a) = \frac{1}{Z(x)} \pi_t^0(a \mid x)\, \mathrm{e}^{-\beta\{V_t + \frac{\sigma^2}{2} V_{xx} + \mu V_x + R(t,x,a)\}}, \tag{A3}$$

where $Z(x)$ is the normalizing constant

$$Z(x) = \int \pi_t^0(a \mid x)\, \mathrm{e}^{-\beta\{V_t + \frac{\sigma^2}{2} V_{xx} + \mu V_x + R(t,x,a)\}}\, \mathrm{d}a. \tag{A4}$$

We remark that (A3) resembles a Gibbs measure where β plays the role of inverse temperature.

By substituting (A3) and (A4) into (A2), the rHJB equation (A2) reduces to

$$-\frac{1}{\beta} \log \int \pi_t^0(a \mid x)\, \mathrm{e}^{-\beta\{V_t + \frac{\sigma^2}{2} V_{xx} + \mu V_x + R(t,x,a)\}}\, \mathrm{d}a = 0 \tag{A5}$$

or equivalently

$$\int \pi_t^0(a \mid x)\, \mathrm{e}^{-\beta\{V_t + \frac{\sigma^2}{2} V_{xx} + \mu V_x + R(t,x,a)\}}\, \mathrm{d}a = 1 \tag{A6}$$

with terminal condition $V(T,x) = g(x)$. We conclude that the optimal 'posterior' distribution π_t^* is given in terms of the value function V and the 'prior' distribution π_t^0 as

$$\pi_t^*(a) = \pi_t^0(a \mid x)\, \mathrm{e}^{-\beta\{V_t + \frac{\sigma^2}{2} V_{xx} + \mu V_x + R(t,x,a)\}}$$

and the rHJB equation (A6) guarantees that the π_t^* given above is indeed a probability distribution.

As an extension to the relative entropy-regularized control problem, one may consider the following *f-divergence regularized* control problem.

$$\min_{\pi} \mathbb{E}\left[g(X_T) + \int_0^T \int \left\{ R(t,X_t,a) + \frac{1}{\beta} f\left(\frac{\pi_t(a)}{\pi_t^0(a)} \right) \right\} \pi_t(a)\, \mathrm{d}a\, \mathrm{d}t \right] \tag{A7}$$

for some function f. We remark that the Kullback–Leibler divergence is recovered by letting $f(x) = \log x$ and the Tsallis divergence by $f(x) = \frac{1}{p-1}(x^{p-1} - 1)$ for $p > 1$.

By the same token, one can show that the value of the control problem (A7) satisfies an rHJB equation. Let $\tilde{f}(x) = xf(x)$, $\mathcal{L}V = \frac{\sigma^2}{2} V_{xx} + \mu V_x + R(t,x,a)$ and λ the Lagrange multiplier. The f-divergence rHJB equation reads

$$\int \left\{ V_t + \mathcal{L}V + \frac{1}{\beta} f \circ (\tilde{f}')^{-1}(-\beta \mathcal{L}V + \beta\lambda) \right\} \times (\tilde{f}')^{-1}(-\beta \mathcal{L}V + \beta\lambda) \pi_t^0(a \mid x)\, \mathrm{d}a = 0,$$

$$\int (\tilde{f}')^{-1}(-\beta \mathcal{L}V + \beta\lambda) \pi_t^0(a \mid x)\, \mathrm{d}a = 1$$

which unfortunately cannot be further simplified. We note that, in the case of relative entropy, i.e. $f(x) = \log x$, the two equations above can be combined into one, as shown in (A5).