

Option market making with hedging-induced market impact

Paulin Aubert* Etienne Chevalier† Vathana Ly Vath‡

November 5, 2025

Abstract

This paper develops a model for option market making in which the hedging activity of the market maker generates price impact on the underlying asset. The option order flow is modeled by Cox processes, with intensities depending on the state of the underlying and on the market maker's quoted prices. The resulting dynamics combine stochastic option demand with both permanent and transient impact on the underlying, leading to a coupled evolution of inventory and price. We first study market manipulation and arbitrage phenomena that may arise from the feedback between option trading and underlying impact. We then establish the well-posedness of the mixed control problem, which involves continuous quoting decisions and impulsive hedging actions. Finally, we implement a numerical method based on policy optimization to approximate optimal strategies and illustrate the interplay between option market liquidity, inventory risk, and underlying impact.

Keywords: Option market making, Cox processes, Mixed stochastic control, Policy optimization, Machine learning.

1 Introduction

Academic research on market making has progressively shifted from stylized microstructure models toward dynamic frameworks able to capture high-frequency trading, inventory risk, and market impact. A key milestone in this evolution is the stochastic-control approach of Avellaneda and Stoikov [4], which formulates the quoting problem in continuous time and solves it through Hamilton–Jacobi–Bellman equations. This framework has been extended in various directions, for instance by Guilbaud and Pham [12], who incorporated both limit and market orders, or by Guéant et al. [11], who derived explicit solutions in an affine setting, enabling practical calibration. Cartea and coauthors [8, 7] further broadened the perspective by linking market making to optimal execution and risk management.

In parallel, a large literature has developed on optimal execution and market impact, starting with Almgren and Chriss [3] and enriched by contributions that distinguish temporary and permanent impact [10] or account for the self-exciting nature of order flow via Hawkes processes [2]. These advances are highly relevant for option market making, where quoting decisions affect not only order flow but also the underlying dynamics through hedging. Models explicitly integrating such feedback effects have emerged more recently. For example, Stoikov and Saglam [16] introduced delta-aware option market making, El

*Laboratoire de Mathématiques et Modélisation d'Évry, Université Évry Paris-Saclay, Exiom Partners, France, paulin.aubert@univ-evry.fr.

†Laboratoire de Mathématiques et Modélisation d'Évry, Université Évry Paris-Saclay, UMR 8071 CNRS, France, etienne.chevalier@univ-evry.fr.

‡Laboratoire de Mathématiques et Modélisation d'Évry, Université Paris-Saclay, ENSIIE, UMR 8071 CNRS, France, vathana.lyvath@ensiie.fr.

Aoud and Abergel [9] modeled joint underlying–option dynamics in a stochastic-control framework, and Baldacci, Bergault and Guéant [5] developed a multi-option setting where order intensities depend on aggregate sensitivities.

While these models provide valuable insight, classical solution techniques become rapidly intractable as soon as one accounts for coupled markets, feedback loops, and state-dependent intensities. This has motivated the use of simulation-based and machine-learning approaches. Spooner et al. [15] applied reinforcement learning to learn quoting strategies without an explicit parametric model, and Buehler et al. [6] introduced the Deep Hedging framework, where neural networks optimize strategies directly on simulated paths.

In this paper, we consider a market maker mandated to ensure the liquidity of a European option, while complying with hedging constraints imposed by a third party (regulator or principal). Unlike other approaches in the literature, notably [16, 9, 5], we make no explicit assumptions about the dynamics of the underlying option price. Instead, we model order arrivals in the options market by Cox processes whose intensities depend on both the quotes offered by the market maker and the state of the underlying. These intensities are modeled in a flexible way, allowing the framework to reflect key market effects such as the dependence of order-flow activity on option moneyness and time to expiry. Furthermore, we assume that the market maker has an impact on the price of the underlying via the consumption of the order book, whose shape is sufficiently general to represent low-liquidity assets. This impact is modeled by a permanent and a resilient component. Finally, we develop a numerical method based on neural networks to approximate the market maker’s optimal quoting strategy, while taking into account the associated liquidity and transaction risks.

The remainder of the paper is structured as follows. Section 2 introduces the modeling framework and formulates the market-making problem. Section 3 establishes the model’s consistency by examining market integrity and the well-posedness of the control problem. It demonstrates the absence of instantaneous and dynamic arbitrage opportunities, rules out transaction-triggered price manipulation, derives a terminal bound in the coupled setting, and proves quadratic-growth estimates ensuring the finiteness of the value function. Finally, Section 4 develops the numerical methodology and reports experiments. We describe the discrete-time simulator and the neural policy training procedure, then present empirical results in a baseline configuration, two asymmetric scenarios, and a low-liquidity regime.

2 Market modeling and the market maker problem

2.1 Modeling the underlying market

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space equipped with a filtration $(\mathcal{F}_t)_{t \geq 0}$ satisfying the usual conditions. Throughout this work, all random variables and stochastic processes are defined on $(\Omega, \mathcal{F}, \mathbb{P})$.

2.1.1 Order book representation

In line with the approach of Predoiu, Shaikhet, and Shreve [14], we aim to represent the order book of the underlying asset while maintaining full flexibility regarding its structure. In particular, the framework accommodates order books with missing price levels, which may result in a discontinuous shape with gaps. We assume that the shape of the order book is time-invariant. It is characterized by two piecewise continuous functions: f_B for the bid side and f_A for the ask side. These functions are required to satisfy Assumptions 2.1 and 2.2.

Assumption 2.1 (Integrability of the order book shape). *The order book densities satisfy*

$$f_B : \mathbb{R}^- \rightarrow [0, \infty), \quad f_A : \mathbb{R}^+ \rightarrow [0, \infty),$$

together with the integrability conditions:

$$\int_{-\infty}^0 (1 + |u|) f_B(u) du < \infty, \quad \int_0^{\infty} (1 + u) f_A(u) du < \infty.$$

Assumption 2.1 ensures that the total quantity and value represented in the order book are finite. This condition reflects the idea that market participants do not perceive the firm's fundamental value as unbounded. We also impose that beyond a certain price level no liquidity is available, which is consistent with empirical observations that liquidity does not extend indefinitely. This requirement is formalized in Assumption 2.2

Assumption 2.2 (Right-truncated depth). *There exist $U_A, U_B \in (0, \infty)$ such that*

$$f_B(u) = 0 \text{ for a.e. } u < -U_B, \quad f_A(u) = 0 \text{ for a.e. } u > U_A.$$

To further characterize the order book, we introduce cumulative distribution functions that quantify the total volume available when liquidity is consumed up to a given depth. For the bid side, given a current best bid price $b > 0$, we set

$$\Phi_B(x) := \int_{-x}^0 f_B(u) du, \quad x \in [0, b]. \quad (2.1)$$

The restriction to $x \in [0, b]$ ensures that the bid side does not extend below zero prices, where liquidity is absent. On the ask side we define

$$\Phi_A(x) := \int_0^x f_A(u) du, \quad x \in \mathbb{R}^+. \quad (2.2)$$

In this case, no analogous restriction is required, since the ask side can in principle extend arbitrarily far above the current price, with the truncation guaranteed only by Assumption 2.2. By Assumptions 2.1 and 2.2, the cumulative functions Φ_B and Φ_A are finite, nondecreasing, and satisfy the bounds

$$\begin{aligned} \forall b > 0, \forall x \in [0, b] : \quad & 0 \leq \Phi_B(x) \leq \Phi_B(b) \leq \Phi_B(U_B) < \infty, \\ \forall x \in \mathbb{R}^+ \cup \{+\infty\} : \quad & 0 \leq \Phi_A(x) \leq \Phi_A(+\infty) = \Phi_A(U_A) < \infty. \end{aligned}$$

The functions Φ_B and Φ_A quantify cumulative liquidity from the best quote up to a given depth: if a market order consumes all liquidity up to level x , the executed volume is $\Phi_B(x)$ or $\Phi_A(x)$. Conversely, given a target volume y , one asks which price interval must be traversed. Since plateaus may occur in Φ_B, Φ_A when there are gaps in the order book, they are not strictly increasing. We therefore use the generalized inverse introduced in Definition 2.1.

Definition 2.1 (Generalized inverse of cumulative liquidity functions). *Let $b > 0$ denote the current best bid price. The generalized inverses of the cumulative liquidity functions Φ_B and Φ_A , denoted respectively by Φ_B^{-1} and Φ_A^{-1} , are defined by*

$$\Phi_B^{-1}(y) = \inf\{x \geq 0 \mid \Phi_B(x) \geq y\}, \quad \forall y \in [0, \Phi_B(b)],$$

and

$$\Phi_A^{-1}(y) = \inf\{x \geq 0 \mid \Phi_A(x) \geq y\}, \quad \forall y \in [0, \Phi_A(+\infty)].$$

The generalized inverse functions Φ_B^{-1} and Φ_A^{-1} are nondecreasing on their respective domains and bounded, with

$$0 \leq \Phi_B^{-1}(y) \leq U_B \quad \text{for } y \in [0, \Phi_B(b)], \quad 0 \leq \Phi_A^{-1}(y) \leq U_A \quad \text{for } y \in [0, \Phi_A(+\infty)]. \quad (2.3)$$

Moreover, they have finite total variation on their domains and are therefore integrable.

2.1.2 Execution price and cost functions

We now define execution cost functions, which link order size to the monetary amount required to consume the book up to a given depth. Let $b > 0$ denote the current best bid price. When an agent sells a quantity $0 \leq q^- \leq \Phi_B(b)$, the order book is consumed starting from b down to the level $b - \Phi_B^{-1}(q^-)$. The total revenue obtained (or, equivalently, the monetary value of the trade from the buyer's perspective) is given by

$$P_B(b, q^-) = \int_{b - \Phi_B^{-1}(q^-)}^b y f_B(y - b) dy.$$

Introducing the change of variables $u = y - b$, so that $y = b + u$ and $u \in [-\Phi_B^{-1}(q^-), 0]$, this expression becomes

$$P_B(b, q^-) = \int_{-\Phi_B^{-1}(q^-)}^0 (b + u) f_B(u) du. \quad (2.4)$$

By symmetry, let $a > 0$ denote the best ask price. If an agent buys a quantity $0 \leq q^+ \leq \Phi_A(+\infty)$, the book is consumed from a up to the level $a + \Phi_A^{-1}(q^+)$. The corresponding execution cost is

$$P_A(a, q^+) = \int_a^{a + \Phi_A^{-1}(q^+)} y f_A(y - a) dy.$$

With the substitution $u = y - a$, so that $y = a + u$ and $u \in [0, \Phi_A^{-1}(q^+)]$, this rewrites as

$$P_A(a, q^+) = \int_0^{\Phi_A^{-1}(q^+)} (a + u) f_A(u) du. \quad (2.5)$$

By Assumption 2.1, the execution costs (2.4)–(2.5) are finite for all admissible order sizes $0 \leq q^- \leq \Phi_B(b)$ and $0 \leq q^+ \leq \Phi_A(+\infty)$.

We next examine structural properties of execution costs, showing that they admit integral representations and satisfy convexity/concavity.

Lemma 2.1 (Analytical structure of execution costs). *Let $b > 0$ the best bid price and $a > 0$ denote the best ask price. For every admissible order size*

$$0 \leq q^- \leq \Phi_B(b) \quad \text{and} \quad 0 \leq q^+ \leq \Phi_A(+\infty),$$

the cost functionals (2.4) and (2.5) admit the representations

$$P_B(b, q^-) = \int_{-\Phi_B^{-1}(q^-)}^0 (b + u) f_B(u) du = bq^- - \int_0^{q^-} \Phi_B^{-1}(y) dy, \quad (2.6)$$

$$P_A(a, q^+) = \int_0^{\Phi_A^{-1}(q^+)} (a + u) f_A(u) du = aq^+ + \int_0^{q^+} \Phi_A^{-1}(y) dy. \quad (2.7)$$

Consequently, $q \mapsto P_B(b, q)$ is concave and $q \mapsto P_A(a, q)$ is convex on their domains. Moreover, both maps are absolutely continuous and \mathcal{C}^1 a.e., with a.e. derivatives

$$\partial_q P_B(b, q) = b - \Phi_B^{-1}(q), \quad \partial_q P_A(a, q) = a + \Phi_A^{-1}(q).$$

In particular, the marginal execution price remains bounded between $b - U_B$ and b on the bid side, and between a and $a + U_A$ on the ask side.

Proof. The equalities (2.6) and (2.7) follow from the following identities:

$$\int_{-x}^0 u f_B(u) du = - \int_0^{\Phi_B(x)} \Phi_B^{-1}(y) dy, \quad \int_0^x u d\Phi_A(u) = \int_0^{\Phi_A(x)} \Phi_A^{-1}(y) dy.$$

Here $q^- = \Phi_B(x)$ and $q^+ = \Phi_A(x)$. Moreover, $d\Phi_A(u) = f_A(u)du$ and $d\Phi_B(u) = f_B(u)du$. Since Φ_B^{-1} and Φ_A^{-1} are nondecreasing, the primitives $q \mapsto -\int_0^q \Phi_B^{-1}(y)dy$ and $q \mapsto \int_0^q \Phi_A^{-1}(y)dy$ are, respectively, convex and concave. Adding the linear terms bq and aq preserves these properties. Absolute continuity and the a.e. derivative formulas are immediate. Equation 2.3 gives, $\Phi_B^{-1}(y) \leq U_B$ and $\Phi_A^{-1}(y) \leq U_A$, which yields the stated bounds on the a.e. derivatives. \square

From an economic perspective, the concavity of P_B captures that the marginal revenue from selling decreases with order size, since progressively lower bid quotes are reached when larger quantities are executed. Conversely, the convexity of P_A reflects that the marginal cost of buying increases with order size, as deeper layers of the ask book are consumed at higher prices.

2.1.3 Dynamics of the underlying

Impact of market orders on quotes

Let $(P_t)_{t \geq 0}$ denote the mid-price process and $(S_t)_{t \geq 0}$ the spread process. At any time $t \geq 0$, the best bid and ask prices are given by

$$B_t = P_t - \frac{1}{2}S_t, \quad A_t = P_t + \frac{1}{2}S_t.$$

Consider a state $(p, s) \in (\mathbb{R}^+)^2$, with corresponding best quotes $b = p - s/2$ and $a = p + s/2$. The submission of market orders consumes liquidity and shifts the best quotes. Specifically, a sell order of size $q^- \in [0, \Phi_B(b)]$ moves the bid down to $b - \Phi_B^{-1}(q^-)$, while a buy order of size $q^+ \in [0, \Phi_A(+\infty))$ pushes the ask up to $a + \Phi_A^{-1}(q^+)$. This consumption widens the spread, which becomes

$$\begin{aligned} S_{t+} &= (p + s/2) + \Phi_A^{-1}(q^+) - (p - s/2) + \Phi_B^{-1}(q^-) \\ &= s + \Phi_A^{-1}(q^+) + \Phi_B^{-1}(q^-). \end{aligned}$$

At the same time, the mid-price is updated as the average of the new best bid and ask:

$$\begin{aligned} P_{t+} &= \frac{1}{2} \left((p + s/2) + \Phi_A^{-1}(q^+) + (p - s/2) + \Phi_B^{-1}(q^-) \right) \\ &= p + \frac{1}{2} \left(\Phi_A^{-1}(q^+) - \Phi_B^{-1}(q^-) \right). \end{aligned}$$

These formulas highlight two key effects of order flow on price formation. First, the updated mid-price incorporates the imbalance between buying and selling pressure, shifting upward when buy orders dominate and downward when sell orders prevail. Second, the spread increases as deeper layers of the order book are consumed, reflecting the greater uncertainty and reduced liquidity that follow large market orders.

Order-flow sources and intensity dynamics

The dynamics of the mid-price are driven by two sources of order flow: exogenous trades submitted by other market participants, and the market maker's own interventions used to hedge option exposures. The latter are modeled as impulse trades occurring at stopping times $(\nu_i)_{i \in \mathbb{N}}$ for sales on the bid side and $(\tau_i)_{i \in \mathbb{N}}$ for purchases on the ask side, with traded sizes $\xi_i^- \geq 0$ and $\xi_i^+ \geq 0$, where ξ_i^- is \mathcal{F}_{ν_i} -measurable and ξ_i^+ is \mathcal{F}_{τ_i} -measurable. The corresponding counting processes of interventions are defined by

$$H_t^- := \sum_{i \geq 1} \mathbf{1}_{\{\nu_i \leq t\}}, \quad H_t^+ := \sum_{i \geq 1} \mathbf{1}_{\{\tau_i \leq t\}}. \quad (2.8)$$

with $H_0^- = H_0^+ = 0$.

Exogenous arrivals on each side are modeled by marked Hawkes processes with exponential kernels. On the bid side (sell arrivals), let $(N_t^-)_{t \geq 0}$ be the counting process of exogenous sales hitting the bid.

Each jump time T_k^- of N^- carries an i.i.d. mark $M_k^- \in [0, 1]$ with density f^- . The mark represents the fraction of the locally available depth consumed by that trade. Similarly, on the ask side (buy arrivals), $(N_t^+)_{t \geq 0}$ counts exogenous buys hitting the ask, each jump time T_ℓ^+ being endowed with an i.i.d. mark $M_\ell^+ \in [0, 1]$ with density f^+ . The intensities $(\lambda_t^-)_{t \geq 0}$ and $(\lambda_t^+)_{t \geq 0}$ evolve according to

$$d\lambda_t^- = \theta^-(\mu^- - \lambda_t^-)dt + \kappa^- dN_t^- + \kappa^- dH_t^-, \quad (2.9)$$

and

$$d\lambda_t^+ = \theta^+(\mu^+ - \lambda_t^+)dt + \kappa^+ dN_t^+ + \kappa^+ dH_t^+. \quad (2.10)$$

with initial conditions $\lambda_0^- \geq 0$ and $\lambda_0^+ \geq 0$, where $\mu^\pm \geq 0$ are baseline levels, $\theta^\pm > 0$ are mean-reversion rates, and $\kappa^\pm \geq 0$ are excitation heights.

When $H^- \equiv 0$ and $H^+ \equiv 0$, the dynamics reduce to standard Hawkes processes which is assumed to respect Assumption 2.3.

Assumption 2.3 (Non-explosivity of the Hawkes intensities). *In the absence of interventions, the Hawkes components are subcritical:*

$$\frac{\kappa^-}{\theta^-} < 1 \quad \text{and} \quad \frac{\kappa^+}{\theta^+} < 1,$$

so that (λ_t^\pm) are finite almost surely.

This stability condition rules out explosion, guarantees the well-posedness of the controlled dynamics, and will be instrumental in proving the well-posedness of the value function and of the optimization problem.

Modeling mid-price and spread dynamics

In line with Alfonsi and Blanc [2], each trade has a permanent impact of fraction $\eta \in [0, 1]$ that durably shifts the mid-price, and a transient impact of fraction $1 - \eta$ that decays over time through a mean-reverting process. Let $(P_t)_{t \geq 0}$ denote the mid-price process. It evolves under the combined effect of exogenous buy and sell orders, as well as the market maker's interventions. The dynamics take the form

$$\begin{cases} dP_t = \frac{\eta}{2} \left(\Phi_A^{-1} \left(M_t^+ \Phi_A(+\infty) \right) dN_t^+ - \Phi_B^{-1} \left(M_t^- \Phi_B(B_t) \right) dN_t^- \right) + dD_t, \\ P_{\nu_i^+} = P_{\nu_i} - \frac{\eta}{2} \Phi_B^{-1}(\xi_i^-), \\ P_{\tau_i^+} = P_{\tau_i} + \frac{\eta}{2} \Phi_A^{-1}(\xi_i^+), \end{cases} \quad (2.11)$$

with initial value $P_0 = p \in \mathbb{R}^+$. $(D_t)_{t \geq 0}$ is the transient component of the impact, which satisfies

$$\begin{cases} dD_t = -rD_t dt + \frac{1-\eta}{2} \left(\Phi_A^{-1} \left(M_t^+ \Phi_A(+\infty) \right) dN_t^+ - \Phi_B^{-1} \left(M_t^- \Phi_B(B_t) \right) dN_t^- \right), \\ D_{\nu_i^+} = D_{\nu_i} - \frac{1-\eta}{2} \Phi_B^{-1}(\xi_i^-), \\ D_{\tau_i^+} = D_{\tau_i} + \frac{1-\eta}{2} \Phi_A^{-1}(\xi_i^+), \end{cases} \quad (2.12)$$

where $D_0 = d \in \mathbb{R}$ and $r > 0$ denotes the rate of resilience.

The spread process $(S_t)_{t \geq 0}$ models the distance between bid and ask. It widens when liquidity is consumed on either side of the book and narrows gradually due to resilience effects. Assuming a minimum spread $\delta \geq 0$ and a mean-reversion rate $\rho > 0$, we set

$$\begin{cases} dS_t = -\rho(S_t - \delta)dt + \Phi_A^{-1} \left(M_t^+ \Phi_A(+\infty) \right) dN_t^+ + \Phi_B^{-1} \left(M_t^- \Phi_B(B_t) \right) dN_t^-, \\ S_{\nu_i^+} = S_{\nu_i} + \Phi_B^{-1}(\xi_i^-), \\ S_{\tau_i^+} = S_{\tau_i} + \Phi_A^{-1}(\xi_i^+), \end{cases} \quad (2.13)$$

with initial condition $S_0 = s \geq \delta$.

Equations (2.11)–(2.13) describe how permanent shifts, transient effects, and spread adjustments drive the underlying dynamics. The conditions ensuring strictly positive quotes under admissible strategies are given in Lemma 2.2.

2.2 Modeling the options market

We consider a European option with maturity $T \in \mathbb{R}^+$ and payoff function φ of linear growth, written on the previously described underlying asset. The market maker is mandated to provide liquidity for this derivative product. Let $e := (p, d, s, \lambda_0^-, \lambda_0^+)$ denote the state vector of the underlying market, where p is the mid-price, d the resilient part of the mid-price, s the spread and λ_0^\pm the initial intensities. We assume that all market participants are able to estimate the same reference price for the option, denoted by b , which is a deterministic function:

$$b : (t, e) \in [0, T] \times (\mathbb{R}^+)^5 \mapsto b(t, e) \in \mathbb{R}^+$$

Function b is chosen to respect Assumption 2.4.

Assumption 2.4 (Linear growth of the option reference price). *There exists a constant $C^{(b)} > 0$ such that, for all $(t, e) \in [0, T] \times (\mathbb{R}^+)^5$,*

$$0 \leq b(t, e) \leq C^{(b)}(1 + \|e\|).$$

Depending on the intended hedging strategy, further assumptions may be made about the regularity of b , such as differentiability with respect to certain variables.

The market maker sets bid and ask quotes on the option, denoted by β and α . Arrivals of option orders are modeled by two Cox processes, N^b for buy orders at the bid and N^a for sell orders at the ask. Their intensities are given by bounded deterministic functions

$$\begin{aligned} \lambda^b : (t, \alpha, \beta, e) &\in [0, T] \times (R^+)^2 \times (\mathbb{R}^+)^5 \mapsto \lambda^b(t, \alpha, \beta, e), \\ \lambda^a : (t, \alpha, \beta, e) &\in [0, T] \times (R^+)^2 \times (\mathbb{R}^+)^5 \mapsto \lambda^a(t, \alpha, \beta, e), \end{aligned}$$

in accordance with Assumption 2.5.

Assumption 2.5 (Bounded option trade intensities). *The Cox intensities for option trades are positive and bounded by finite constants $\bar{\lambda}^b, \bar{\lambda}^a > 0$, leading to*

$$0 \leq \lambda^b(t, \alpha, \beta, e) \leq \bar{\lambda}^b, \quad 0 \leq \lambda^a(t, \alpha, \beta, e) \leq \bar{\lambda}^a.$$

Both intensities are assumed to satisfy the linear growth condition

$$\sup_{\alpha, \beta} \left\{ \lambda^a(t, \alpha, \beta, e)(\alpha - b(t, e)) - \lambda^b(t, \alpha, \beta, e)(\beta - b(t, e)) \right\} \leq C^{(\lambda)}(1 + \|e\|),$$

for some constant $C^{(\lambda)} > 0$. This ensures that the expected P&L contribution of option trades remains uniformly bounded, with at most linear growth in the state variables.

Assumption 2.5 formalizes how order intensities depend on the market maker's quotes relative to the reference price. Uncompetitive quotes reduce order flow as investors turn to alternative venues, boundedness expresses the finite size of the market, and the growth condition ensures consistency with the benchmark price while preventing unbounded gains or losses. Together, these conditions provide a coherent and realistic description of option order flow.

2.3 The market maker's optimization problem

2.3.1 Market maker inventories

By setting bid and ask quotes (β_t, α_t) on the option market, the market maker generates order arrivals modeled by the counting processes $(N_t^b)_{t \geq 0}$ and $(N_t^a)_{t \geq 0}$. The resulting option inventory process $(I_t)_{t \geq 0}$ records the net position in derivative contracts and evolves according to

$$\begin{cases} dI_t = dN_t^b - dN_t^a, \\ I_0 = i \in \mathbb{Z}. \end{cases} \quad (2.14)$$

To hedge this exposure, the market maker trades in the underlying through sequences of sell and buy orders $(\nu_i)_{i \in \mathbb{N}}$ and $(\tau_i)_{i \in \mathbb{N}}$ at mid-price p and spread s . Each intervention involves a quantity $\xi_i^- \in [0, \Phi_B(p - s/2)]$, \mathcal{F}_{ν_i} -measurable, or $\xi_i^+ \in [0, \Phi_A(+\infty)]$, \mathcal{F}_{τ_i} -measurable. The underlying inventory process $(Q_t)_{t \geq 0}$ is piecewise constant, updated only at these intervention times

$$\begin{cases} dQ_t = 0, \\ Q_{\nu_i^+} = Q_{\nu_i} - \xi_i^-, \\ Q_{\tau_i^+} = Q_{\tau_i} + \xi_i^+, \\ Q_0 = q \in \mathbb{R}. \end{cases} \quad (2.15)$$

Finally, these activities generate cash flows. Each underlying trade incurs a fixed cost $c > 0$, while option transactions occur at the bid and ask quotes $(\alpha_t, \beta_t)_{t \geq 0}$. The cash balance process $(X_t)_{t \geq 0}$ combines these option trades with the execution costs of underlying interventions:

$$\begin{cases} dX_t = \alpha_t dN_t^a - \beta_t dN_t^b, \\ X_{\nu_i^+} = X_{\nu_i} + P_B\left(P_{\nu_i} - \frac{1}{2}S_{\nu_i}, \xi_i^-\right) - c, \\ X_{\tau_i^+} = X_{\tau_i} - P_A\left(P_{\tau_i} + \frac{1}{2}S_{\tau_i}, \xi_i^+\right) - c, \\ X_0 = x \in \mathbb{R}. \end{cases} \quad (2.16)$$

P_B and P_A denote the execution cost functionals defined in (2.4)–(2.5).

2.3.2 Admissible strategies

Admissible strategies must satisfy both economic and technical requirements. On the option market, quotes must be ordered so that $\beta_t \leq \alpha_t$ at all times. On the underlying market, trades are constrained by liquidity and short-selling. Executed quantities must lie within the available depth and inventories cannot be driven below $-\Phi_A(+\infty)$, ensuring positions remain coverable. Finally, technical conditions guarantee well-posedness: quotes (β_t, α_t) must be \mathcal{F} -predictable, intervention times strictly increasing stopping times, and traded sizes \mathcal{F} -measurable. We also require square integrability of the number of interventions. Together, these conditions define the admissible set $\mathcal{A}(p, s, q)$ which is described in Definition (2.2).

Definition 2.2 (Admissible strategies). *Given a state (p, s, q) with mid-price p , spread s and inventory*

q , the set of admissible strategies is defined by

$$\begin{aligned}
\mathcal{A}(p, s, q) := & \{(\alpha, \beta, (\nu_i, \xi_i^-)_{i \geq 1}, (\tau_i, \xi_i^+)_{i \geq 1}) : \\
& (\alpha, \beta) \text{ } \mathcal{F}\text{-predictable, with } 0 \leq \beta \leq \alpha, \\
& (\nu_i), (\tau_i) \text{ increasing sequences of stopping times,} \\
& \xi_i^- \text{ } \mathcal{F}_{\nu_i}\text{-measurable, } \xi_i^+ \text{ } \mathcal{F}_{\tau_i}\text{-measurable,} \\
& 0 \leq \xi_i^- \leq \Phi_B(p - s/2), \quad 0 \leq \xi_i^+ \leq \Phi_A(+\infty), \\
& q - \xi_i^- > -\Phi_A(+\infty) \text{ for all } i, \\
& \mathbb{E}[H_T^\pm] < \infty, \quad \mathbb{E}[(H_T^\pm)^2] < \infty\}.
\end{aligned} \tag{2.17}$$

The next lemma shows that, under suitable initial conditions, bid and ask quotes remain nonnegative for all times and all admissible strategies.

Lemma 2.2 (Non-negativity preservation of bid–ask quotes). *Let $(P_t, S_t)_{t \geq 0}$ be the mid-price and spread defined in (2.11)–(2.13), and set $B_t := P_t - \frac{1}{2}S_t$ and $A_t := P_t + \frac{1}{2}S_t$. For any admissible strategy in $\mathcal{A}(p, s, q)$ with initial conditions*

$$B_0 = p - \frac{s}{2} \geq 0 \quad \text{and} \quad B_0 \geq D_0,$$

we have

$$\mathbb{P}(B_t \geq 0, A_t \geq 0, \forall t \geq 0) = 1.$$

Proof. By (2.13), we have $S_t \geq \delta \geq 0$ for all $t \geq 0$. Since $A_t = B_t + S_t$, it is enough to prove that $B_t \geq 0$ for all t . We first show that jumps cannot drive the bid below zero, and then verify that the inter-jump dynamics preserve non-negativity. From (2.11)–(2.13), an ask-side jump of size q produces $\Delta P_t = +\frac{1}{2}\Phi_A^{-1}(q)$ and $\Delta S_t = +\Phi_A^{-1}(q)$, which implies $\Delta B_t = 0$. A bid-side jump of size q yields $\Delta P_t = -\frac{1}{2}\Phi_B^{-1}(q)$ and $\Delta S_t = +\Phi_B^{-1}(q)$, so that $\Delta B_t = -\Phi_B^{-1}(q) \leq 0$. Moreover, by construction of the model and by admissibility of the strategy, the consumed quantity on the bid side is bounded above by the available depth: $q \leq \Phi_B(B_{t-})$. Using the monotonicity of the generalized inverse, this leads to

$$B_{t+} = B_{t-} - \Phi_B^{-1}(q) \geq B_{t-} - \Phi_B^{-1}(\Phi_B(B_{t-})) \geq 0,$$

which shows that jumps cannot make the bid negative.

On any interval $[u, t]$ with no jumps, the dynamics are given by $dP_t = dD_t = -rD_t dt$ and $dS_t = -\rho(S_t - \delta)dt$. Integrating yields

$$D_t = D_u e^{-r(t-u)}, \quad S_t = \delta + (S_u - \delta)e^{-\rho(t-u)},$$

and therefore

$$B_t = P_t - \frac{1}{2}S_t = (P_u - D_u) + D_u e^{-r(t-u)} - \frac{1}{2}\delta - \frac{1}{2}(S_u - \delta)e^{-\rho(t-u)}.$$

This yields the following explicit inter-jump representation:

$$B_t = B_u - D_u(1 - e^{-r(t-u)}) + \frac{1}{2}(S_u - \delta)(1 - e^{-\rho(t-u)}). \tag{2.18}$$

Since $S_u \geq \delta$ and $1 - e^{-\rho(\cdot)} \geq 0$, the last term is nonnegative, and we obtain

$$B_t \geq B_u - D_u(1 - e^{-r(t-u)}) \geq B_u - D_u.$$

In particular, on the initial inter-jump interval $[0, T_1)$, the assumption $B_0 \geq D_0$ implies $B_t \geq B_0 - D_0 \geq 0$.

Taken together, these arguments show that the bid remains nonnegative both at jump times and during inter-jump periods. As a result, $B_t \geq 0$ holds for all $t \geq 0$, and consequently $A_t = B_t + S_t \geq 0$ as well. \square

Lemma 2.2 ensures that the model never generates negative quotes. The condition $B_0 \geq D_0$ excludes initial states where a residual transient impact would not be reflected in the mid-price, which could drive the bid below zero during resilience.

2.3.3 Performance criterion and value function

Given an initial state $(t, x, q, i, e) \in [0, T] \times \mathbb{R}^3 \times (\mathbb{R}^+)^5$, the market maker chooses an admissible strategy γ to maximize expected cash holdings plus the liquidation value of residual positions, adjusted by running penalties. The penalty function g penalizes deviations from hedging, and the incentive function h enforces market-making activity. We assume that the penalty and incentive functions g and h satisfy Assumption 2.6.

Assumption 2.6 (Quadratic growth bounds for penalty and incentive functions). *There exist constants $C^{(g)}, C^{(h)} > 0$ such that for all (t, q, i, e) ,*

$$0 \leq g(t, q, i, e) \leq C^{(g)}(1 + q^2 + i^2 + \|e\|^2), \quad 0 \leq h(t, q, i, e) \leq C^{(h)}(1 + q^2 + i^2 + \|e\|^2).$$

Let $E_t = (P_t, D_t, S_t, \lambda_t^-, \lambda_t^+)$ with $E_0 = e$ and $\gamma \in \mathcal{A}(p, s, q)$. The objective function is defined by

$$J_\gamma(t, x, q, i, e) = \mathbb{E} \left[X_T + L(T, Q_T, I_T, P_T, S_T) - \int_t^T (g + h)(u, Q_u, I_u, E_u) du \right], \quad (2.19)$$

where the liquidation function is given by

$$\begin{aligned} L(T, q, i, p, s) &= \mathbb{1}_{\{q \geq 0\}} P_B(p - s/2, \min\{q, \Phi_B(p - s/2)\}) \\ &\quad - \mathbb{1}_{\{q < 0\}} P_A(p + s/2, -q) + i\varphi(p). \end{aligned} \quad (2.20)$$

By admissibility, a short position can always be covered at maturity, while excess long inventory beyond Φ_B is liquidated at zero value, discouraging overly aggressive hedging. The option inventory settles through φ .

A natural specification for g is to enforce delta-hedging, penalizing deviations between the actual underlying inventory q and the hedged position $i\partial_p b(t, e)$. Similarly, the role of h is to discourage purely passive quoting without generating trades, by penalizing intensities below a target level $\underline{\lambda} > 0$. Typical quadratic forms are therefore

$$g(t, q, i, e) = (q - i\partial_p b(t, e))^2, \quad h(t, q, i, e) := \left(\max\{0, \underline{\lambda} - (\lambda^a(t, \alpha, \beta, e) + \lambda^b(t, \alpha, \beta, e))\} \right)^2,$$

which combine hedging discipline with a requirement to sustain a minimum level of trading activity.

The market maker seeks to maximize her expected performance criterion:

$$\tilde{v}(t, x, q, i, e) := \sup_{\gamma \in \mathcal{A}(p, s, q)} J_\gamma(t, x, q, i, e)$$

Proposition 2.1 shows that x separates additively, so the effective value function is v , independent of initial cash.

Proposition 2.1 (Reduction by cash additivity). *For all (t, x, q, i, e) , the value function satisfies*

$$\tilde{v}(t, x, q, i, e) = x + v(t, q, i, e),$$

where v is given by

$$\begin{aligned} v(t, q, i, e) &= \sup_{\gamma \in \mathcal{A}} \mathbb{E} \left[\int_t^T [\alpha_u \lambda^a - \beta_u \lambda^b](u, \alpha_u, \beta_u, E_u) du \right. \\ &\quad + \sum_{i=1}^{+\infty} \mathbb{1}_{\{\nu_i \in [t, T]\}} \left(P_B(P_{\nu_i} - S_{\nu_i}/2, \xi_i^-) - c \right) \\ &\quad + \sum_{i=1}^{+\infty} \mathbb{1}_{\{\tau_i \in [t, T]\}} \left(-P_A(P_{\tau_i} + S_{\tau_i}/2, \xi_i^+) - c \right) \\ &\quad \left. - \int_t^T (g + h)(u, Q_u, I_u, E_u) du + L(T, Q_T, I_T, P_T, S_T) \right] \end{aligned} \quad (2.21)$$

Proof. Integrating the dynamics of the cash process over $[t, T]$ and using the compensated martingales for the option order arrival processes then taking the expectations yields the required result. \square

3 Model consistency and well-posedness

The purpose of this section is to examine possible forms of arbitrage and manipulation within our framework, in order to ensure the model's internal consistency and to identify whether manipulative strategies could arise as optimal candidates.

3.1 Underlying market: absence of arbitrage and manipulation

In Lemma 3.3, we first rule out arbitrage in the simplest case of an instantaneous round-trip. This result confirms that immediate buy–sell cycles are loss-making, with the spread acting as a lower bound on trading frictions.

Lemma 3.3 (No instantaneous round-trip). *For any market state (p, s) , any admissible order book satisfying Assumptions 2.1–2.2, and any admissible trade size $q \geq 0$, the execution cost difference satisfies*

$$P_B(p - s/2, q) - P_A(p + s/2, q) - 2c \leq -\delta q - 2c < 0,$$

where $2c$ denotes the total fixed cost of the two transactions.

Proof. From the integral representations of P_A and P_B (Equations 2.7–2.6), we have

$$\begin{aligned} P_B(p - s/2, q) - P_A(p + s/2, q) &= (p - s/2)q - \int_0^q \Phi_B^{-1}(y) dy - (p + s/2)q + \int_0^q \Phi_A^{-1}(y) dy \\ &= \int_0^q \left[-s - \Phi_B^{-1}(y) - \Phi_A^{-1}(y) \right] dy. \end{aligned}$$

Since $s \geq \delta$, the right-hand side is bounded above by $-\delta q$. An instantaneous round-trip involves two trades, each incurring a fixed cost c . This contributes an additional $-2c$, which yields the strict inequality. \square

Beyond this elementary case, one must also rule out more elaborate manipulations over finite horizons. In line with Huberman and Stanzl's no-dynamic-arbitrage condition [13], the next result shows that microstructure frictions in our model prevent any admissible strategy from generating a strictly positive pure execution profit.

Proposition 3.2 (No round-trip arbitrage). *Let $(\nu_i, \xi_i^-)_{i \geq 1}$ and $(\tau_j, \xi_j^+)_{j \geq 1}$ denote the sequences of admissible trades executed by the market maker on the underlying up to a finite horizon T , where ξ_i^- (resp. ξ_j^+) is the size of the i -th sale (resp. j -th purchase). Define the total traded volume and the number of interventions by*

$$\mathcal{V}_T := \sum_{\nu_i \leq T} \xi_i^- + \sum_{\tau_j \leq T} \xi_j^+, \quad \bar{H}_T := \sum_{\nu_i \leq T} 1 + \sum_{\tau_j \leq T} 1.$$

Then, for any round trip completed over $[0, T]$ such that $Q_T = Q_0$, the pure execution profit and loss Π satisfies

$$\Pi_T - \Pi_0 \leq -\frac{\delta}{2} \mathcal{V}_T - c \bar{H}_T \leq 0,$$

with strict inequalities whenever at least one trade is executed.

Proof. Let C_T be the cumulative underlying cash flows up to T . Using (2.6)–(2.7) at pre-trade quotes,

$$\begin{aligned}
C_T &= \sum_{i=1}^{+\infty} \mathbb{1}_{\{\nu_i \in [0, T]\}} \left(P_B(P_{\nu_i} - \tfrac{1}{2}S_{\nu_i}, \xi_i^-) - c \right) + \sum_{j=1}^{+\infty} \mathbb{1}_{\{\tau_j \in [0, T]\}} \left(-P_A(P_{\tau_j} + \tfrac{1}{2}S_{\tau_j}, \xi_j^+) - c \right) \\
&= \sum_{i=1}^{+\infty} \left(\mathbb{1}_{\{\nu_i \in [0, T]\}} \left(P_{\nu_i} - \tfrac{1}{2}S_{\nu_i} \right) \xi_i^- - \int_0^{\xi_i^-} \Phi_B^{-1}(y) dy \right) \\
&\quad - \sum_{j=1}^{+\infty} \left(\mathbb{1}_{\{\tau_j \in [0, T]\}} \left(P_{\tau_j} + \tfrac{1}{2}S_{\tau_j} \right) \xi_j^+ + \int_0^{\xi_j^+} \Phi_A^{-1}(y) dy \right) - c\bar{H}_T \\
&= \sum_{\nu_i \leq T} P_{\nu_i} \xi_i^- - \sum_{\tau_j \leq T} P_{\tau_j} \xi_j^+ - \frac{1}{2} \sum_{\nu_i \leq T} S_{\nu_i} \xi_i^- - \frac{1}{2} \sum_{\tau_j \leq T} S_{\tau_j} \xi_j^+ - \sum_{\nu_i \leq T} \int_0^{\xi_i^-} \Phi_B^{-1}(y) dy \\
&\quad - \sum_{\tau_j \leq T} \int_0^{\xi_j^+} \Phi_A^{-1}(y) dy - c\bar{H}_T.
\end{aligned}$$

We note that:

$$\sum_{\nu_i \leq T} P_{\nu_i} \xi_i^- - \sum_{\tau_j \leq T} P_{\tau_j} \xi_j^+ = - \int_{(0, T]} P_u^- dQ_u,$$

and introduce the pure execution P&L:

$$\Pi_T = C_T + \int_{(0, T]} P_u^- dQ_u + Q_0 P_0.$$

The additional term $\int_{(0, T]} P_u^- dQ_u$ accounts for the neutralisation of the variations of the mid-price. The total pure execution profit and loss can be expressed as

$$\Pi_T = -\frac{1}{2} \sum_{\nu_i \leq T} S_{\nu_i} \xi_i^- - \frac{1}{2} \sum_{\tau_j \leq T} S_{\tau_j} \xi_j^+ - \sum_{\nu_i \leq T} \int_0^{\xi_i^-} \Phi_B^{-1}(y) dy - \sum_{\tau_j \leq T} \int_0^{\xi_j^+} \Phi_A^{-1}(y) dy - c\bar{H}_T + Q_0 P_0. \quad (3.1)$$

Spread satisfy $S_t \geq \delta$ for all t , $\Phi_B^{-1}, \Phi_A^{-1} \geq 0$ by definition and $\Pi_0 = Q_0 P_0$. Therefore, every term on the right-hand side of (3.1) is nonpositive, which yields

$$\Pi_T - \Pi_0 \leq -\frac{\delta}{2} \left(\sum_{\nu_i \leq T} \xi_i^- + \sum_{\tau_j \leq T} \xi_j^+ \right) - c\bar{H}_T = -\frac{\delta}{2} \mathcal{V}_T - c\bar{H}_T.$$

Since $\delta \geq 0$ and $c > 0$, the inequality is strict whenever at least one trade occurs, which completes the proof. \square

Taken together, these results show that neither instantaneous nor multi-step round trips can be profitable once P&L is evaluated on a pure execution basis, which isolates trading costs from inventory revaluations. This condition parallels the classical notion of no price manipulation of Huberman and Stanzl [13], later extended to limit-order-book models [10, 1]. Still, one must also exclude transaction-triggered strategies, where preliminary trades are used to influence prices before unwinding a position. The next proposition shows that such manipulations are likewise ruled out in our framework.

Proposition 3.3 (No transaction-triggered price manipulation). *Let $0 \leq \nu < \tau \leq T$. Consider a two-step trading scheme consisting of a pre-trade followed by a target trade: a pre-purchase of size $z \geq 0$ at time ν , followed by a target sale of size $q \geq 0$ at time τ . Denote by $\Pi_T^{(z)}$ the pure execution profit and loss generated by this pair of trades, and by $\Pi_T^{(0)}$ the corresponding P&L when the pre-trade is omitted.*

Then, for any admissible z, q , the incremental profit satisfies

$$\Pi_T^{(z)} - \Pi_T^{(0)} \leq -\frac{\delta}{2}z - c < 0.$$

In particular, pre-trading to influence prices prior to execution of a target trade necessarily reduces the pure execution P&L.

Proof. According to Equation 3.1, with pre-trade, the pure execution P&L is

$$\Pi_T^{(z)} = -\frac{1}{2}S_\nu z - \int_0^z \Phi_A^{-1}(y)dy - \frac{\eta}{2}z\Phi_A^{-1}(z) - \frac{1}{2}S_\tau^{(z)}q - \int_0^q \Phi_B^{-1}(y)dy - \frac{\eta}{2}q\Phi_B^{-1}(q) - 2c.$$

In the absence of the pre-trade, the pure execution P&L reduces to

$$\Pi_T^{(0)} = -\frac{1}{2}S_\tau^{(0)}q - \int_0^q \Phi_B^{-1}(y)dy - \frac{\eta}{2}q\Phi_B^{-1}(q) - c.$$

Subtracting the two expressions yields

$$\Pi_T^{(z)} - \Pi_T^{(0)} = -\frac{1}{2}S_\nu z - \int_0^z \Phi_A^{-1}(y)dy - \frac{\eta}{2}z\Phi_A^{-1}(z) - \frac{1}{2}q(S_\tau^{(z)} - S_\tau^{(0)}) - c.$$

Since $S_t \geq \delta$ for all t . Since the inequality $S_\tau^{(z)} \geq S_\tau^{(0)}$ holds pathwise because the spread dynamics (2.13) are monotone in their initial condition. It follows that

$$\Pi_T^{(z)} - \Pi_T^{(0)} \leq -\frac{\delta}{2}z - c < 0.$$

The strict inequality holds because $c > 0$, which completes the proof. \square

The framework excludes strategies based on preliminary trades to influence later executions: such manipulative tactics are necessarily loss-making, which ensures economic consistency by preventing price distortions through strategically timed trading and preserving the robustness of the market-making model.

Corollary 3.1 (No arbitrage and no price manipulation on the underlying). *Under the assumptions of Lemma 3.3, Proposition 3.2 and Proposition 3.3, the proposed market model precludes any form of arbitrage or price manipulation on the underlying. In particular, every admissible trading strategy necessarily yields a nonpositive pure execution P&L.*

3.2 Coupled markets and terminal manipulation

In the standalone option market, trades do not affect quotes and the spread ($\beta_t \leq \alpha_t$) rules out profitable round-trips: any purchase at the ask must be unwound at the bid, yielding a nonpositive outcome. From the market maker's perspective, attempts to manipulate the underlying to shift the reference price and option order intensities are also ineffective, as intensities are uniformly bounded by Assumption 2.5 and frictions in the underlying make such schemes unprofitable.

In coupled markets, identifying manipulative strategies is more intricate. A salient possibility in our setting is end-of-maturity manipulation: near T , a market maker holding a nonzero terminal option inventory i may attempt to move the underlying to tilt the option payoff in her favor. Proposition 3.4 quantifies this risk under a Lipschitz payoff and uniformly bounded order-book depth. From a control standpoint, however, this need not undermine well-posedness: the resulting gains are at most bounded at maturity, so the optimization problem and its value function remain viable. A fuller analysis of manipulation on coupled markets is an interesting direction for future research.

Proposition 3.4 (Terminal no-arbitrage condition in the coupled market). *Let \overline{H}_T be the total number of interventions of the market maker on underlying market. Assume that there exists $m > 0$ such that $f_A(u) \leq m$ for all $u \geq 0$ and $f_B(u) \leq m$ for all $u \leq 0$, and that the payoff function φ is $L^{(\varphi)}$ -Lipschitz in p . Let $i \in \mathbb{Z}$ denote the terminal option inventory and $\Delta_T := |P_T^{(H)} - P_T^{(0)}|$ the terminal price distortion generated by underlying impulses. Then the terminal P&L difference satisfies*

$$\Delta \hat{\Pi}_T := (\Pi_T^{(H)} + i\varphi(P_T^{(H)})) - (\Pi_T^{(0)} + i\varphi(P_T^{(0)})) \leq -(\delta m - |i|L^{(\varphi)})\Delta_T - \frac{2m}{\overline{H}_T}\Delta_T^2 - c\overline{H}_T. \quad (3.2)$$

Proof. We compare two paths driven by the same exogenous order flow: one with the market maker's underlying impulses (superscript (H)) and one without (superscript (0)). By (2.11)–(2.12), a buy impulse of size ξ at time u shifts the mid price at time T by

$$\left(\frac{\eta}{2} + \frac{1-\eta}{2}e^{-r(T-u)}\right)\Phi_A^{-1}(\xi),$$

while a sell impulse of the same size shifts it by the negative of

$$\left(\frac{\eta}{2} + \frac{1-\eta}{2}e^{-r(T-u)}\right)\Phi_B^{-1}(\xi).$$

Since $\frac{\eta}{2} + \frac{1-\eta}{2}e^{-r(T-u)} \leq \frac{1}{2}$, summing the contributions of all impulses and taking absolute values yields the elementary bound

$$\Delta_T := |P_T^{(H)} - P_T^{(0)}| \leq \frac{1}{2} \left(\sum_{\tau_j \leq T} \Phi_A^{-1}(\xi_j^+) + \sum_{\nu_i \leq T} \Phi_B^{-1}(\xi_i^-) \right). \quad (3.3)$$

We now turn to the cashflow side. Using the pure execution representation (3.1) and the bound $S_t \geq \delta$, the difference between the impacted and non-impacted pure execution P&L satisfies

$$\Pi_T^{(H)} - \Pi_T^{(0)} \leq -\frac{\delta}{2} \sum_{\nu_i \leq T} \xi_i^- - \frac{\delta}{2} \sum_{\tau_j \leq T} \xi_j^+ - \sum_{\nu_i \leq T} \int_0^{\xi_i^-} \Phi_B^{-1}(y) dy - \sum_{\tau_j \leq T} \int_0^{\xi_j^+} \Phi_A^{-1}(y) dy - c\overline{H}_T.$$

Finally, since φ is L -Lipschitz in p , we have

$$i(\varphi(P_T^{(H)}) - \varphi(P_T^{(0)})) \leq |i|L^{(\varphi)}\Delta_T.$$

Adding this to the previous inequality gives the stated bound

$$\begin{aligned} \Delta \hat{\Pi}_T &\leq |i|L^{(\varphi)}\Delta_T - \frac{\delta}{2} \sum_{\nu_i \leq T} \xi_i^- - \frac{\delta}{2} \sum_{\tau_j \leq T} \xi_j^+ \\ &\quad - \sum_{\nu_i \leq T} \int_0^{\xi_i^-} \Phi_B^{-1}(y) dy - \sum_{\tau_j \leq T} \int_0^{\xi_j^+} \Phi_A^{-1}(y) dy - c\overline{H}_T, \end{aligned} \quad (3.4)$$

which holds pathwise.

In particular, if $|i|L^{(\varphi)} \leq \delta m$ and at least one impulse occurs so that $\overline{H}_T \geq 1$, the right-hand side is strictly negative, and the claim follows.

We have the local depth bound $0 \leq f_A(u) \leq m$ (and symmetrically for f_B). Hence $\Phi_A(x) \leq mx$ for every admissible quantity x , which implies $\Phi_A^{-1}(y) \geq y/m$ for $y \in [0, \Phi_A(x)]$, and analogously for Φ_B^{-1} .

$$\Phi_A^{-1}(y) \geq \frac{y}{m}, \quad \Phi_B^{-1}(y) \geq \frac{y}{m}.$$

Substituting into (3.3) gives

$$\Delta_T \leq \frac{1}{2m} \left(\sum_{\tau_j \leq T} \xi_j^+ + \sum_{\nu_i \leq T} \xi_i^- \right). \quad (3.5)$$

For any admissible $\xi \geq 0$, the integral inequality $\int_0^\xi \Phi^{-1}(y) dy \geq \xi^2/(2m)$. Applying it term by term and grouping the sums gives

$$\Pi_T^{(H)} - \Pi_T^{(0)} \leq -\frac{\delta}{2} \left(\sum_{\nu_i \leq T} \xi_i^- + \sum_{\tau_j \leq T} \xi_j^+ \right) - \frac{1}{2m} \left(\sum_{\nu_i \leq T} (\xi_i^-)^2 + \sum_{\tau_j \leq T} (\xi_j^+)^2 \right) - c\bar{H}_T.$$

From Cauchy—Schwarz inequality, $\sum(\xi)^2 \geq \frac{(\sum \xi)^2}{\bar{H}_T}$. Using this and (3.5) yields

$$\Pi_T^{(H)} - \Pi_T^{(0)} \leq -\delta m \Delta_T - \frac{2m}{\bar{H}_T} \Delta_T^2 - c\bar{H}_T.$$

Hence Inequality 3.4 leads to

$$\Delta \hat{\Pi}_T \leq -(\delta m - |i|L^{(\varphi)})\Delta_T - \frac{2m}{\bar{H}_T} \Delta_T^2 - c\bar{H}_T,$$

which holds pathwise. In particular, the inequality holds for all $\bar{H}_T \geq 1$. Hence, the right-hand side is strictly negative, and the claim follows. \square

Proposition 3.4 shows that an end-of-maturity manipulation channel does exist, yet under natural Lipschitz and depth bounds the associated gains remain bounded, so that the control problem retains its well-posedness, which will be the focus of the next section.

3.3 Well-posedness of the control problem

Before introducing the numerical method, we first need to ensure that the control problem is meaningful and non-trivial. In particular, we must verify that the formulation given in (2.21) defines a well-posed optimization problem whose value function is finite. Establishing this result guarantees that the model does not admit degenerate or unbounded solutions, and thus provides a solid foundation for the numerical analysis developed in the next section. The following theorem formalizes these properties.

Theorem 3.1 (Well-posedness and quadratic bounds). *Under Assumptions 2.1, 2.2, 2.3, 2.4, 2.5 and 2.6, the value function v in (2.21) is well defined and finite for all $(t, q, i, e) \in [0, T] \times \mathbb{R} \times \mathbb{Z} \times (\mathbb{R}^+)^5$. Moreover, there exist constants $C_T^{(-)}, C_T^{(+)} > 0$, depending only on T and model parameters, such that*

$$-C_T^{(-)}(1 + q^2 + i^2 + \|e\|^2) \leq v(t, q, i, e) \leq C_T^{(+)}(1 + |i|^2 + \|e\|^2).$$

The argument relies on establishing stability and moment bounds for the model's driving processes, which are then used to control the expected performance functional. Lemma A.4 provides moment estimates for the option inventory, while Lemma A.5 ensures uniform bounds for the Hawkes intensities and their counting processes. These results are propagated to the full market state in Proposition A.5, which guarantees integrability and uniform moment control for all state variables on $[0, T]$. Based on these estimates, a global lower bound for the value function is obtained in Proposition A.6, and an upper bound in Proposition A.7. Together, these results establish the finiteness and quadratic growth of the value function, proving Theorem 3.1. The detailed arguments are presented in Appendix A.

4 Numerical methodology and experiments

We propose a neural, simulation-based approach to approximate optimal quoting and hedging in our market-making model. Inspired by Deep Hedging [6], we use direct policy search: a network maps the current state to quotes and an inventory-normalized hedge, and its parameters are trained on a differentiable simulator to maximize the objective function.

4.1 Simulation and neural policy learning

Discrete-time model and simulator

Recalling that $T > 0$ is the time horizon for the market making problem. We discretize the interval $[0, T]$ into $N > 0$ time steps. Let $\delta_t = T/N$ denote the time step size, and define the discrete time grid as

$$\mathbb{T} = \{t_0 = 0, \dots, t_i = i\delta_t, \dots, t_N = T\}.$$

Let π denote the agent's discrete policy, which prescribes a decision at each point in the grid \mathbb{T} . For every $t_i \in \mathbb{T}$, the market maker chooses the controls $(\Gamma^\pi, \alpha^\pi, \beta^\pi)_{t_i}$. Inspired by the work of Buehler et al. [6], we use an inventory-normalized hedge ratio $\Gamma^\pi_{t_i} \in [-1, 1]$.

Here, I^π denotes the discretized option inventory process, which is controlled by the market maker's option quotes $\beta^\pi_{t_i}$ and $\alpha^\pi_{t_i}$. These quotes influence the arrival intensities of the counting processes N^a and N^b . Define the option-market counting increments

$$\Delta N^a_{t_i} := N^a_{t_{i+1}} - N^a_{t_i}, \quad \Delta N^b_{t_i} := N^b_{t_{i+1}} - N^b_{t_i}, \quad (4.1)$$

then the option inventory position evolves as

$$I^\pi_{t_{i+1}} = I^\pi_{t_i} + \Delta N^a_{t_i} - \Delta N^b_{t_i},$$

and the hedged underlying

$$Q^\pi_{t_{i+1}} = \Gamma^\pi_{t_i} I^\pi_{t_{i+1}},$$

with initial conditions $I^\pi_0 = i \in \mathbb{Z}$ and $Q^\pi_0 = q \in \mathbb{R}$. In the following, we write $\Delta Q^\pi_{t_i} := Q^\pi_{t_{i+1}} - Q^\pi_{t_i}$. In our numerical implementation, the underlying inventory adjustment is clipped to remain within admissible trading volumes, according to:

$$\Delta Q^\pi_{t_i} = \begin{cases} \max \left\{ -\Phi_B(P^\pi_{t_i} - S^\pi_{t_i}/2), \Delta Q^\pi_{t_i} \right\} & \Delta Q^\pi_{t_i} < 0, \\ \min \left\{ \Phi_A(+\infty), \Delta Q^\pi_{t_i} \right\} & \Delta Q^\pi_{t_i} > 0, \\ 0 & \text{otherwise.} \end{cases}$$

This ensures that the market maker does not attempt to trade beyond available liquidity. In practice, Φ_A and Φ_B are chosen such that the desired hedging action rarely reaches these bounds, reflecting realistic market behavior.

We now express the evolution of the market maker's cash process X^π along the grid. It is given by

$$\begin{aligned} X^\pi_{t_{i+1}} = & X^\pi_{t_i} + \alpha^\pi_{t_i} \Delta N^a_{t_i} - \beta^\pi_{t_i} \Delta N^b_{t_i} + P_B \left(P^\pi_{t_i} - \frac{1}{2} S^\pi_{t_i}, (\Delta Q^\pi_{t_i})^- \right) \\ & - P_A \left(P^\pi_{t_i} + \frac{1}{2} S^\pi_{t_i}, (\Delta Q^\pi_{t_i})^+ \right) - c \mathbb{1}_{\{\Delta Q^\pi_{t_i} \neq 0\}}. \end{aligned} \quad (4.2)$$

The inventory-normalized hedge ratio Γ^π allows the policy to be specified in relative terms, ensuring scalability across inventory levels with fewer parameters and improved stability. Together with the discretized dynamics, this yields a self-contained simulator on the grid \mathbb{T} , directly amenable to Monte Carlo estimation and gradient-based training. All other state variables are discretized by standard schemes. Since their dynamics are driven by Hawkes arrivals, this introduces no additional numerical difficulty. We next specify the measurability of the discrete processes and the information available to the agent, completing the definition of admissible strategies in discrete time.

Measurability of processes

Let $t_i \in \mathbb{T}$ denote a discrete time point, we define the state of the system at that time as:

$$Z_{t_i} := (P_{t_i}^\pi, D_{t_i}^\pi, S_{t_i}^\pi, \lambda_{t_i}^{\pi,-}, \lambda_{t_i}^{\pi,+}, I_{t_i}^\pi, Q_{t_i}^\pi, N_{t_i}^a, N_{t_i}^b, N_{t_i}^-, N_{t_i}^+).$$

We introduce the filtration $(\mathcal{F}_{t_i})_{i \in \mathbb{N}}$ describing the information available up to time t_i :

$$\mathcal{F}_{t_i} := \sigma(Z_{t_0}, \dots, Z_{t_i}).$$

At each time t_i , the market state variables $(P_{t_i}^\pi, D_{t_i}^\pi, S_{t_i}^\pi, \lambda_{t_i}^{\pi,-}, \lambda_{t_i}^{\pi,+})$ are \mathcal{F}_{t_i} -measurable, as they are entirely determined by past realizations. Similarly, the agent's controls $(\alpha_{t_i}^\pi, \beta_{t_i}^\pi, \Gamma_{t_i}^\pi)$ are chosen on the basis of \mathcal{F}_{t_i} , and are therefore \mathcal{F}_{t_i} -measurable. By contrast, the order-flow increments $(\Delta N_{t_i}^a, \Delta N_{t_i}^b, \Delta N_{t_i}^-, \Delta N_{t_i}^+)$ are only revealed over the interval $[t_i, t_{i+1})$ and hence belong to $\mathcal{F}_{t_{i+1}}$. These properties ensure the internal consistency of the discrete-time model and clarify the informational structure available to the agent at each step of the simulation.

Objective function

Within the discrete-time setting and the measurability framework established above, we now introduce a tractable version of the objective originally defined in (2.19). For a given policy π , the corresponding criterion reads

$$J_\pi(x, q, i, e) = \mathbb{E} \left[X_T^\pi - \delta_t \sum_{k=0}^{N-1} (g + h)(t_{k+1}, Q_{t_{k+1}}^\pi, I_{t_{k+1}}^\pi, E_{t_{k+1}}^\pi) + L(T, Q_T^\pi, I_T^\pi, P_T^\pi, S_T^\pi) \right]. \quad (4.3)$$

The function L denotes the terminal liquidation function introduced in (2.20), while g and h collect the running penalty terms reflecting hedging frictions and quoting incentives. The vector $e = (p, d, s, \lambda_0^-, \lambda_0^+)$ specifies the initial exogenous market configuration.

Using the same arguments than in Proposition 2.1, the objective function in (4.3) admits the following decomposition:

$$\begin{aligned} J_\pi(x, q, i, e) = x + \mathbb{E} & \left[\delta_t \sum_{k=0}^{N-1} \left(\alpha_{t_k}^\pi \lambda^a - \beta_{t_k}^\pi \lambda^b \right) (t_k, \alpha_{t_k}^\pi, \beta_{t_k}^\pi, E_{t_k}^\pi) \right. \\ & + \sum_{k=0}^{N-1} \left(P_B \left(P_{t_k}^\pi - \frac{1}{2} S_{t_k}^\pi, (\Delta Q_{t_k}^\pi)^- \right) \right. \\ & \left. \left. - P_A \left(P_{t_k}^\pi + \frac{1}{2} S_{t_k}^\pi, (\Delta Q_{t_k}^\pi)^+ \right) - c \mathbb{1}_{\{\Delta Q_{t_k}^\pi \neq 0\}} \right) \right. \\ & \left. - \delta_t \sum_{k=0}^{N-1} (g + h)(t_{k+1}, Q_{t_{k+1}}^\pi, I_{t_{k+1}}^\pi, E_{t_{k+1}}^\pi) + L(T, Q_T^\pi, I_T^\pi, P_T^\pi, S_T^\pi) \right] \end{aligned}$$

The problem faced by the agent is to select an admissible policy π^* that maximizes the expected objective, namely

$$\pi^* = \arg \max_{\pi \in \Pi} J_\pi(x, q, i, e).$$

This optimization naturally leads to the definition of the discrete-time value function,

$$\hat{v}(t, q, i, e) = \sup_{\pi \in \Pi} J_\pi(x, q, i, e)$$

which represents the maximal attainable performance given the current state.

Training by Monte Carlo and autodifferentiation

The policy π_θ is parameterized by a neural network with weights θ . Given θ , we simulate M independent market trajectories under the discrete-time dynamics. For each trajectory, the agent sequentially selects at each step a bid quote, a spread (with the ask quote set to bid plus spread, ensuring $\text{bid} \leq \text{ask}$), and a normalized hedge ratio Γ^π . The resulting controls generate order arrivals and inventory updates, and the simulator propagates the impacts so that all state variables evolve consistently with the model. The realized objective values $J_\pi^{(m)}(\theta)$ are then collected, and their empirical average

$$\hat{J}_\pi(\theta) := \frac{1}{M} \sum_{m=1}^M J_\pi^{(m)}(\theta)$$

provides a Monte Carlo estimate of the expected objective.

To render the simulator almost everywhere differentiable, we assume that the bid and ask intensity functions are differentiable almost everywhere. Under this assumption, automatic differentiation yields unbiased gradients $\nabla_\theta \hat{J}_\pi(\theta)$, which are used to update the network parameters via stochastic gradient ascent. In practice, we employ a standard multilayer perceptron architecture, together with feature transformations that enhance training stability and improve sample efficiency. Potential pointwise non-differentiabilities are handled by standard machine-learning techniques, ensuring that gradient-based optimization remains effective in practice.

4.2 Experimental results

4.2.1 Parameters

Underlying market

We consider a trading horizon of five trading weeks, corresponding to

$$T = \frac{25}{252} \text{ years.}$$

In our numerical experiments we assume that the agent has ten trading opportunities per day, so that the interval $[0, T]$ is discretized into $N = 250$ steps. To describe the underlying market, we adopt a linear order book representation with constant depth densities on both sides up to finite cutoffs. On the bid side, liquidity is specified as

$$f_B(u) = c_B \mathbf{1}_{\{-U_B \leq u \leq 0\}},$$

yielding the cumulative depth $\Phi_B(x) = c_B x$ for $0 \leq x \leq U_B$, with generalized inverse $\Phi_B^{-1}(y) = y/c_B$ for $y \in [0, c_B U_B]$. On the ask side, we similarly set

$$f_A(u) = c_A \mathbf{1}_{\{0 \leq u \leq U_A\}},$$

so that $\Phi_A(x) = c_A x$ for $0 \leq x \leq U_A$, saturating at $\Phi_A(+\infty) = c_A U_A$, with inverse $\Phi_A^{-1}(y) = y/c_A$ for $y \in [0, c_A U_A]$. These explicit forms make execution costs analytically tractable. If $b > 0$ denotes the best bid and $a > 0$ the best ask, then selling a quantity $q^- \in [0, c_B U_B]$ generates the revenue

$$P_B(b, q^-) = \int_{-q^-/c_B}^0 (b + u) c_B du = bq^- - \frac{(q^-)^2}{2c_B},$$

while purchasing a quantity $q^+ \in [0, c_A U_A]$ entails the cost

$$P_A(a, q^+) = \int_0^{q^+/c_A} (a + u) c_A du = aq^+ + \frac{(q^+)^2}{2c_A}.$$

We fix linear book densities at $(c_A, c_B) = (100, 100)$ and finite depths $(U_A, U_B) = (0.5, 0.5)$, implying a total liquidity of 50 units on each side. At the start of the experiment, the underlying mid-price is

set to $P_0 = 100$, with an initial bid-ask spread $S_0 = 0.10$ and a minimal admissible spread $\delta = 0.02$. Furthermore, we assume that $D_0 = 0$.

For simplicity, we model exogenous buy and sell orders arrive at the best quotes according to independent homogeneous Poisson processes with intensities $\lambda^- > 0$ and $\lambda^+ > 0$. This tractable specification is particularly convenient for the construction of the option reference price, since it yields explicit expressions for the variance of the mid-price dynamics. Each order is associated with a random mark $M_t^\pm \in [0, 1]$ representing the fraction of available depth consumed. We assume $M^- \sim \text{Beta}(a_-, b_-)$ and $M^+ \sim \text{Beta}(a_+, b_+)$, independent of the Poisson clocks and of the past. The Beta distribution provides a convenient way to generate bounded marks in $[0, 1]$, with the parameters controlling the typical trade size. In our calibration we use $(a_\pm, b_\pm) = (2, 5)$, producing a majority of small trades.

Exogenous order flow is calibrated to produce, on average, 30 events per trading day on each side. Annualized with 252 trading days, this corresponds to $\lambda^+ = \lambda^- = 30 \times 252 = 7560$ arrivals per year. We assume that 30% of price impact is permanent, while the remaining 70% mean-reverts at resilience $r = 60$ events per day (that is, $r = 60 \times 252 = 15,120$ annually). The spread reverts to its lower bound at rate $\rho = 200$ per day, or $\rho = 200 \times 252 = 50,400$ annually. For the baseline calibration, we normalize the fixed cost of each impulse trade in the underlying to zero, i.e., $c = 0$. All parameter values are collected in Table 1.

P_0	D_0/Q_0	S_0	δ	U_A, U_B	c_A, c_B	$\lambda^{+/-}$	(a_\pm, b_\pm)	η	r	ρ	T
100	0	0.10	0.02	0.5	100	7560	(2, 5)	0.3	15,120	50,400	25/252

Table 1: Baseline parameters of the underlying market.

Option market

The market maker is assumed to provide liquidity on a European call option with strike $K = 98$ and payoff

$$(x - K)_+ = \max(x - K, 0).$$

To model the option order arrival rates, we specify the following functional form for the bid and ask intensities, which allows for flexible calibration through scale and shift parameters $\mu_a, \mu_b \in \mathbb{R}$. Specifically, for $(t, \alpha, \beta, e) \in [0, T] \times \mathbb{R}^5$, we set

$$\lambda^b(t, \alpha, \beta, e) = \bar{\lambda}^b \sigma(-k_b(b(t, e) - \beta) + \mu_b), \quad \lambda^a(t, \alpha, \beta, e) = \bar{\lambda}^a \sigma(k_a(b(t, e) - \alpha) + \mu_a), \quad (4.4)$$

where $\sigma(x) = (1 + e^{-x})^{-1}$ is the logistic sigmoid, $\bar{\lambda}^a, \bar{\lambda}^b > 0$ and $k_a, k_b > 0$. As before, b denotes the option reference price. Its construction is detailed below.

On a daily scale, we set $\bar{\lambda}^a = \bar{\lambda}^b = 200$ arrivals at competitive quotes. Annualized, this corresponds to $\bar{\lambda}^a = \bar{\lambda}^b = 200 \times 252 = 50,400$ arrivals per year. The exponential slopes are fixed at $k_a = k_b = 50$, so that intensities decay rapidly once quotes deviate from the efficient level by more than a few ticks. We set the initial option inventory to $I_0 = 0$, and summarize the full set of parameters of the option order flow model in Table 2.

I_0	$\bar{\lambda}^a$	$\bar{\lambda}^b$	k_a, k_b	K
0	50,400	50,400	50	98

Table 2: Baseline parameters of the option order flow model.

Reference price and penalties

The option reference price b is defined as a Black–Scholes value computed with an effective volatility extracted from the order-driven mid-price dynamics. The idea is to match the instantaneous variance of the compound-Poisson mid-price process with that of a diffusion, and then to value the option under the corresponding proxy diffusion. With independent Poisson flows on the bid and ask at intensities (λ^-, λ^+) and marks $M_t^\pm \in [0, 1]$, a jump on the bid shifts the mid by $-\frac{1}{2}\Phi_B^{-1}(M_t^- \Phi_B(B_{t-}))$ and a jump on the ask by $+\frac{1}{2}\Phi_A^{-1}(M_t^+ \Phi_A(+\infty))$. In the linear book,

$$\Phi_A(+\infty) = c_A U_A, \quad \Phi_A^{-1}(y) = \frac{y}{c_A} \Rightarrow \Phi_A^{-1}(M^+ \Phi_A(+\infty)) = M^+ U_A,$$

and similarly $\Phi_B^{-1}(M^- \Phi_B(\cdot)) = M^- U_B$ (with U_B small in price units, $B_{t-} \geq U_B$ is natural in practice). Over $[0, t]$, the quadratic variation is the sum of squared jumps. Taking expectations and using that the numbers of bid/ask arrivals are Poisson with means $\lambda^\pm t$, while the marks are i.i.d. and independent of the counts, the expected quadratic variation is the mean count times the mean squared jump size on each side. Hence

$$\mathbb{E}[[P]_t] = \frac{1}{4} \left(\lambda^+ \mathbb{E}[(M^+ U_A)^2] + \lambda^- \mathbb{E}[(M^- U_B)^2] \right) t,$$

so the variance rate of the mid is

$$\nu = \frac{1}{t} \mathbb{E}[[P]_t] = \frac{1}{4} \left(\lambda^+ U_A^2 \mathbb{E}[(M^+)^2] + \lambda^- U_B^2 \mathbb{E}[(M^-)^2] \right).$$

As $M^\pm \sim \text{Beta}(a_\pm, b_\pm)$, we have

$$\mathbb{E}[(M^\pm)^2] = \frac{a_\pm(a_\pm + 1)}{(a_\pm + b_\pm)(a_\pm + b_\pm + 1)}.$$

The effective volatility is then defined as

$$\sigma_{\text{eff}}(t, e) = \frac{\sqrt{\nu}}{P_t} = \frac{1}{2P_t} \sqrt{\lambda^+ U_A^2 \mathbb{E}[(M^+)^2] + \lambda^- U_B^2 \mathbb{E}[(M^-)^2]}. \quad (4.5)$$

Given time-to-maturity $T - t$ and strike K , the option reference price is defined as

$$b(t, e) = \text{BS}(P_t, K, T - t, r, \sigma_{\text{eff}}(t, e)),$$

where BS denotes the standard Black-Scholes formula, evaluated at the effective volatility σ_{eff} obtained from the mid-price dynamics. In our numerical implementation we set the risk-free rate $r = 0$, which is consistent with the symmetric order book assumption.

To enforce hedging discipline, we introduce a quadratic penalty anchored to the Black-Scholes delta,

$$g(t, q, i, e) = \kappa_{\text{hedge}} \left(q + i \partial_p b(t, e) \right)_+^2,$$

where $\partial_p b$ is the option delta under σ_{eff} . To complement the hedging penalty, we introduce an incentive term that only penalizes under-provision of liquidity. Let $\bar{\lambda}^a, \bar{\lambda}^b$ denote the maximal ask and bid intensities, and set the benchmark

$$\Lambda := \theta_{\text{flow}} (\bar{\lambda}^a + \bar{\lambda}^b),$$

with $\theta \in (0, 1)$ a fixed proportion of the maximal total flow. Given quotes (α, β) generating order flow intensities (λ^a, λ^b) , the activity penalty is then defined as

$$h(t, \alpha, \beta, e) = \kappa_{\text{act}} \left(\Lambda - (\lambda^a(t, \alpha, \beta, e) + \lambda^b(t, \alpha, \beta, e)) \right)_+^2,$$

with $\kappa_{\text{act}} > 0$ a penalty coefficient and $(x)_+ = \max(x, 0)$. This formulation ensures that the penalty vanishes whenever the total captured flow exceeds the benchmark Λ , while increasing quadratically otherwise. Unless specified otherwise, all subsequent experiments use these parameter values; they are

consistent with Assumption 2.6 and preserve the qualitative behavior described in Section 2.3. In our numerical applications, we adopt the following parameter values:

$$\kappa_{\text{hedge}} = 4, \quad \theta_{\text{flow}} = 5\%, \quad \kappa_{\text{act}} = 0.1.$$

These choices provide a sufficiently strong incentive for hedging discipline while ensuring a minimal level of liquidity provision without excessively penalizing the agent.

4.2.2 Baseline: zero option inventory

In this first experiment we adopt the baseline calibration, with the initial option inventory fixed at $I_0 = 0$. This neutral starting point serves as a benchmark, in later experiments we will vary the initial inventory to analyze how it shapes the market maker’s optimal behavior. Training is conducted on batches of 10,000 simulated paths over 500 epochs, with a learning rate of 10^{-4} . The policy network is implemented as a standard multilayer perceptron with ReLU activations and input normalization.

Figure 1 documents the training dynamics across episodes: the return rises steadily and then plateaus, indicating that the policy discovers a stable quoting and hedging regime. The hedging penalty declines markedly as the agent learns to align its underlying position with the option exposure, thereby reducing costly discrepancies. The incentive penalty quickly drops toward low levels and remains stable, which shows that the agent consistently meets the activity requirement without being punished for supplying additional liquidity.

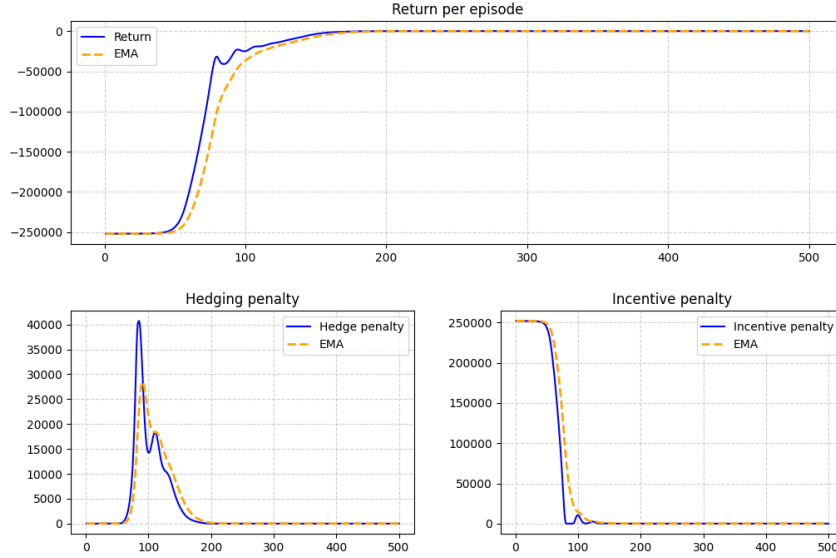


Figure 1: Learning metrics for linear order book with $I_0 = 0$.

To better understand the policy learned in this setting, we now examine the average evolution of both the strategy and the environment across 10,000 simulated paths. Figure 2 shows that, on average, the agent learns to quote around the option reference price, with spreads that gradually decrease as maturity approaches. This tightening reflects the fact that option values converge toward their intrinsic component near expiration, making order flow more sensitive to quoted prices. To continue attracting sufficient trading activity and avoid the incentive penalty, the agent reduces its spread, while still capturing a margin on transactions.

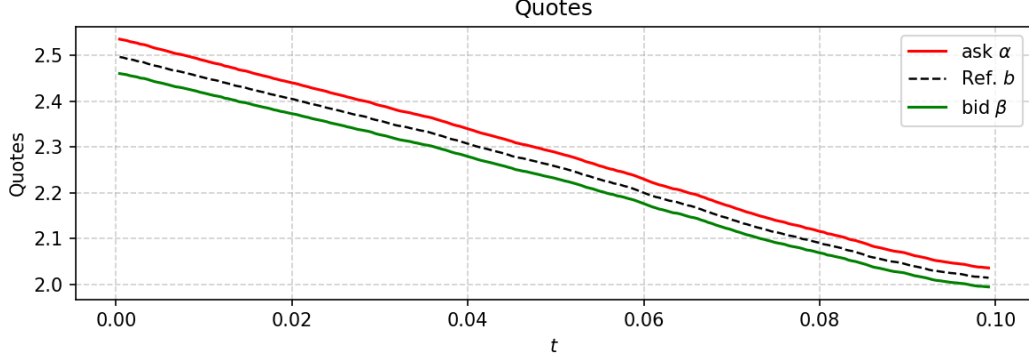


Figure 2: Average quoting strategy ($I_0 = 0$).

As illustrated in Figure 3a, the agent's strategy is to maintain its option inventory close to zero. This behavior is desirable for two reasons: first, it allows the agent to consistently capture spread revenues on the option market while generating sufficient order flow to avoid the activity penalty. Second, by keeping the inventory nearly balanced, the agent has no need to trade in the underlying, which simultaneously avoids transaction costs and eliminates hedging penalties.

Figure 3b illustrates the evolution of the market maker's cash along the path. We observe that the agent purchases more underlying than it sells during the first half of the horizon, before progressively unwinding this position and selling back underlying contracts as maturity approaches. This rebalancing behavior drives the P&L into positive levels toward the end of the trajectory.

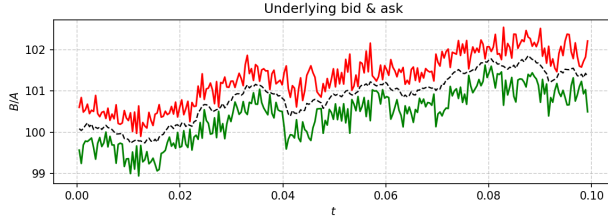


(a) Average option inventory and hedging position over time ($I_0 = 0$). (b) Average P&L trajectory of the market maker ($I_0 = 0$).

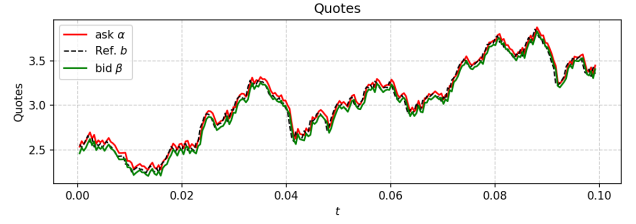
Figure 3: Average inventories and P&L along trajectory.

In the early part of the horizon, the agent tends to accumulate a slightly negative option inventory, which is naturally offset by purchases of the underlying to maintain a balanced delta exposure. As maturity approaches, however, the incentive to reduce residual risk dominates: the agent progressively unwinds its hedge and sells back underlying units in order to converge toward a flat terminal position. This two-phase pattern explains why the cash balance initially decreases due to net purchases, before recovering and turning positive once the accumulated option spread revenues are realized and the underlying position is closed out.

Figure 4 provides an illustration based on a single simulated path. Panel (a) displays the evolution of the underlying mid-price, while Panel (b) reports the bid-ask quotes set by the market maker relative to the option reference price. This example highlights the typical joint dynamics of the market environment and the agent's quoting behavior in the symmetric baseline case.



(a) Underlying mid-price path ($I_0 = 0$).



(b) Quoted bid-ask spread relative to reference price ($I_0 = 0$).

Figure 4: Sample path: underlying price and corresponding quoting behavior of the agent ($I_0 = 0$).

In this symmetric setting, the market maker succeeds in maintaining its option inventory close to zero, which in turn results in very limited trading activity on the underlying market. This outcome is consistent with the symmetry of the setup, since order flow intensities are centered and do not generate persistent directional pressure. In the next set of experiments, we deliberately introduce asymmetries in the option market to analyze how the market maker adapts its quoting and hedging behavior in response.

4.2.3 Asymmetry: negative initial inventory

We now investigate the impact of an asymmetric initial condition by setting the option inventory to $I_0 = -100$. All training parameters are kept identical to the symmetric reference case (batches of 10,000 simulated paths, 500 epochs, learning rate 10^{-4}). The only difference is the negative initial option position, which forces the market maker to adopt a biased strategy, as the outstanding inventory would generate a loss if carried to maturity.

Figure 5 confirms that training converges in this setting as well, with declining hedging and activity penalties indicating that the agent succeeds in managing both risks despite the asymmetric starting point.

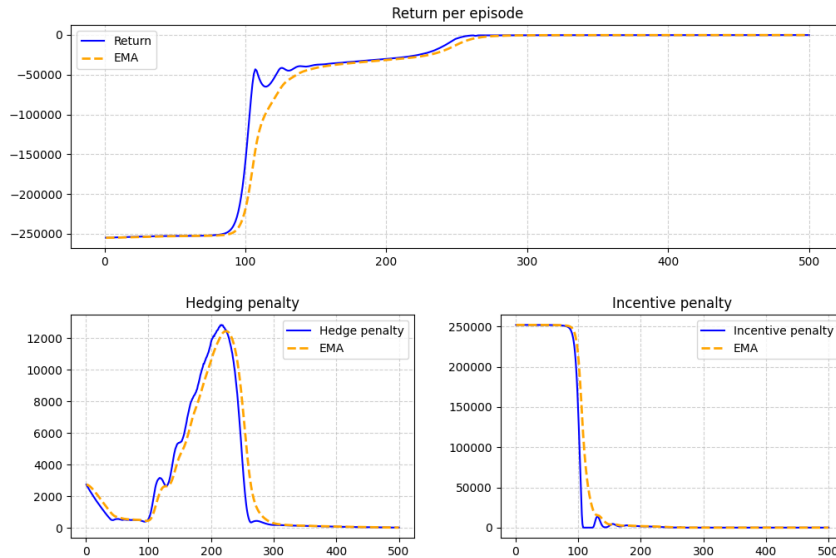


Figure 5: Learning metrics for linear order book with $I_0 = -100$.

A key feature of the learned policy is the immediate reaction of the market maker on the underlying market. As shown in Figure 6, the agent rapidly buys underlying contracts to offset the negative option

position and reduce the hedging penalty. This generates an upward pressure on the option reference price via hedging-induced impact, which subsequently decays as resilience restores market balance. The initial spike in the mid-price is therefore a direct consequence of hedging activity triggered by the initial short option inventory.

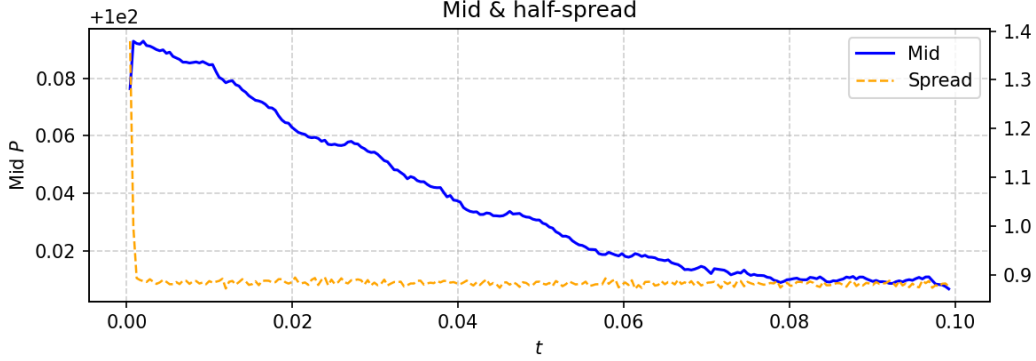


Figure 6: Average mid-price and spread ($I_0 = -100$).

Turning to the quoting strategy, Figure 7 shows that the agent sets asymmetric spreads, with wider quotes on the ask side than on the bid side. This asymmetry indicates that the market maker is willing to pay relatively more to attract sell orders (thereby reducing the short option position), while discouraging trades that would increase it. In practice, the agent biases order flow in the direction that helps liquidate its initial imbalance.

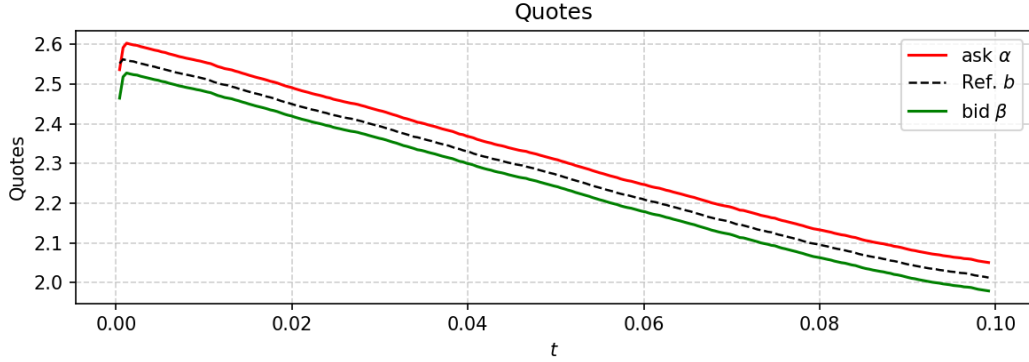


Figure 7: Average quoting strategy ($I_0 = -100$).

The inventory dynamics reported in Figure 8a confirm this interpretation. The option inventory progressively converges toward zero as maturity approaches, while the underlying hedge position is gradually unwound. This two-step process ensures that both the hedging penalty and terminal risk are controlled.

Finally, Figure 8b reports the cash trajectory. Unlike the symmetric case, the path features a pronounced initial drawdown: the agent aggressively buys underlying at the ask to hedge the short option position, incurring immediate execution and impact costs. Subsequently, as the policy skews quotes to buy back options and gradually unwinds the hedge, the cash recovers thanks to option spread capture and liquidation proceeds. In this configuration, however, the recovery is incomplete: the P&L converges toward break-even from below and does not turn positive by maturity. The early hedging cost dominates the spread revenues that can be earned over the remaining horizon.



(a) Average option inventory and hedging position over time ($I_0 = -100$). (b) Average P&L trajectory of the market maker ($I_0 = -100$).

Figure 8: Average inventories and P&L along trajectory.

4.2.4 Asymmetry: shifted order intensities

In this second asymmetric configuration, we do not impose a negative initial option inventory. Instead, we distort the option market itself by shifting the order arrival intensities away from symmetry. Specifically, we use the specification of Equation (4.4) with $\mu_b = 3/2$ and $\mu_a = -1/2$, and reduce the maximal ask intensity to 80% of the reference value. As shown in Figure 9, this parametrization makes bid intensities systematically higher than ask intensities near the reference price, with the two curves intersecting only at quotes placed well below b . As a consequence, client buy orders (hitting the bid) occur more frequently than sell orders, so the market maker is structurally pushed toward a negative option inventory. This imbalance in option flows then translates into a persistent exposure in the underlying market.

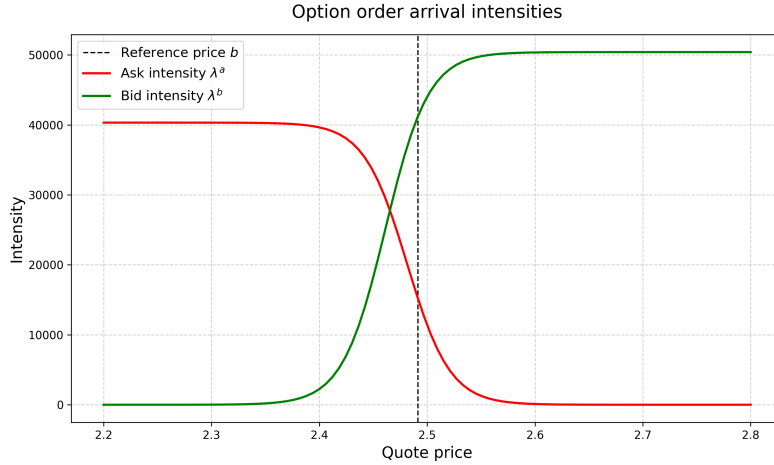


Figure 9: Shape of asymmetric option order flow intensities.

The average quoting strategy is shown in Figure 10. The agent systematically quotes with a wider spread on the bid side than on the ask side. This asymmetry reflects an attempt to discourage client buy orders, which are structurally more frequent due to the shifted intensities, while still keeping competitive ask quotes to facilitate inventory reduction. In other words, the market maker adapts its quotes to slow down the accumulation of negative inventory while creating opportunities to sell back options.

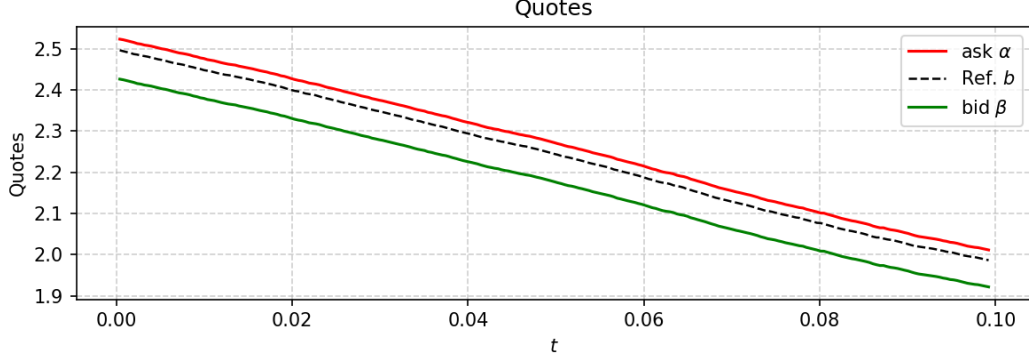


Figure 10: Average quoting strategy ($I_0 = 0$ and asymmetric intensities).

Figure 11a confirms this mechanism at the level of inventories. Starting from $I_0 = 0$, the market maker's option inventory drifts downward as client buy orders dominate, reaching significantly negative levels around the midpoint of the horizon. In the second half, the agent progressively buys back options in order to converge to a flat position at maturity. The hedging activity on the underlying mirrors this evolution: the agent initially purchases underlying contracts to offset the negative option delta, and then unwinds this hedge as the option inventory is reduced.

Finally, Figure 11b shows the average P&L trajectory. The market maker suffers an initial cash outflow when building its hedge against the growing negative option position. Subsequently, spread revenues on the option market and the unwinding of the hedge contribute to a recovery, with the cash balance turning positive before maturity and remaining in surplus at the terminal date. This pattern illustrates the two-phase adaptation of the market maker in an unbalanced environment: absorbing an early cost to manage exposures, then progressively recovering through adjusted quoting and hedging.



(a) Average option inventory and hedging position over time ($I_0 = 0$ and asymmetric intensities).

(b) Average P&L trajectory of the market maker ($I_0 = 0$ and asymmetric intensities).

Figure 11: Average inventories and P&L along trajectory.

4.2.5 Liquidity-constrained environment

In this final experiment, we initialize the market maker with a negative option inventory $I_0 = -100$ and reduce the available liquidity. Specifically, we set $c_A = c_B = 2.5$ and $U_A = U_B = 2$.

Increasing the depth parameters U_A and U_B amplifies the volatility of the underlying asset, since individual order arrivals trigger larger price movements. In the baseline setting the effective volatility was around 10%, whereas under the present low-liquidity configuration it rises to approximately 40%.

The combination of a negative initial option inventory and reduced liquidity induces a markedly different regime compared to the baseline experiments. At the start of the horizon, the inventory constraint dominates the agent's behavior: as shown in Figure 13a, the market maker immediately engages in aggressive purchases of the underlying in order to hedge its short option exposure. Because the available depth is low, this sequence of trades strongly impacts the underlying mid-price.

The quoting behavior, reported in Figure 12, exhibits very tight spreads around the reference price. This reflects the agent’s attempt to maximize order flow and thereby adjust its option inventory as rapidly as possible. A slight asymmetry is visible, with quotes skewed so as to favor option purchases, consistent with the objective of unwinding the initial short position.

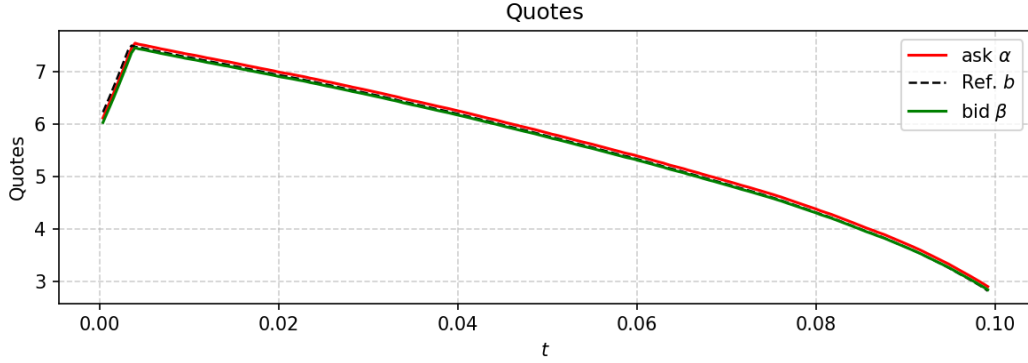
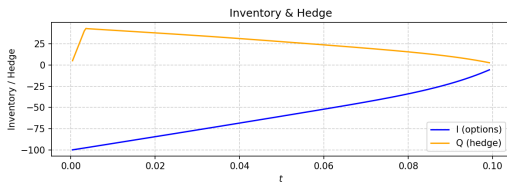


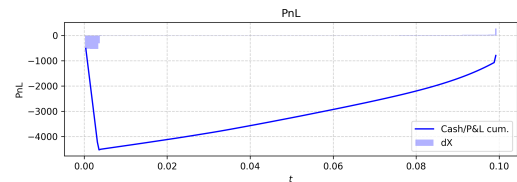
Figure 12: Average quoting strategy ($I_0 = -100$ and low liquidity).

As the horizon progresses, the agent gradually reduces its short option position, converging toward a nearly flat inventory at maturity (Figure 13a). This adjustment is accompanied by a progressive liquidation of the hedge position in the underlying, once the bulk of the initial coverage has been achieved.

The cash dynamics are displayed in Figure 13b. During the first part of the horizon, the P&L decreases sharply due to costly hedging trades executed in a thin market. Although the agent subsequently stabilizes its performance by collecting option spreads and progressively unwinding its positions, the initial imbalance leaves a lasting mark. The final P&L remains negative, reflecting the fact that the temporary mismatch between option and underlying exposures generates losses that cannot be fully compensated within the limited horizon.



(a) Average option inventory and hedging position over time ($I_0 = -100$ and low liquidity).



(b) Average P&L trajectory of the market maker ($I_0 = -100$ and low liquidity).

Figure 13: Average inventories and P&L along trajectory.

Bibliography

- [1] Aurélien Alfonsi and José Infante Acevedo. “Optimal execution and price manipulations in time-varying limit order books”. In: *Applied Mathematical Finance* 21.3 (2014), pp. 201–237. DOI: 10.1080/1350486X.2013.845471.
- [2] Aurélien Alfonsi and Pierre Blanc. “Dynamic optimal execution in a mixed-market-impact Hawkes price model”. In: *Finance and Stochastics* 20.1 (2016), pp. 183–218. ISSN: 1432-1122. DOI: 10.1007/s00780-015-0282-y.
- [3] Robert Almgren and Neil Chriss. “Optimal execution of portfolio transactions”. In: *Journal of Risk* 3 (2001), pp. 5–40. DOI: 10.21314/JOR.2001.041.

- [4] Marco Avellaneda and Sasha Stoikov. “High-frequency trading in a limit order book”. In: *Quantitative Finance* 8.3 (2008), pp. 217–224. DOI: 10.1080/14697680701381228.
- [5] Bastien Baldacci, Philippe Bergault, and Olivier Guéant. “Algorithmic market making for options”. In: *Quantitative Finance* 21.1 (2021), pp. 85–97. DOI: 10.1080/14697688.2020.1766099.
- [6] Hans Buehler et al. “Deep hedging”. In: *Quantitative Finance* 19.8 (2019), pp. 1271–1291. DOI: 10.1080/14697688.2019.1571683.
- [7] Álvaro Cartea and Sebastian Jaimungal. “Optimal execution with limit and market orders”. In: *Quantitative Finance* 15 (May 2015), pp. 1–13. DOI: 10.1080/14697688.2015.1032543.
- [8] Álvaro Cartea, Sebastian Jaimungal, and Jason Ricci. “Buy Low, Sell High: A High Frequency Trading Perspective”. In: *SIAM Journal on Financial Mathematics* 5.1 (2014), pp. 415–444. DOI: 10.1137/130911196.
- [9] Sofiene El Aoud and Frédéric Abergel. “A Stochastic Control Approach to Option Market Making”. In: *SSRN Electronic Journal* 01 (July 2015). DOI: 10.2139/ssrn.2491446.
- [10] Jim Gatheral and Alexander Schied. “Dynamical Models of Market Impact and Algorithms for Order Execution”. In: *SSRN Electronic Journal* (Jan. 2013), pp. 579–599. DOI: 10.2139/ssrn.2034178.
- [11] Olivier Guéant, Charles-Albert Lehalle, and Joaquin Fernandez Tapia. “Dealing with the Inventory Risk. A solution to the market making problem”. In: *Mathematics and Financial Economics* 7.4 (2013), pp. 477–507. DOI: 10.1007/s11579-012-0087-0.
- [12] Fabien Guilbaud and Huyên Pham. “Optimal High Frequency Trading with Limit and Market Orders”. In: *Quantitative Finance* 13 (June 2011). DOI: 10.1080/14697688.2012.708779.
- [13] Gur Huberman and Werner Stanzl. “Price manipulation and quasi-arbitrage”. In: *Econometrica* 72.4 (2004), pp. 1247–1275. DOI: 10.1111/j.1468-0262.2004.00531.x.
- [14] Silviu Predoiu, Gennady Shaikhet, and Steven Shreve. “Optimal Execution in a General One-Sided Limit-Order Book”. In: *SIAM J. Financial Math.* 2 (Jan. 2011), pp. 183–212. DOI: 10.1137/10078534X.
- [15] Thomas Spooner et al. “Market Making via Reinforcement Learning”. In: *Proceedings of the 17th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2018)*. 2018. URL: <https://dl.acm.org/doi/10.5555/3237383.3237450>.
- [16] Sasha Stoikov and Mehmet Sağlam. “Option market making under inventory risk”. In: *Review of Derivatives Research* 12.1 (Apr. 2009), pp. 55–79. DOI: 10.1007/s11147-009-9036-3.

A Technical proofs

A.1 Stability and moment estimates

We establish stability and moment bounds for the Hawkes intensities and the state variables, which are then propagated to the price, spread, and resilience dynamics. These estimates provide the probabilistic control required to prove boundedness of the value function.

Lemma A.4 (Moment bounds for the option inventory). *Consider the inventory process defined in Equation (2.14) satisfying Assumption 2.5. Then, there exist constants $C_{1,T}^{(I)}, C_{2,T}^{(I)} > 0$, depending only on T and on $(\bar{\lambda}^b, \bar{\lambda}^a)$, such that*

$$\sup_{t \in [0, T]} \mathbb{E}[|I_t|] \leq C_{1,T}^{(I)}(1 + |i|), \quad \sup_{t \in [0, T]} \mathbb{E}[|I_t|^2] \leq C_{2,T}^{(I)}(1 + i^2). \quad (\text{A.1})$$

Proof. By (2.14), the inventory satisfies $I_t = i + N_t^b - N_t^a$ for $t \in [0, T]$. Let Λ_t^b and Λ_t^a be the compensators of the counting processes and define the compensated martingales $M_t^b := N_t^b - \Lambda_t^b$ and $M_t^a := N_t^a - \Lambda_t^a$. Assumption 2.5 gives $0 \leq \lambda^b \leq \bar{\lambda}^b$ and $0 \leq \lambda^a \leq \bar{\lambda}^a$, hence $0 \leq \Lambda_t^b \leq \bar{\lambda}^b t$ and $0 \leq \Lambda_t^a \leq \bar{\lambda}^a t$. For the first moment,

$$\begin{aligned}\mathbb{E}[|I_t|] &\leq |i| + \mathbb{E}[N_t^a] + \mathbb{E}[N_t^b] \\ &= |i| + \mathbb{E}[\Lambda_t^a] + \mathbb{E}[\Lambda_t^b] \\ &\leq |i| + t(\bar{\lambda}^a + \bar{\lambda}^b),\end{aligned}$$

and taking $\sup_{t \in [0, T]}$ yields

$$\sup_{t \in [0, T]} \mathbb{E}[|I_t|] \leq |i| + T(\bar{\lambda}^a + \bar{\lambda}^b) \leq C_{1, T}^{(I)}(1 + |i|).$$

For the second moment, $(a + b + c)^2 \leq 3(a^2 + b^2 + c^2)$ gives $|I_t|^2 \leq 3i^2 + 3(N_t^a)^2 + 3(N_t^b)^2$. Writing $N_t^k = M_t^k + \Lambda_t^k$ ($k \in \{a, b\}$) and using $(a + b)^2 \leq 2a^2 + 2b^2$,

$$\mathbb{E}[(N_t^k)^2] \leq 2\mathbb{E}[(M_t^k)^2] + 2\mathbb{E}[(\Lambda_t^k)^2].$$

Since $\langle M^k \rangle_t = \Lambda_t^k$, the martingale isometry yields $\mathbb{E}[(M_t^k)^2] = \mathbb{E}[\Lambda_t^k] \leq \bar{\lambda}^k t$, while $\mathbb{E}[(\Lambda_t^k)^2] \leq (\bar{\lambda}^k t)^2$. Therefore

$$\mathbb{E}[(N_t^k)^2] \leq 2\bar{\lambda}^k t + 2(\bar{\lambda}^k t)^2,$$

and consequently

$$\sup_{t \in [0, T]} \mathbb{E}[|I_t|^2] \leq 3i^2 + 6(\bar{\lambda}^a + \bar{\lambda}^b)T + 6((\bar{\lambda}^a)^2 + (\bar{\lambda}^b)^2)T^2 \leq C_{2, T}^{(I)}(1 + i^2),$$

This proves (A.1) and the square-integrability of $(I_t)_{t \in [0, T]}$. \square

Lemma A.5 (Moment bounds for Hawkes processes with interventions). *Consider the intensity process $(\lambda_t^\pm)_{t \geq 0}$ defined in (2.9)–(2.10), with parameters satisfying Assumption 2.3, and let $(N_t^\pm)_{t \geq 0}$ be the associated counting process. Assume the intervention processes $(H_t^\pm)_{t \geq 0}$ are adapted, nondecreasing, càdlàg, $H_0^\pm = 0$, and admit bounded first moment on $[0, T]$, i.e.*

$$\overline{H}_T^\pm = \sup_{t \in [0, T]} \mathbb{E}[H_t^\pm] < \infty.$$

Then there exists two constants $C_{1, T}^{(H^\pm)}, \tilde{C}_{1, T}^{(H^\pm)} > 0$, depending only on T , the model parameters (θ, μ, κ) , and the intervention moments \overline{H}_T^\pm , such that

$$\sup_{t \in [0, T]} \mathbb{E}[\lambda_t^\pm] \leq C_{1, T}^{(H^\pm)}(1 + \lambda_0), \quad \sup_{t \in [0, T]} \mathbb{E}[N_t^\pm] \leq \tilde{C}_{1, T}^{(H^\pm)}(1 + \lambda_0^\pm). \quad (\text{A.2})$$

If in addition, we suppose that

$$\overline{H}_{2, T}^\pm = \sup_{t \in [0, T]} \mathbb{E}[H_t^2] < \infty.$$

Then there exist constants $C_{2, T}^{(H^\pm)}, \tilde{C}_{2, T}^{(H^\pm)} > 0$, depending only on T , the model parameters (θ, μ, κ) , and the intervention moments $\overline{H}_T^\pm, \overline{H}_{2, T}^\pm$, such that

$$\sup_{t \in [0, T]} \mathbb{E}[(\lambda_t^\pm)^2] \leq C_{2, T}^{(H^\pm)}(1 + \lambda_0^2), \quad \sup_{t \in [0, T]} \mathbb{E}[(N_t^\pm)^2] \leq \tilde{C}_{2, T}^{(H^\pm)}(1 + (\lambda_0^\pm)^2).$$

Proof. For simplicity, we drop the \pm superscripts, but the proof is valid for both cases. Let $M_t := N_t - \int_0^t \lambda_u du$ be the compensated martingale associated with N . From (2.9)–(2.10), the intensity satisfies the linear SDE

$$d\lambda_t = \theta \mu dt - (\theta - \kappa) \lambda_t dt + \kappa dM_t + \kappa dH_t,$$

with $\lambda_0 \geq 0$. By variation of constants, taking expectations and noting that $\mathbb{E}\left[\int_0^t dM_u\right] = 0$,

$$\mathbb{E}[\lambda_t] = e^{-(\theta-\kappa)t} \lambda_0 + \frac{\theta\mu}{\theta-\kappa} (1 - e^{-(\theta-\kappa)t}) + \kappa \mathbb{E}\left[\int_0^t e^{-(\theta-\kappa)(t-u)} dH_u\right].$$

Thanks to Assumption 2.3 we have $\theta - \kappa > 0$ and this allows us to prove the first claim:

$$\mathbb{E}[\lambda_t] \leq \lambda_0 + \frac{\theta\mu}{\theta-\kappa} + \kappa \overline{H}_T \leq C_{1,T}^{(H)} (1 + \lambda_0).$$

Furthermore, for $t \in [0, T]$,

$$\mathbb{E}[N_t] = \int_0^t \mathbb{E}[\lambda_u] du \leq \frac{\lambda_0}{\theta-\kappa} + \frac{\theta\mu}{\theta-\kappa} t + \kappa \int_0^t \mathbb{E}\left[\int_0^u e^{-(\theta-\kappa)(u-s)} dH_s\right] du.$$

Applying Tonelli–Fubini for nonnegative integrands pathwise (since H is nondecreasing),

$$\int_0^t \left(\int_0^u e^{-(\theta-\kappa)(u-s)} dH_s\right) du = \int_0^t \left(\int_s^t e^{-(\theta-\kappa)(u-s)} du\right) dH_s \leq \frac{H_t}{\theta-\kappa}.$$

Taking expectations, recalling that $\sup_{u \in [0, T]} \mathbb{E}[H_u] = \overline{H}_T$ and maximizing over $t \in [0, T]$,

$$\sup_{t \in [0, T]} \mathbb{E}[N_t] \leq \frac{\theta\mu}{\theta-\kappa} T + \frac{\lambda_0}{\theta-\kappa} + \frac{\kappa}{\theta-\kappa} \overline{H}_T \leq \tilde{C}_{1,T}^{(H)} (1 + \lambda_0),$$

for some constant $\tilde{C}_{1,T}^{(H)}$ depending only on T , (θ, μ, κ) , and the intervention moments \overline{H}_T . This proves (A.2). For the square-integrability we start by calculating

$$\begin{aligned} \mathbb{E}[\lambda_t^2] &= \mathbb{E}\left[\lambda_0^2 e^{-2(\theta-\kappa)t} + \left(\theta\mu \int_0^t e^{-(\theta-\kappa)(t-u)} du\right)^2 + 2\lambda_0\theta\mu \int_0^t e^{-(\theta-\kappa)(2t-u)} du \right. \\ &\quad + \left(\kappa \int_0^t e^{-(\theta-\kappa)(t-u)} dM_u\right)^2 + \left(\kappa \int_0^t e^{-(\theta-\kappa)(t-u)} dH_u\right)^2 \\ &\quad + 2\lambda_0\kappa \int_0^t e^{-(\theta-\kappa)(2t-u)} dH_u + 2\left(\theta\mu \int_0^t e^{-(\theta-\kappa)(t-u)} du\right) \left(\kappa \int_0^t e^{-(\theta-\kappa)(t-u)} dH_u\right) \\ &\quad \left. + 2\left(\kappa \int_0^t e^{-(\theta-\kappa)(t-u)} dM_u\right) \left(\kappa \int_0^t e^{-(\theta-\kappa)(t-u)} dH_u\right)\right]. \end{aligned}$$

Applying Cauchy–Schwarz and using $\theta - \kappa > 0$ (Assumption 2.3), which implies $e^{-(\theta-\kappa)(\cdot)} \leq 1$, we obtain

$$\begin{aligned} \mathbb{E}[\lambda_t^2] &\leq \lambda_0^2 + (\theta\mu)^2 t^2 + 2\lambda_0\theta\mu t + \kappa^2 \mathbb{E}[M_t^2] + \kappa^2 \mathbb{E}[H_t^2] + 2\lambda_0\kappa \mathbb{E}[H_t] \\ &\quad + 2\theta\mu\kappa t \mathbb{E}[H_t] + 2\kappa^2 \sqrt{\mathbb{E}[M_t^2]} \sqrt{\mathbb{E}[H_t^2]}. \end{aligned}$$

We now absorb the cross terms by Young’s inequality and Jensen:

$$2\lambda_0\theta\mu t \leq \lambda_0^2 + (\theta\mu t)^2, \quad 2\lambda_0\kappa \mathbb{E}[H_t] \leq \lambda_0^2 + \kappa^2 (\mathbb{E}[H_t])^2 \leq \lambda_0^2 + \kappa^2 \mathbb{E}[H_t^2],$$

and

$$2\theta\mu\kappa t \mathbb{E}[H_t] \leq (\theta\mu t)^2 + \kappa^2 \mathbb{E}[H_t^2], \quad 2\kappa^2 \sqrt{\mathbb{E}[M_t^2]} \sqrt{\mathbb{E}[H_t^2]} \leq \kappa^2 \mathbb{E}[M_t^2] + \kappa^2 \mathbb{E}[H_t^2].$$

Plugging these bounds above yields

$$\mathbb{E}[\lambda_t^2] \leq 3\lambda_0^2 + 3(\theta\mu)^2 t^2 + 2\kappa^2 \mathbb{E}[M_t^2] + 4\kappa^2 \mathbb{E}[H_t^2].$$

Since $\mathbb{E}[M_t^2] = \mathbb{E}[\int_0^t \lambda_u du] = \mathbb{E}[N_t]$, the first-moment bound from (A.2) gives $\sup_{t \leq T} \mathbb{E}[M_t^2] \leq \tilde{C}_{1,T}^{(H)}(1 + \lambda_0)$. Together with $\sup_{t \leq T} \mathbb{E}[H_t^2] \leq \overline{H}_{2,T} < \infty$ and $t \leq T$ we conclude that

$$\sup_{t \in [0, T]} \mathbb{E}[\lambda_t^2] \leq C_{2,T}^{(H)}(1 + \lambda_0^2),$$

for some constant $C_{2,T}^{(H)}$ depending only on T , (θ, μ, κ) , and the intervention moments $\overline{H}_T, \overline{H}_{2,T}$. Finally, for the counting process we write $N_t = M_t + \Lambda_t$ with $\Lambda_t := \int_0^t \lambda_u du$. Then $N_t^2 \leq 2M_t^2 + 2\Lambda_t^2$, hence

$$\mathbb{E}[N_t^2] \leq 2\mathbb{E}[M_t^2] + 2\mathbb{E}[\Lambda_t^2].$$

By the martingale isometry, $\mathbb{E}[M_t^2] = \mathbb{E}[\Lambda_t] \leq \tilde{C}_{1,T}^{(H)}(1 + \lambda_0)$, and by Cauchy-Schwarz,

$$\mathbb{E}[\Lambda_t^2] = \mathbb{E}\left[\left(\int_0^t \lambda_u du\right)^2\right] \leq t \int_0^t \mathbb{E}[\lambda_u^2] du \leq t^2 \sup_{0 \leq u \leq t} \mathbb{E}[\lambda_u^2] \leq T^2 C_{2,T}^{(H)}(1 + \lambda_0^2).$$

Therefore,

$$\sup_{t \in [0, T]} \mathbb{E}[N_t^2] \leq \tilde{C}_{2,T}^{(H)}(1 + \lambda_0^2),$$

for some finite constant $\tilde{C}_{2,T}^{(H)}$ depending only on T , (θ, μ, κ) , and the intervention moments. \square

Lemma A.5 provides uniform first and second-moment control of the Hawkes intensities and their counting processes, which are the sole source of randomness in the model. Combined with the integrated representations of P , D , and S , these bounds yield uniform moment estimates for the full state on $[0, T]$, ensuring integrability of all terms and supporting the finiteness of the value function.

Proposition A.5 (Moment bounds for the state process). *Let $(E_u)_{u \geq 0} := (P_u, D_u, S_u, \lambda_u^-, \lambda_u^+)$ denote the market state process with initial condition $E_0 = e := (p, d, s, \lambda_0^-, \lambda_0^+)$. We assume Assumptions 2.2 and 2.3 hold. Assume moreover that the intervention processes $(H_t^\pm)_{t \geq 0}$ are adapted, nondecreasing, càdlàg, satisfy $H_0^\pm = 0$, and admit bounded first moments on $[0, T]$, i.e.*

$$\overline{H}_T^\pm = \sup_{t \in [0, T]} \mathbb{E}[H_t^\pm] < \infty.$$

Then there exists a finite constant $C_{1,T}^{(E)}$, depending only on T , the model parameters and on \overline{H}_T^\pm , such that,

$$\sup_{u \in [0, T]} \mathbb{E}\left[|P_u| + |D_u| + S_u + \lambda_u^- + \lambda_u^+\right] \leq C_{1,T}^{(E)}(1 + \|e\|).$$

If, in addition,

$$\overline{H}_{2,T}^\pm = \sup_{t \in [0, T]} \mathbb{E}[(H_t^\pm)^2] < \infty,$$

then there exists a finite constant $C_{2,T}^{(E)}$, depending only on T , the model parameters and on $\overline{H}_T^\pm, \overline{H}_{2,T}^\pm$, such that,

$$\sup_{u \in [0, T]} \mathbb{E}\left[|P_u|^2 + |D_u|^2 + S_u^2 + (\lambda_u^-)^2 + (\lambda_u^+)^2\right] \leq C_{2,T}^{(E)}(1 + \|e\|^2).$$

Proof. Set $U_\star := \max\{U_A, U_B\}$. By Assumption 2.2 and Proposition 2.1, the price impacts per single trade/impulse are uniformly bounded by U_\star .

Resilient component. Using the integrated representation of Equation (2.12), obtained via variation of constants, and noting that each jump has magnitude at most U_\star ,

$$|D_u| \leq e^{-ru}|d| + \frac{1-\eta}{2} \int_0^u e^{-r(u-s)} d(N_s^+ + N_s^- + H_s^+ + H_s^-) U_\star. \quad (\text{A.3})$$

Taking expectations and dropping the exponential yields

$$\mathbb{E}[|D_u|] \leq |d| + \frac{1-\eta}{2} U_\star (\mathbb{E}[N_u^+] + \mathbb{E}[N_u^-] + \mathbb{E}[H_u^+] + \mathbb{E}[H_u^-]).$$

Spread. Similarly, from (2.13),

$$S_u \leq \delta + e^{-\rho u}(s - \delta) + \int_0^u e^{-\rho(u-s)} d(N_s^+ + N_s^- + H_s^+ + H_s^-) U_\star, \quad (\text{A.4})$$

hence

$$\mathbb{E}[S_u] \leq \delta + s + U_\star (\mathbb{E}[N_u^+] + \mathbb{E}[N_u^-] + \mathbb{E}[H_u^+] + \mathbb{E}[H_u^-]).$$

Mid-price. From Equation 2.11, P is the sum of its permanent jump part and D . Using again the bound U_\star for each jump,

$$|P_u| \leq |p| + |D_u| + \frac{\eta}{2} U_\star (N_u^+ + N_u^- + H_u^+ + H_u^-), \quad (\text{A.5})$$

and thus

$$\mathbb{E}[|P_u|] \leq |p| + \mathbb{E}[|D_u|] + \frac{\eta}{2} U_\star (\mathbb{E}[N_u^+] + \mathbb{E}[N_u^-] + \mathbb{E}[H_u^+] + \mathbb{E}[H_u^-]).$$

Combining the three displays for $\mathbb{E}[|D_u|]$, $\mathbb{E}[S_u]$, and $\mathbb{E}[|P_u|]$, inserting the bounds on $\mathbb{E}[H_u^\pm]$ and using bounds of $\mathbb{E}[N_u^\pm]$ and $\mathbb{E}[\lambda_u^\pm]$ obtained in Lemma A.5, we obtain

$$\sup_{u \in [0, T]} \mathbb{E}[|P_u| + |D_u| + S_u + \lambda_u^- + \lambda_u^+] \leq C_{1,T}^{(E)} (1 + |p| + |d| + s + \lambda_0^- + \lambda_0^+) = C_{1,T}^{(E)} (1 + \|e\|),$$

for a constant $C_{1,T}^{(E)} < \infty$ depending only on T , the model parameters and on $\sup_{u \in [0, T]} \mathbb{E}[H_u^\pm]$. This proves the first claim.

Now, for the square integrability we use the integrated representations together with $e^{-r(\cdot)}, e^{-\rho(\cdot)} \leq 1$ and Young's inequality, we bound each component on $[0, T]$.

Resilient component. From (A.3) we obtain:

$$|D_u|^2 \leq 2d^2 + 2 \left(\frac{1-\eta}{2} \right)^2 U_\star^2 \left(\int_0^u d(N_s^+ + N_s^- + H_s^+ + H_s^-) \right)^2$$

Using $(x_1 + x_2 + x_3 + x_4)^2 \leq 4 \sum_{i=1}^4 x_i^2$ and taking expectations,

$$\mathbb{E}[|D_u|^2] \leq 2d^2 + 8 \left(\frac{1-\eta}{2} \right)^2 U_\star^2 (\mathbb{E}[(N_u^+)^2] + \mathbb{E}[(N_u^-)^2] + \mathbb{E}[(H_u^+)^2] + \mathbb{E}[(H_u^-)^2]).$$

Spread. From (A.4) and $s \geq \delta$,

$$S_u^2 \leq 2s^2 + 2U_\star^2 \left(\int_0^u d(N_s^+ + N_s^- + H_s^+ + H_s^-) \right)^2$$

and therefore

$$\mathbb{E}[S_u^2] \leq 2s^2 + 8U_\star^2 (\mathbb{E}[(N_u^+)^2] + \mathbb{E}[(N_u^-)^2] + \mathbb{E}[(H_u^+)^2] + \mathbb{E}[(H_u^-)^2]).$$

Mid-price. From (A.5),

$$|P_u|^2 \leq 4p^2 + 4D_u^2 + 2\left(\frac{\eta}{2}\right)^2 U_\star^2 \left(\int_0^u d(N_s^+ + N_s^- + H_s^+ + H_s^-)\right)^2$$

and taking expectations,

$$\mathbb{E}[P_u^2] \leq 4p^2 + 4\mathbb{E}[D_u^2] + 8\left(\frac{\eta}{2}\right)^2 U_\star^2 \left(\mathbb{E}[(N_u^+)^2] + \mathbb{E}[(N_u^-)^2] + \mathbb{E}[(H_u^+)^2] + \mathbb{E}[(H_u^-)^2]\right).$$

Finally, by Lemma A.5 we have the uniform bounds

$$\sup_{u \leq T} \mathbb{E}[(N_u^\pm)^2] \leq \tilde{C}_{2,T}^{(H^\pm)} (1 + (\lambda_0^\pm)^2) \quad \text{and} \quad \sup_{u \leq T} \mathbb{E}[(\lambda_u^\pm)^2] \leq C_{2,T}^{(H^\pm)} (1 + (\lambda_0^\pm)^2),$$

while by assumption $\sup_{u \leq T} \mathbb{E}[(H_u^\pm)^2] \leq \overline{H}_{2,T}^\pm < \infty$. Plugging these estimates into the bounds obtained above for D , S , and P , and enlarging constants if necessary, we conclude that there exists a finite constant $C_{2,T}^{(E)}$ such that

$$\sup_{u \in [0, T]} \mathbb{E}[|P_u|^2 + |D_u|^2 + S_u^2 + (\lambda_u^-)^2 + (\lambda_u^+)^2] \leq C_{2,T}^{(E)} (1 + p^2 + d^2 + s^2 + (\lambda_0^-)^2 + (\lambda_0^+)^2).$$

This completes the proof of the second-order bound. \square

A.2 Proofs of the value-function bounds

Building on the previous results, we now establish the main statements that ensure the well-posedness of the value function. We begin by proving a global lower bound.

Proposition A.6 (Lower bound for the value function). *For all $(t, q, i, e) \in [0, T] \times \mathbb{R} \times \mathbb{Z} \times (\mathbb{R}^+)^5$, there exists a constant $C_T^{(-)} > 0$, depending only on T and on the model parameters, such that*

$$v(t, q, i, e) \geq -C_T^{(-)} (1 + q^2 + i^2 + \|e\|^2).$$

Proof of Proposition A.6. Fix an admissible strategy with no impulses in the underlying and constant option quotes $\beta \equiv 0$ and $\alpha \geq 0$. Then $Q_u \equiv q$ for all $u \in [t, T]$, and the order-flow term $\int_t^T [\alpha \lambda^a - \beta \lambda^b] du = \alpha \int_t^T \lambda^a du$ is nonnegative. From (2.21) and (2.20),

$$v(t, q, i, e) \geq \mathbb{E} \left[- \int_t^T (g + h)(u, Q_u, I_u, E_u) du + I_T \varphi(P_T) - P_A \left(P_T + \frac{1}{2} S_T, |q| \right) \right].$$

By Assumption 2.6, g and h satisfy a quadratic growth bound; since $Q_u \equiv q$,

$$- \int_t^T (g + h) du \geq -C \int_t^T (1 + q^2 + I_u^2 + \|E_u\|^2) du.$$

Lemma A.4 and Proposition A.5 with $H_u^\pm \equiv 0$ for all $u \in [t, T]$, yield

$$\mathbb{E} \left[- \int_t^T (g + h) du \right] \geq - \left(C_T^{(g)} + C_T^{(h)} \right) (1 + q^2 + i^2 + \|e\|^2). \quad (\text{A.6})$$

The payoff has linear growth, so $I_T \varphi(P_T) \geq -C^{(\varphi)} (|I_T| + |I_T| |P_T|)$. Taking expectations and applying Cauchy—Schwarz together with the same moment bounds gives

$$\mathbb{E}[I_T \varphi(P_T)] \geq -C_T (1 + i^2 + \|e\|^2). \quad (\text{A.7})$$

For the execution cost, the integral representation (2.7) and the bound $\Phi_A^{-1} \leq U_A$ imply

$$P_A \left(P_T + \frac{1}{2} S_T, |q| \right) \leq \left(P_T + \frac{1}{2} S_T \right) |q| + U_A |q|.$$

Taking expectations, and since $|q|$ is deterministic under the chosen strategy yields

$$\mathbb{E} \left[P_A \left(P_T + \frac{1}{2} S_T, |q| \right) \right] \leq |q| \mathbb{E} [P_T + \frac{1}{2} S_T] + U_A |q|.$$

Proposition A.5 gives $\mathbb{E}[P_T + S_T/2] \leq C_T(1 + \|e\|)$, hence

$$\mathbb{E} \left[P_A \left(P_T + \frac{1}{2} S_T, |q| \right) \right] \leq C_T |q| (1 + \|e\|) + U_A |q| \leq C_T (1 + q^2 + \|e\|^2), \quad (\text{A.8})$$

where the last inequality uses $|q| \leq 1 + q^2$.

Combining (A.6), (A.7), and (A.8) yields

$$v(t, q, i, e) \geq -C_T^{(-)} (1 + q^2 + i^2 + \|e\|^2),$$

for a constant $C_T^{(-)}$ depending only on T and the model parameters. \square

We now establish an upper bound for the value function, which implies its finiteness.

Proposition A.7 (Quadratic-growth of the value function). *The value function defined in Equation 2.21 is well-defined for all*

$$(t, q, i, e) \in [0, T] \times \mathbb{R} \times \mathbb{Z} \times (\mathbb{R}^+)^5.$$

Moreover, there exists a generic constant $C_T^{(+)} > 0$, depending only on the time horizon T , model parameters, such that

$$v(t, q, i, e) \leq C_T^{(+)} (1 + |i|^2 + \|e\|^2) \quad (\text{A.9})$$

Proof of Theorem A.7. Let us recall that any admissible strategy ensures that:

$$\sup_{t \in [0, T]} \mathbb{E}[H_t^{\pm}] < \infty, \quad \sup_{t \in [0, T]} \mathbb{E}[H_t^2] < \infty,$$

hence we can apply all results of Section A.1.

By definition of the value function and positivity of g , h , P_A and c :

$$\begin{aligned} v(t, q, i, e) &\leq \sup_{\gamma \in \mathcal{A}} \mathbb{E} \left[\int_t^T [\alpha_u \lambda^a - \beta_u \lambda^b] (u, \alpha_u, \beta_u, E_u) du \right. \\ &\quad \left. + \sum_{i=1}^{+\infty} \mathbb{1}_{\{\nu_i \in [t, T]\}} \left(P_B(P_{\nu_i} - S_{\nu_i}/2, \xi_i^-) \right) + L(T, Q_T, I_T, P_T, S_T) \right]. \end{aligned}$$

We first derive a strategy-independent bound for the option order-flow term. Under Assumptions 2.4 and 2.5, for any admissible (α, β) and $t \leq T$,

$$\begin{aligned} \mathbb{E} \left[\int_t^T (\alpha_u \lambda^a - \beta_u \lambda^b)(u, \alpha_u, \beta_u, E_u) du \right] &= \mathbb{E} \left[\int_t^T (\lambda^a(u, \alpha_u, \beta_u, E_u)(\alpha_u - b(u, E_u)) \right. \\ &\quad \left. - \lambda^b(u, \alpha_u, \beta_u, E_u)(\beta_u - b(u, E_u)) + b(u, E_u)(\lambda^a - \lambda^b)(u, \alpha_u, \beta_u, E_u)) du \right] \\ &\leq \mathbb{E} \left[\int_t^T (C^{(\lambda)}(1 + \|E_u\|) + b(u, E_u)(\lambda^a - \lambda^b)(u, \alpha_u, \beta_u, E_u)) du \right] \\ &\leq \mathbb{E} \left[\int_t^T (C^{(\lambda)}(1 + \|E_u\|) + (\bar{\lambda}^a + \bar{\lambda}^b) C^{(b)}(1 + \|E_u\|)) du \right] \end{aligned}$$

By the state first-moment estimate of Proposition A.5, we conclude that there exists $C_T^{(\lambda,b)} > 0$ such that

$$\sup_{\alpha,\beta} \mathbb{E} \left[\int_t^T (\alpha_u \lambda^a - \beta_u \lambda^b)(u, \alpha_u, \beta_u, E_u) du \right] \leq C_T^{(\lambda,b)} (1 + \|e\|).$$

Now, for the liquidation function. By Lemma 2.2, $P_T - \frac{1}{2}S_T \geq 0$ a.s. Using finite depth (Assumption 2.2) and Lemma 2.1, we have

$$0 \leq P_B \left(P_T - \frac{1}{2}S_T, \min\{Q_T, \Phi_B(B_T)\} \right) \leq \Phi_B(U_B)(P_T - \frac{1}{2}S_T).$$

Moreover, $-P_A(\cdot) \leq 0$, so for an upper bound we may drop the ask term. With the linear growth of φ ,

$$|I_T \varphi(P_T)| \leq C^{(\varphi)}(|I_T| + |I_T||P_T|).$$

Therefore

$$\mathbb{E}[L(T, Q_T, I_T, P_T, S_T)] \leq \Phi_B(U_B)\mathbb{E}[P_T - \frac{1}{2}S_T] + C^{(\varphi)}\mathbb{E}[|I_T| + |I_T||P_T|].$$

Apply the elementary inequalities $|x| \leq \frac{1}{2}(1 + x^2)$ and Young's inequality $|xy| \leq \frac{1}{2}(x^2 + y^2)$ to obtain

$$\mathbb{E}[P_T - \frac{1}{2}S_T] \leq \frac{1}{2} + \frac{1}{2}\mathbb{E}[P_T^2] + \frac{1}{4}\mathbb{E}[S_T^2], \quad \mathbb{E}[|I_T| + |I_T||P_T|] \leq \frac{1}{2} + \mathbb{E}[I_T^2] + \frac{1}{2}\mathbb{E}[P_T^2].$$

Hence

$$\mathbb{E}[L(T, Q_T, I_T, P_T, S_T)] \leq K \left(1 + \mathbb{E}[I_T^2] + \mathbb{E}[P_T^2] + \mathbb{E}[S_T^2] \right).$$

By Lemma A.4 and by Proposition A.5, enlarging constants we conclude that

$$\mathbb{E}[L(T, Q_T, I_T, P_T, S_T)] \leq C_T^{(L)}(1 + i^2 + \|e\|^2).$$

Finally as $P_B(b, q) \leq bq$:

$$\begin{aligned} v(t, q, i, e) &\leq \sup_{\gamma \in \mathcal{A}} \mathbb{E} \left[\int_t^T [\alpha_u \lambda^a - \beta_u \lambda^b](u, \alpha_u, \beta_u, E_u) du \right. \\ &\quad \left. + \sum_{i=1}^{+\infty} \mathbf{1}_{\{\nu_i \in [t, T]\}} \left(P_B(P_{\nu_i} - S_{\nu_i}/2, \xi_i^-) \right) + L(T, Q_T, I_T, P_T, S_T) \right] \\ &\leq \sup_{\gamma \in \mathcal{A}} \mathbb{E} \left[C_T^{(\lambda,b)}(1 + \|e\|) + \Phi_B(U_B) \sum_{i=1}^{+\infty} \mathbf{1}_{\{\nu_i \in [t, T]\}} P_{\nu_i} + C_T^{(L)}(1 + i^2 + \|e\|^2) \right] \\ &\leq C_T^{(\lambda,b)}(1 + \|e\|) + \Phi_B(U_B)\mathbb{E}[H_t^-] (1 + \|e\|) + C_T^{(L)}(1 + i^2 + \|e\|^2) \end{aligned}$$

Hence:

$$v(t, q, i, e) \leq C_T^{(+)} (1 + i^2 + \|e\|^2)$$

Which shows that $v(t, q, i, e) < +\infty$. Recalling that, by Proposition A.6 we also have $v(t, q, i, e) > -\infty$ proves that the value function is well defined. \square