

# A Games-in-Games Paradigm for Strategic Hybrid Jump-Diffusions: Hamilton-Jacobi-Isaacs Hierarchy and Spectral Structure

Yunian Pan & Quanyan Zhu  
 ECE Department  
 New York University  
 Brooklyn, NY  
 {yp1170, qz494}@nyu.edu

## ABSTRACT

This paper develops a hierarchical games-in-games control architecture for hybrid stochastic systems governed by regime-switching jump-diffusions. We model the interplay between continuous state dynamics and discrete mode transitions as a bilevel differential game: an inner layer solves a robust stochastic control problem within each regime, while a strategic outer layer modulates the transition intensities of the underlying Markov chain. A Dynkin-based analysis yields a system of coupled Hamilton-Jacobi-Isaacs (HJI) equations. We prove that for the class of Linear-Quadratic games and Exponential-Affine games, this hierarchy admits tractable semi-closed form solutions via coupled matrix differential equations. The framework is demonstrated through a case study on adversarial market microstructure, showing how the outer layer's strategic switching pre-emptively adjusts inventory spreads against latent regime risks, which leads to a hyper-alert equilibrium.

## 1 Introduction

Hybrid systems, characterized by the interplay between continuous state dynamics and discrete event transitions, constitute a fundamental modeling paradigm for complex socio-economic engineering systems. Within this broad class, *regime-switching jump-diffusions* occupy a central role, capturing systems where continuous stochastic trajectories are modulated by a hidden or observable Markov chain. Applications range from fault-tolerant control in networked systems [21, 17, 18] and cyber-physical security [22, 19, 16] to economic systems [14].

In these settings, decision-making is rarely monolithic. It operates hierarchically: fast-timescale controllers regulate the continuous state (e.g., stabilizing a plant or hedging a portfolio), while slow-timescale policies influence the discrete operating modes (e.g., system reconfiguration or regime induction). However, classical literature typically decouples these layers. The theory of *Piecewise-Deterministic Markov Processes* (PDMPs) [11, 9] and switching diffusions [21] generally treats the regime transition mechanism as either exogenous (governed by nature) or subject to a single controller's optimization (optimal switching) [12].

A critical gap exists in modeling *adversarial hybrid interactions*, where the regime transitions themselves are the outcome of a strategic game. For instance, in cyber-physical systems, an attacker may seek to destabilize the system by inducing transitions to vulnerable modes [8], while a defender attempts to harden the transition logic. Existing differential game theory [6, 7] provides robust tools for the continuous layer but typically assumes a fixed or purely stochastic discrete structure. Conversely, impulse games [1] focus on discrete interventions but often abstract away the continuous-time feedback loops.

This paper bridges this gap by developing a hierarchical *Games-in-Games* control architecture for regime-switching jump-diffusions. We construct a bilevel system where a fast inner layer solves a robust stochastic differential game within each mode, while a strategic outer layer actively modulates the *transition intensity kernel* of the underlying Markov chain. This structure, illustrated in Figure 1, formalizes the problem of *strategic regime control*, applicable to scenarios ranging from adversarial market microstructure to multi-modal resilient control.

Our contributions are threefold:

1. We provide a unified formulation for the games-in-games architecture on jump-diffusion spaces. By leveraging a unified Dynkin formula for switching diffusions [21], we decompose the bilevel problem into a hierarchy of

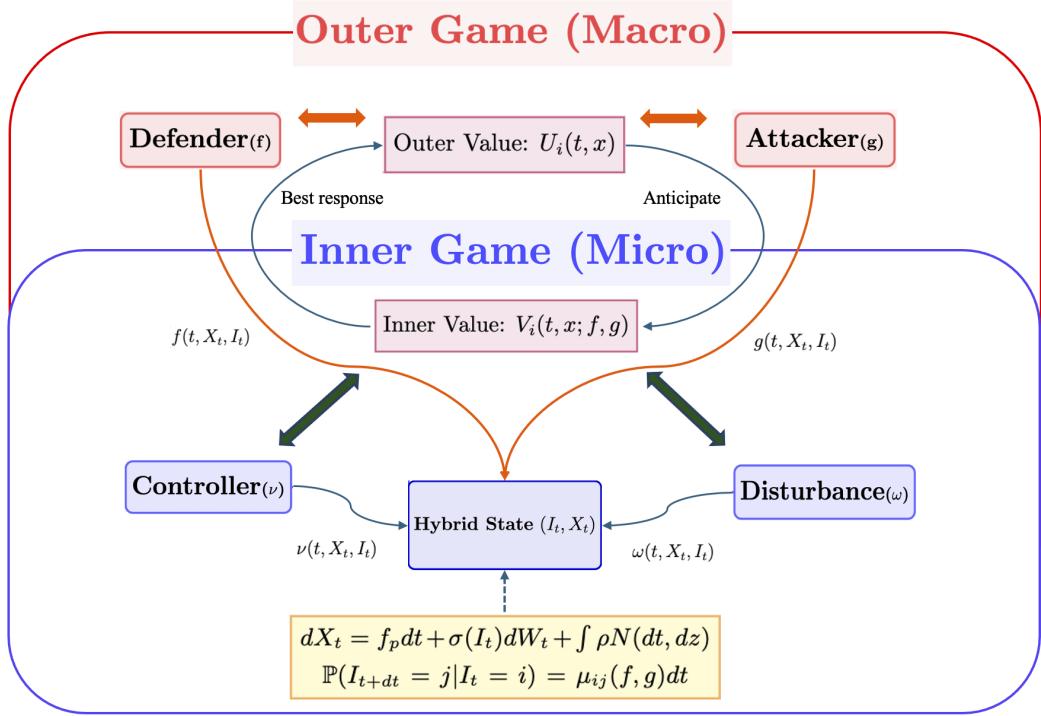


Figure 1: Compact Games-in-Games diagram: macro players shape regime switching; micro controls act on continuous dynamics under the active regime.

coupled Hamilton-Jacobi-Isaacs (HJI) equations. This separates the continuous control (inner Isaacs equation) from the strategic switching (outer Hamilton-Jacobi equation) without circular dependencies.

2. While general HJI equations are computationally demanding, we prove that for the class of Linear-Quadratic (LQ) and CARA-type exponential transformations (as in the market microstructure case study), the hierarchy collapses into a system of *coupled matrix differential equations* under certain conditions. This extends the classical coupled Riccati theory [13, 10] to the game-theoretic setting with endogenous transition rates.
3. We demonstrate the framework's efficacy through an inventory game case study on adversarial market making. We derive a risk isomorphism principle that allows us to characterize equilibrium policies that pre-emptively adjust control gains (inventory spreads) based on the stability gap between regimes.

The remainder of this paper is organized as follows. Section 2 formulates the two-layer hybrid game. Section 3 derives the viscosity solution hierarchy and the Dynkin transformation. Section 4 presents the spectral solution for the LQ case. Section 5 details the market microstructure application, and Section 6 concludes.

## 2 Problem Formulation

We consider a hybrid decision architecture in which a continuous state evolves according to mode-dependent stochastic dynamics, while a discrete mode process switches between a finite collection of regimes. Two layers of strategic decision-making interact: a fast-timescale controller/disturbance pair regulating the continuous state, and a slow-timescale pair of agents whose actions influence the transition rates among the discrete modes. This structure captures a wide range of multi-layer hybrid systems, including resilient infrastructure networks, multi-agent cyber-physical systems, and robust control under regime uncertainty.

Let  $(\Omega, \mathcal{F}, \mathbb{P})$  be a complete probability space equipped with a filtration  $\mathbb{F} = (\mathcal{F}_t)_{t \geq 0}$  satisfying the usual conditions of right-continuity and completeness. The time horizon is finite,  $T < \infty$ . We define the following primitive stochastic processes adapted to  $\mathbb{F}$ :

1.  $W = (W_t)_{t \geq 0}$  is a standard  $d$ -dimensional Brownian motion.

2.  $N(dt, dz)$  is a Poisson random measure on  $[0, T] \times \mathcal{Z}$  with intensity measure  $\lambda(dz)dt$ , where  $\mathcal{Z} \subseteq \mathbb{R}^k$ . We denote the compensated measure by  $\tilde{N}(dt, dz) = N(dt, dz) - \lambda(dz)dt$ .

**Definition 1** (Two-layer hybrid decision system). Let  $\mathcal{I} = \{1, \dots, N\}$  be a finite set of regimes (modes),  $U$  and  $W$  compact convex sets representing continuous-layer control and disturbance actions, and  $\mathcal{A}_D, \mathcal{A}_A$  finite sets of actions available to the two agents governing the mode transitions. A two-layer hybrid system is a tuple  $\Gamma = (\mathcal{X}, \mathcal{I}, \mathcal{V}, \mathcal{W}, \mathcal{F}, \mathcal{G})$ , where:

1.  $\mathcal{X} \subseteq \mathbb{R}^n$  is the continuous state space and  $\mathcal{I}$  is the finite mode space;
2. The continuous-layer policies

$$\nu : [0, T] \times \mathcal{X} \times \mathcal{I} \rightarrow U, \quad \omega : [0, T] \times \mathcal{X} \times \mathcal{I} \rightarrow W$$

are Borel measurable and  $\mathbb{F}$ -progressively measurable; the associated admissible policy classes are denoted  $\mathcal{V}$  and  $\mathcal{W}$ ;

3. The mode-selection policies

$$f : [0, T] \times \mathcal{X} \times \mathcal{I} \rightarrow \Delta(\mathcal{A}_D), \quad g : [0, T] \times \mathcal{X} \times \mathcal{I} \rightarrow \Delta(\mathcal{A}_A)$$

assign mixed actions from  $\mathcal{A}_D$  and  $\mathcal{A}_A$ . The admissible classes are denoted  $\mathcal{F}$  and  $\mathcal{G}$ .

Given  $(f, g)$  and  $(\nu, \omega)$ , the hybrid state process  $(X_t, I_t)_{t \in [0, T]}$  evolves as a Regime-Switching Jump-Diffusion defined by:

- (a) *Continuous dynamics:* Between jumps of the mode  $I_t$ , the continuous state  $X_t$  evolves according to the stochastic differential equation (SDE):

$$\begin{aligned} dX_t &= f_p(t, X_t, \nu_t, \omega_t; I_t)dt + \sigma(t, X_t; I_t)dW_t \\ &\quad + \int_{\mathcal{Z}} \rho(t, X_t, z; I_t)\tilde{N}(dt, dz), \end{aligned} \tag{1}$$

with  $X_0 = x_0 \in \mathcal{X}$ . Here,  $\nu_t = \nu(t, X_t, I_t)$  and  $\omega_t = \omega(t, X_t, I_t)$ .

- (b) *Discrete dynamics:* The mode process  $I_t \in \mathcal{I}$  is a controlled continuous-time Markov chain with generator matrix  $\Pi_t = [\mu_{ij}(t)]_{i,j \in \mathcal{I}}$  modulated by the outer policies:

$$\mathbb{P}(I_{t+dt} = j \mid I_t = i, X_t = x) = \mu_{ij}(f(t, x, i), g(t, x, i))dt + o(dt), \tag{2}$$

where  $\mu_{ii} = -\sum_{j \neq i} \mu_{ij}$ . The mode-selection policies  $(f, g)$  determine the regime transition intensities  $\mu_{ij} : \Delta(\mathcal{A}_D) \times \Delta(\mathcal{A}_A) \rightarrow [0, \infty)$ , yielding the instantaneous rate  $\mu_{ij}(f(t, X_t, I_t), g(t, X_t, I_t))$  at each  $t$ .

Above defines a measure  $\mathbb{P}^{\nu, \omega, f, g}$  on the path space  $D([0, T]; \mathcal{X} \times \mathcal{I})$ .

The performance of a policy tuple is evaluated through the cost functional:

$$J(f, g; \nu, \omega) = \mathbb{E} \left[ q_f(X_T, I_T) + \int_0^T c(t, X_t, \nu_t, \omega_t, I_t) dt \right]. \tag{3}$$

The two-layer decision architecture is expressed as the bi-level optimization problem

$$\begin{aligned} &\min_{f \in \mathcal{F}} \max_{g \in \mathcal{G}} \Phi(f, g; \nu^*, \omega^*) \\ \text{s.t. } &\nu^*, \omega^* \in \inf_{\nu \in \mathcal{V}} \sup_{\omega \in \mathcal{W}} J(f, g; \nu, \omega) \end{aligned} \tag{4}$$

where the inner minimax determines the continuous-layer value and the outer minimax governs mode manipulation. The functional  $\Phi(\cdot, \cdot; \nu^*, \omega^*) : \Delta(\mathcal{A}_D) \times \Delta(\mathcal{A}_A) \rightarrow \mathbb{R}$  represents the outer-layer preference and will be specified later.

**Assumption 1** (Standing assumptions).

(i) (Generalized Isaacs Condition) For each fixed mode  $i \in \mathcal{I}$  and any smooth test function  $\phi \in C^2(\mathcal{X})$ , we define the generalized Hamiltonian  $\mathcal{H}_i[\phi]$  acting on the state  $x$ , gradient  $p = \nabla \phi(x)$ , Hessian  $M = \nabla^2 \phi(x)$ , and controls  $(u, w)$ :

$$\begin{aligned} \mathcal{H}_i[\phi](t, x, p, M, u, w) &= c(t, x, u, w, i) + p^\top f_p(t, x, u, w; i) \\ &\quad + \frac{1}{2} \text{Tr}(\sigma(t, x; i) \sigma(t, x; i)^\top M) \\ &\quad + \int_{\mathcal{Z}} (\phi(x + \rho(t, x, z; i)) - \phi(x) - p^\top \rho(t, x, z; i)) \nu(dz; u, w). \end{aligned}$$

We assume that the minimax condition holds for all valid inputs:

$$\min_{u \in U} \max_{w \in W} \mathcal{H}_i[\phi](t, x, p, M, u, w) = \max_{w \in W} \min_{u \in U} \mathcal{H}_i[\phi](t, x, p, M, u, w).$$

(ii) The coefficients  $f_p, \sigma, \rho$  satisfy the standard Lipschitz and linear growth conditions in  $x$ , uniformly in  $(u, w)$ . The measure  $\nu(dz; u, w)$  is a bounded kernel satisfying appropriate integrability conditions. The costs  $c, q_f$  satisfy quadratic growth conditions.

**Lemma 1** (Existence and Estimates). *Under Assumption 1, for any admissible policies, the hybrid system (1)–(2) admits a unique strong solution  $(X_t, I_t)_{t \in [0, T]}$ . Furthermore, for any  $p \geq 1$ , there exists a constant  $C_p > 0$  such that:*

$$\mathbb{E} \left[ \sup_{t \in [0, T]} \|X_t\|^p \right] \leq C_p (1 + \|x_0\|^p).$$

*Proof.* Since the transition rates are bounded, the mode process  $I_t$  undergoes finitely many jumps in  $[0, T]$  almost surely. Between any two jump times, the system evolves as a standard SDE with jumps. Under the Lipschitz and linear growth conditions (Assumption 1), a unique strong solution exists for each interval (see e.g., [2, 15]). We construct the global solution  $(X_t)_{t \geq 0}$  by concatenating these trajectory segments at the jump times of  $I_t$ .

We apply Itô's formula to the function  $\phi(x) = \|x\|^p$ . The linear growth assumption on the drift  $f_p$  and jump intensity implies that the generator is bounded by  $\mathcal{L}\phi(x) \leq C(1 + \|x\|^p)$ . To bound the expectation of the supremum, we handle the martingale terms (diffusion and compensated jumps) using the Burkholder-Davis-Gundy (BDG) inequalities, which control the maximum of the stochastic integrals. Combining these bounds yields an integral inequality for  $g(t) = \mathbb{E}[\sup_{s \leq t} \|X_s\|^p]$ . The final result follows immediately from Grönwall's inequality.  $\square$

### 3 Cross-Layer Viscosity Solution

Once the hybrid architecture  $\Gamma$  in Definition 1 is equipped with Assumption 1, the coupled state process  $(X_t, I_t)$  is a Regime-Switching Jump-Diffusion. Between jumps of the discrete mode  $I_t$ , the continuous state  $X_t$  follows the stochastic evolution generated by the drift  $f_p$ , diffusion  $\sigma$ , and jump measure defined in (1). For any probe function

$$\phi : \mathcal{X} \times \mathcal{I} \times [0, T] \rightarrow \mathbb{R}, \quad \phi(\cdot, i, \cdot) \in C^{2,1}(\mathcal{X} \times [0, T]),$$

the infinitesimal generator of  $(X_t, I_t)$  under continuous-layer policies  $(\nu, \omega)$  and mode-selection policies  $(f, g)$  is given by the sum of the diffusion generator, the inner jump generator, and the regime-switching operator:

$$\begin{aligned} (\mathcal{L}^{f,g,\nu,\omega}\phi)(t, x, i) &= \frac{\partial \phi}{\partial t}(t, x, i) + \nabla_x \phi(t, x, i)^\top f_p(t, x, \nu, \omega; i) \\ &\quad + \frac{1}{2} \text{Tr}(\sigma(t, x; i) \sigma(t, x; i)^\top \nabla_{xx}^2 \phi(t, x, i)) \\ &\quad + \int_{\mathcal{Z}} (\phi(t, x + \rho(t, x, z; i), i) - \phi(t, x, i) - \nabla_x \phi^\top \rho(t, x, z; i)) \nu(dz) \\ &\quad + \sum_{j \neq i} \mu_{ij} (f(t, x, i), g(t, x, i)) [\phi(t, x, j) - \phi(t, x, i)]. \end{aligned} \tag{5}$$

Because  $(f, g)$  may depend explicitly on time and on the instantaneous continuous state, the generator (5) captures the full state-time dependence of the mode-switching rates.

**Lemma 2** (Dynkin identity). *Under Assumption 1 and fixed admissible policies  $(f, g, \nu, \omega)$ , the process*

$$M_\phi(\tau) = \phi(\tau, X_\tau, I_\tau) - \phi(t, x, i) - \int_t^\tau (\mathcal{L}^{f,g,\nu,\omega}\phi)(s, X_s, I_s) ds$$

*is a local martingale. If  $\phi$  and its derivatives are bounded, it is a martingale for any bounded stopping time  $\tau \leq T$ . Consequently,*

$$\mathbb{E}[\phi(\tau, X_\tau, I_\tau)] = \phi(t, x, i) + \mathbb{E} \int_t^\tau (\mathcal{L}^{f,g,\nu,\omega}\phi)(s, X_s, I_s) ds. \quad (6)$$

*Proof.* Let  $(T_k)_{k \geq 0}$  denote the jump times of the mode process  $I_t$  with  $T_0 = t$ . On each random interval  $[T_k, T_{k+1})$ , the mode  $I_t = i$  remains constant. The evolution of  $\phi(t, X_t, i)$  is governed by Itô's formula for semimartingales with jumps [2, Thm 4.4.7]:

$$\phi(T_{k+1}^-, X_{T_{k+1}^-}, i) - \phi(T_k, X_{T_k}, i) = \int_{T_k}^{T_{k+1}} \mathcal{L}_{inner}\phi ds + \mathcal{M}_{k,k+1},$$

where  $\mathcal{L}_{inner}$  represents the continuous drift, diffusion, and inner jump parts of the generator, and  $\mathcal{M}_{k,k+1}$  collects the stochastic integrals with respect to  $dW_t$  and  $\tilde{N}(dt, dz)$ , which are zero-mean martingales under the boundedness assumptions.

At the jump time  $T_{k+1}$ , the mode switches from  $i$  to  $j$  with intensity  $\mu_{ij}$ . The compensator for this discrete transition is exactly the regime-switching sum in (5). Summing these contributions over all intervals up to  $\tau$  and taking expectations eliminates the martingale terms, yielding the claimed identity.  $\square$

Lemma 2 connects the stochastic dynamics in Definition 1 to the value functions developed below. It justifies infinitesimal expansions of probe functions and supports the viscosity-solution formulations of the coupled Hamilton-Jacobi-Isaacs (HJI) equations [20].

### 3.1 Inner-Layer Hamilton-Jacobi-Isaacs Equation

Definition 1 shows that once the mode-selection policies  $(f, g)$  are frozen, the continuous-layer control and disturbance interact through a zero-sum stochastic differential game. The state evolves as a controlled Jump-Diffusion, coupled with mode-dependent switching intensities.

For fixed  $(f, g)$ , define the inner-layer value functions by

$$V_i(x, t; f, g) = \inf_{\nu \in \mathcal{V}} \sup_{\omega \in \mathcal{W}} \mathbb{E} \left[ \int_t^T c(s, X_s, \nu_s, \omega_s, I_s) ds + c_T(X_T, I_T) \middle| \mathcal{F}_t \right]. \quad (7)$$

**Lemma 3** (Inner-layer HJI). *For fixed mode-selection policies  $(f, g)$ , the family of inner-layer value functions  $\{V_i(\cdot, \cdot; f, g)\}_{i \in \mathcal{I}}$  is the unique viscosity solution (with appropriate growth conditions) of the following system of Partial Integro-Differential Equations (PIDE):*

$$\begin{aligned} -\partial_t V_i(t, x) &= \min_{u \in U} \max_{w \in W} \left\{ c(t, x, u, w, i) + \mathcal{L}_i^{u,w} V_i(t, x) \right. \\ &\quad \left. + \sum_{j \neq i} \mu_{ij} (f(t, x, i), g(t, x, i)) [V_j(t, x) - V_i(t, x)] \right\}, \\ V_i(T, x) &= c_T(x, i), \end{aligned} \quad (8)$$

for all  $(t, x) \in [0, T] \times \mathcal{X}$ . Here,  $\mathcal{L}_i^{u,w}$  is the local integro-differential operator:

$$\begin{aligned} \mathcal{L}_i^{u,w} \phi(x) &= \nabla_x \phi(x)^\top f_p(t, x, u, w; i) + \frac{1}{2} \text{Tr}(\sigma(t, x; i) \sigma(t, x; i)^\top \nabla_{xx}^2 \phi(x)) \\ &\quad + \int_{\mathcal{Z}} (\phi(x + \rho(t, x, z; i)) - \phi(x) - \nabla_x \phi(x)^\top \rho(t, x, z; i)) \nu(dz). \end{aligned}$$

*Proof.* We proceed in two steps: first establishing that the value function  $V_i$  is a viscosity solution (satisfying the subsolution and supersolution properties), and second invoking a comparison principle for uniqueness.

We only demonstrate the subsolution property as the supersolution argument is symmetric. Let  $\phi(t, x) \in C^{1,2}([0, T] \times \mathcal{X})$  be a smooth test function such that  $V_i(t, x) - \phi(t, x)$  achieves a local maximum at  $(\hat{t}, \hat{x})$  with  $V_i(\hat{t}, \hat{x}) = \phi(\hat{t}, \hat{x})$ .

From the Dynamic Programming Principle (DPP), for small  $h > 0$ :

$$V_i(\hat{t}, \hat{x}) \leq \inf_{\nu} \sup_{\omega} \mathbb{E} \left[ \int_{\hat{t}}^{\hat{t}+h} c(s, X_s, \nu_s, \omega_s, i) ds + V_{I_{\hat{t}+h}}(\hat{t} + h, X_{\hat{t}+h}) \right].$$

We decompose the expectation based on whether the mode  $I_s$  jumps during  $[\hat{t}, \hat{t} + h]$ . With probability  $1 - O(h)$ , no jump occurs. In this case,  $I_{\hat{t}+h} = i$ . Using  $V_i \leq \phi$  and applying Itô's formula to  $\phi$ :  $\mathbb{E}[\phi(\hat{t} + h, X_{\hat{t}+h}) - \phi(\hat{t}, \hat{x})] = \mathbb{E} \int_{\hat{t}}^{\hat{t}+h} (\partial_t \phi + \mathcal{L}_i^{u,w} \phi) ds$ . With probability rate  $\mu_{ij}(f, g)$ , the mode switches to  $j$ . The contribution to the expected value change is dominated by the difference  $V_j - V_i \approx V_j - \phi$ , the jump contribution is then:  $\int_{\hat{t}}^{\hat{t}+h} \sum_{j \neq i} \mu_{ij}(f, g) [V_j(s, X_s) - \phi(s, X_s)] ds + o(h)$ .

Substituting these expansions back into the DPP inequality and using  $V_i(\hat{t}, \hat{x}) = \phi(\hat{t}, \hat{x})$  to cancel the zero-order terms:

$$0 \leq \inf_{\nu} \sup_{\omega} \mathbb{E} \left[ \int_{\hat{t}}^{\hat{t}+h} \left( c + \partial_t \phi + \mathcal{L}_i^{u,w} \phi + \sum_{j \neq i} \mu_{ij}(f, g) [V_j - \phi] \right) ds \right] + o(h).$$

Dividing by  $h$  and letting  $h \downarrow 0$ , the mean value theorem applies. Since the inequality holds for all controls, we obtain:

$$-\partial_t \phi(\hat{t}, \hat{x}) - \inf_u \sup_w \{c + \mathcal{L}_i^{u,w} \phi\} - \sum_{j \neq i} \mu_{ij}(f, g) [V_j(\hat{t}, \hat{x}) - V_i(\hat{t}, \hat{x})] \leq 0.$$

This confirms the viscosity subsolution condition. The supersolution argument is symmetric using a local minimum.

The system (8) is a system of coupled non-linear PIDEs. Under Assumption 1 (Lipschitz coefficients, quadratic growth), the comparison principle for viscosity solutions of such systems holds [5, Thm 3.4]. Specifically, if  $U$  is a subsolution and  $V$  is a supersolution with  $U(T) \leq V(T)$ , then  $U \leq V$  on  $[0, T]$ . Since our value function is both, it is the unique solution.  $\square$

The family  $\{V_i\}$  encodes the inner-layer response to any fixed choice of mode-selection strategies  $(f, g)$ . In particular,  $V_i(x, t; f, g)$  can be regarded as the effective performance index associated with starting in mode  $i$  at state  $x$  and time  $t$ , factoring in the optimal continuous-time response to the induced regime uncertainty.

### 3.2 Outer-Layer Hamilton-Jacobi-Isaacs System

We now return to the outer-level problem (4), in which the two mode-selection agents choose strategies  $(f, g)$  that influence the mode-transition dynamics while anticipating the optimal inner-layer responses captured by Lemma 3. We interpret the outer-layer interaction as a *committed* (Stackelberg-type) Markov game: at time  $t$ , the macro-agents choose Markov (state-feedback) policies  $(f, g)$  on  $[t, T]$ , anticipating that the micro-layer subsequently plays the induced inner saddle-point feedback  $(\nu^*, f, g, \omega^*, f, g)$  associated with  $(f, g)$ .

For each fixed pair of mode-selection policies  $(f, g)$ , the inner-layer value functions  $V_i(\cdot, \cdot; f, g)$  are determined by the system (8). Let the optimal inner feedback policies be denoted by:

$$(u^*, f, g, w^*, f, g)(t, x, i) \in \arg \min_{u \in U} \max_{w \in W} \mathcal{H}_i[V_i](t, x, \nabla_x V_i, \nabla^2 V_i, u, w)$$

Define the corresponding coefficients:  $\bar{f}_p(t, x, i; f, g) := f_p(t, x, u^*, f, g, w^*, f, g; i)$ ,  $\bar{\sigma}(t, x, i; f, g) := \sigma(t, x; i)$ ,  $\bar{\nu}(dz; t, x, i, f, g) := \nu(dz; u^*, f, g, w^*, f, g)$ . Under  $(f, g)$ , the hybrid state  $(X_t, I_t)$  therefore evolves as a Jump-Diffusion driven by these effective coefficients, coupled with the regime transition rates  $\mu_{ij}(f, g)$ .

We model the outer-layer objective as a path integral whose running cost depends on the inner-layer value functions. Let  $\varphi : [0, T] \times \mathcal{X} \times \mathcal{I} \times \mathbb{R} \rightarrow \mathbb{R}$  denote a bounded, continuous outer-layer running cost, where the final argument will be instantiated with the scalar quantity  $V_i(t, x; f, g)$ .

For an initial condition  $(t, x, i)$  and mode-selection policies  $(f, g)$ , define the outer-layer performance functional

$$J_{\text{out}}(t, x, i; f, g) := \mathbb{E}_{f,g}^{t,x,i} \left[ \int_t^T \varphi(s, X_s, I_s, V_{I_s}(s, X_s; f, g)) ds + c_T(X_T, I_T) \right], \quad (9)$$

where  $\mathbb{E}_{f,g}^{t,x,i}$  denotes expectation with respect to the probability law induced by the closed-loop Jump-Diffusion dynamics and the transition rates  $\mu_{ij}(f, g)$ .

The outer-layer value functions are then

$$U_i(t, x) := \inf_{f \in \mathcal{F}} \sup_{g \in \mathcal{G}} J_{\text{out}}(t, x, i; f, g). \quad (10)$$

## Outer-layer Isaacs condition and HJI system

We impose an Isaacs condition at the outer layer.

**Assumption 2** (Outer-layer Isaacs condition). *For each tuple  $(t, x, i)$  and test function  $\psi$ , consider the outer-layer Hamiltonian*

$$\begin{aligned}\mathcal{H}_{\text{out}}[\psi](t, x, i, \alpha, \beta) := & \varphi(t, x, i, V_i^{\alpha, \beta}(t, x)) + \mathcal{L}_{\text{eff}}^{\alpha, \beta} \psi(x) \\ & + \sum_{j \neq i} \mu_{ij}(\alpha, \beta) [\psi(t, x, j) - \psi(t, x, i)],\end{aligned}\quad (11)$$

where  $\mathcal{L}_{\text{eff}}^{\alpha, \beta}$ , similar to what is defined in (5), is the generator associated with the closed-loop coefficients  $\bar{f}_p, \bar{\sigma}, \bar{\nu}$  (evaluated at mixed actions  $\alpha, \beta$ ). We assume that the minimax condition holds:

$$\inf_{\alpha \in \Delta(\mathcal{A}_D)} \sup_{\beta \in \Delta(\mathcal{A}_A)} \mathcal{H}_{\text{out}} = \sup_{\beta \in \Delta(\mathcal{A}_A)} \inf_{\alpha \in \Delta(\mathcal{A}_D)} \mathcal{H}_{\text{out}}. \quad (12)$$

Under Assumption 2 and the regularity inherited from Assumption 1, the outer-layer value functions satisfy the following system of HJI equations.

**Lemma 4** (Outer-layer HJI). *For each  $i \in \mathcal{I}$ , the outer-layer value function  $U_i$  is the unique viscosity solution of*

$$\begin{aligned}-\partial_t U_i(t, x) = & \inf_{\alpha \in \Delta(\mathcal{A}_D)} \sup_{\beta \in \Delta(\mathcal{A}_A)} \left\{ \varphi(t, x, i, V_i^{\alpha, \beta}(t, x)) \right. \\ & \left. + \mathcal{L}_{\text{eff}}^{\alpha, \beta} U_i(t, x) + \sum_{j \neq i} \mu_{ij}(\alpha, \beta) [U_j(t, x) - U_i(t, x)] \right\}, \\ U_i(T, x) = & c_T(x, i),\end{aligned}\quad (13)$$

for all  $(t, x) \in [0, T] \times \mathcal{X}$ .

The proof follows the same argument as Lemma 3. We observe that the outer optimization problem (10) is a generalized Bolza problem for a Regime-Switching Jump-Diffusion, where the drift, diffusion, and jump measure are determined by the closed-loop coefficients  $\bar{f}_p, \bar{\sigma}, \bar{\nu}$ .

Under Assumption 1(iii) (strict convex-concavity of the inner Hamiltonian), the inner feedback maps  $(u^{*, f, g}, w^{*, f, g})$  are continuous with respect to the gradient  $\nabla_x V_i$  [4]. Although  $\nabla_x V_i$  exists only in the generalized viscosity sense, the effective outer dynamics satisfy the necessary growth and continuity conditions to apply the standard dynamic programming principle. Consequently,  $U_i$  is characterized as the unique viscosity solution to (13) via the same expansion of the Dynkin identity employed in Lemma 3. Thus the outer-layer HJI system (13) mirrors the “generator-plus-minimax” structure of the inner-layer HJI (8), but with effective dynamics that incorporate the optimal inner-layer response.

Having established the necessary conditions for the inner layer (Lemma 3) and the outer layer (Lemma 4) independently, we now characterize the solution to the full bi-level problem (4).

**Theorem 1** (Feedback Stackelberg Equilibrium). *Let  $\mathcal{V}$  and  $\mathcal{U}$  be the spaces of admissible feedback strategies for the inner and outer layers, respectively. A strategy profile pair  $((u^*, w^*) \in \mathcal{V} \times \mathcal{W}, (f^*, g^*) \in \mathcal{F} \times \mathcal{G})$  constitutes a **Feedback Stackelberg Equilibrium** for the bi-level problem (4) if there exist value functions  $V(t, x, i)$  and  $U(t, x, i)$  that simultaneously satisfy the coupled HJI system:*

1. Given the outer strategy  $(f^*, g^*)$ , the inner value  $V$  satisfies the inner Isaacs equation (8), with  $(u^*, w^*)$  attaining the minimax of the local Hamiltonian  $\mathcal{H}_i$ . This guarantees that the inner layer is optimal for the current topology.
2. Given the inner value field  $V$ , the outer value  $U$  satisfies the outer Isaacs equation (13), where  $(f^*, g^*)$  attains the saddle point of the switching Hamiltonian. This guarantees that the outer layer optimizes the system objective subject to the inner layer’s best response.

Since the value functions  $V, U$  satisfy the dynamic programming equations over the entire domain  $[0, T] \times \mathbb{R}^n$ , the equilibrium strategy is time-consistent and constitutes a Subgame Perfect Equilibrium.

*Proof sketch.* Fix  $(f^*, g^*)$ . By the dynamic programming principle for the inner layer and the verification theorem for Isaacs equations, if  $V$  is a (sufficiently regular) solution of (8) and  $(u^*, w^*)$  attains the saddle condition (8), then the

associated controlled state-regime process achieves the inner game value and no admissible deviation of  $(u, w)$  can improve the follower's objective; hence  $(u^*, w^*)$  is a best response to  $(f^*, g^*)$ . Next, treat the resulting inner value field  $V$  as the induced continuation payoff entering the outer running cost. By the dynamic programming principle for the outer layer and the corresponding verification theorem, if  $U$  solves (13) and  $(f^*, g^*)$  attains (13), then no admissible deviation of  $(f, g)$  can improve the leader's objective given the follower's best-response mapping encoded by  $V$ ; hence  $(f^*, g^*)$  is optimal at the outer layer. Because both layers are characterized by HJI equations posed on  $[0, T] \times \mathbb{R}^n \times \mathcal{I}$  and the equilibrium strategies are feedback (Markov) and obtained from pointwise saddle conditions, the resulting policy is time-consistent on every subgame starting at any  $(t, x, i)$ , i.e., it is subgame-perfect.  $\square$

## 4 Case Study: Mode-Controlled Markov Jump Linear System

We now examine the important case in which the inner-layer differential game admits closed-form solutions. We focus on the *Linear-Quadratic-Gaussian (LQG)* setting without stochastic inner-layer jump, which yields a coupled family of matrix Riccati equations.

Fix a mode  $i \in \mathcal{I}$ , and assume the continuous state  $X_t \in \mathbb{R}^n$  evolves as a linear SDE controlled by affine drifts, with the control and disturbance action space being  $U \subseteq \mathbb{R}^{d_1}, W \subseteq \mathbb{R}^{d_2}$ :

$$dX_t = (A_i X_t + B_i u + D_i w)dt + \Sigma_i dW_t,$$

where  $A_i, B_i$ , and  $D_i$  are system matrices of proper dimensions,  $\Sigma_i \Sigma_i^\top \succ 0$  captures the regime-dependent volatility. Throughout this section, we assume that, The running and terminal costs are quadratic:  $c(t, x, u, w, i) = x^\top Q_i x + u^\top R_i u - w^\top S_i w$ ,  $c_T(x, i) = x^\top Q_{T,i} x$ , with  $Q_i \in \mathbb{R}^{n \times n}$ ,  $R_i \in \mathbb{R}^{d_1 \times d_1}$ , and  $S_i \in \mathbb{R}^{d_2 \times d_2}$ ,  $S_i \succ 0$ . Under the conditions that  $Q_i \succeq 0$ ,  $R_i \succ 0$ , and  $S_i \succ 0$ , the generalized Isaacs condition (Assumption 1) holds. Although the diffusion adds a constant trace term to the Hamiltonian, the saddle-point feedback  $(u_i^*, w_i^*)$  remains unique and affine in the gradient  $\nabla_x V_i$  ( $V_i(t, x) = \frac{1}{2}x^\top P_i(t)x$  is the value function):

$$u_i^*(t, x) = -R_i^{-1} B_i^\top \nabla_x V_i(x, t; f, g), \quad w_i^*(t, x) = S_i^{-1} D_i^\top \nabla_x V_i(x, t; f, g). \quad (14)$$

Substituting these into the inner-layer HJI (8) yields the Stochastic LQ-specialized HJI:

$$\begin{aligned} -\partial_t V_i &= x^\top Q_i x - \nabla_x V_i^\top B_i R_i^{-1} B_i^\top \nabla_x V_i + \nabla_x V_i^\top D_i S_i^{-1} D_i^\top \nabla_x V_i \\ &\quad + \nabla_x V_i^\top A_i x + \frac{1}{2} \text{Tr}(\Sigma_i \Sigma_i^\top \nabla_{xx}^2 V_i) + \sum_{j \neq i} \mu_{ij}(f, g)[V_j - V_i]. \end{aligned} \quad (15)$$

We adopt the quadratic ansatz  $V_i(t, x) = x^\top P_i(t)x + r_i(t)$ . The second-order term  $\nabla_{xx}^2 V_i = 2P_i(t)$  results in a trace term that decouples from the quadratic optimization. Consequently, the quadratic weight matrices  $\{P_i\}_{i \in \mathcal{I}}$  satisfy the *Coupled Riccati Differential Equations*:

$$-\dot{P}_i = Q_i + A_i^\top P_i + P_i A_i - P_i \Sigma_i^{ctrl} P_i + \sum_{j \neq i} \mu_{ij}(f, g)(P_j - P_i), \quad (16)$$

with boundary conditions  $P_i(T) = Q_{T,i}$  for  $i \in \mathcal{I}$ , where the control matrces  $\Sigma_i^{ctrl} = (B_i R_i^{-1} B_i^\top - D_i S_i^{-1} D_i^\top)$ . Note that while the stochastic noise affects the scalar offset  $-\dot{r}_i(t) = \text{Tr}(\Sigma_i \Sigma_i^\top P_i(t)) + \sum_{j \neq i} \mu_{ij}(f, g)(r_j(t) - r_i(t))$ , with  $r_i(T) = 0$ . We define the Metzler operator  $\mathcal{M}$ , trace operator  $\mathcal{T}$  and Riccati operator  $\mathcal{R}$ , the coupled Riccati flow can be written as:

$$\begin{aligned} -\dot{\mathbf{P}} &= \mathcal{R}(\mathbf{P}) + \mathcal{M}\mathbf{P}, \quad \mathbf{P}(T) = \mathbf{Q}_T \\ -\dot{\mathbf{r}} &= \mathcal{T}(\mathbf{P}) + \mathcal{M}\mathbf{r}, \quad \mathbf{r}(T) = 0 \end{aligned}$$

where  $\mathbf{r}, \mathbf{P}$  are stacked vector and matrix.

For outer players, we let there be discrete action sets  $\mathcal{A}_A$  and  $\mathcal{A}_D$ , and each action pair  $(a_A, a_D) \in \mathcal{A}_A \times \mathcal{A}_D$  corresponds to one jump rate matrix  $(\Lambda_{ij}^{(a_A, a_D)})_{ij \in \mathcal{I}}$ , let  $f, g$  be chosen from simplices  $\Delta(\mathcal{A}_A), \Delta(\mathcal{A}_D)$ , hence the resulting controlled transition rates are given by the expected jump intensities  $\mu_{ij}(f, g) = \bar{\mu}_{ij} + f^\top \Lambda_{ij} g$ , where  $\bar{\mu}_{ij}$  denotes the nominal baseline rate and  $\Lambda_{ij}$  collects the action-pair-dependent perturbations.

### 4.1 Hierarchical Solution and Outer-Layer Structure

To render the hierarchical game tractable, we define the outer running cost as the function of the instantaneous value matrices  $\varphi(t, x, i, V_i) = \varphi(P_i)$ , e.g., the trace  $\varphi(t, x, i) = \text{Tr}(P_i(t))$ , This cost captures the magnitude of the risk

exposure in regime  $i$ , incorporating both the Hessian of the value function and the covariance of the noise. Crucially, by using the trace, we project the matrix-valued risk into a scalar cost, allowing us to adopt a state-independent ansatz for the outer value function:  $U_i(t, x) = k_i(t)$ . Under this ansatz, the outer HJI equation reduces to a scalar differential equation driven by the inner risk source:

$$-\dot{k}_i(t) = \min_f \max_g \left( \text{Tr}(P_i(t)) + \sum_{j \neq i} \mu_{ij}(f, g)(k_j(t) - k_i(t)) \right), \quad (17)$$

with boundary condition  $k_i(T) = 0$ .

Crucially, while the term  $\text{Tr}(P_i(t))$  evolves dynamically based on the switching history, it is *exogenous* to the instantaneous optimization at time  $t$  (i.e., it does not depend explicitly on  $f, g$  at that instant). Under the bilinear transition model  $\mu_{ij}(f, g) = \bar{\mu}_{ij} + f^\top \Lambda_{ij} g$ , the optimization becomes the saddle-point solution of the state-independent game matrix  $\mathbf{M}_i(t) = \sum_{j \neq i} \Lambda_{ij} (k_j(t) - k_i(t))$ . The equilibrium policies  $(f^*(t), g^*(t))$  are the mixed-strategy saddle-point of this matrix game.

**Theorem 2.** Consider the hierarchical LQG case study with mode-only outer mixed strategies  $(f(t, i), g(t, i))$  inducing bilinear switching rates  $\mu_{ij}(f, g) = \bar{\mu}_{ij} + f^\top \Lambda_{ij} g$ , and assume the inner Isaacs regularity conditions so that the dynamic programming principle applies. A feedback equilibrium is characterized by the coupled backward system on  $[0, T]$ :

1. **Outer value and equilibrium switching.** Let  $\mathbf{k}(t) = (k_i(t))_{i \in \mathcal{I}}$  and define the (forced) outer flow

$$-\dot{\mathbf{k}}(t) = \varphi(t) + \mathcal{M}(\mu^*(t)) \mathbf{k}(t), \quad \mathbf{k}(T) = 0, \quad (18)$$

where  $(\mathcal{M}(\mu)\mathbf{z})_i := \sum_{j \neq i} \mu_{ij}(z_j - z_i)$ . At each  $(t, i)$ , define the local matrix game payoff

$$\mathbf{M}_i(t) := \sum_{j \neq i} \Lambda_{ij} (k_j(t) - k_i(t)), \quad (19)$$

and let  $(f_i^*(t), g_i^*(t))$  be any mixed saddle point of  $\mathbf{M}_i(t)$ . Then the equilibrium rates  $\mu^*(t)$  are obtained row-wise by

$$\mu_{ij}^*(t) = \bar{\mu}_{ij} + f_i^*(t)^\top \Lambda_{ij} g_i^*(t), \quad j \neq i, \quad \text{with } \mu_{ii}^*(t) = - \sum_{j \neq i} \mu_{ij}^*(t). \quad (20)$$

2. **Inner LQG (Riccati) equilibrium.** Given  $\mu^*(t)$ , the inner value admits the quadratic form  $V_i(t, x) = x^\top P_i(t)x + r_i(t)$  and the matrices  $\mathbf{P}(t) = (P_i(t))_{i \in \mathcal{I}}$  satisfy the coupled Riccati flow

$$-\dot{\mathbf{P}}(t) = \mathcal{R}(\mathbf{P}(t)) + \mathcal{M}(\mu^*(t)) \mathbf{P}(t), \quad \mathbf{P}(T) = \mathbf{Q}_T,$$

with the standard LQG saddle feedback laws  $u_i^*(t, x) = -R_i^{-1} B_i^\top (P_i(t)x)$  and  $w_i^*(t, x) = S_i^{-1} D_i^\top (P_i(t)x)$ .

*Proof sketch.* With mode-only outer strategies, the outer layer is a finite-state zero-sum switching game; dynamic programming yields (18). Under the bilinear rate model, the outer Hamiltonian separates row-wise and depends on  $(f_i, g_i)$  only through the matrix game (19); any mixed saddle point induces the equilibrium rates (20). Fixing  $\mu^*(t)$  reduces the inner layer to a Markov jump LQG Isaacs problem; the quadratic ansatz closes and the usual verification argument yields the coupled Riccati flow and feedback saddle controls.  $\square$

## 4.2 Spectral Structure and Adaptive Filtering

The coupled system defined in Theorem 2 describes a nonlinear feedback loop where the switching topology adapts to the magnitude of the underlying risk. We can interpret this physically using spectral operator theory, viewing the outer layer as an *adaptive filter* acting on the risk signals generated by the inner layer.

Let  $\mathcal{L}_{\Pi(t)}$  denote the graph Laplacian associated with the equilibrium switching rates  $\mu^*(t)$ . The outer flow equation can be rewritten as a forced heat equation on an evolving graph. The inner-layer  $\mathbf{P}(t)$  acts as the exogenous heat source, generating risk cost based on the local control authority and noise covariance. The outer-layer acts as the thermal medium, diffusing this risk across the network via the strategic coupling  $\mathcal{M}(\mu^*)$ .

#### 4.2.1 Adaptive Spectral Gap Modulation

Unlike a passive diffusion process, the strategic coupling creates an active feedback mechanism. The equilibrium strategy effectively modulates the *spectral gap* of the switching graph to match the heterogeneity of the inner risk source: (i) When the risk difference between regimes is large ( $k_j \gg k_i$ ), the game solver drives the transition rates  $\mu_{ij}^*$  high. This increases the *algebraic connectivity* (the second eigenvalue  $\lambda_2$ ) of the Laplacian  $\mathcal{L}_{\Pi^*}$ , where  $\Pi^* = (\mu_{ij}^*)_{ij \in \mathcal{I}}$  is the equilibrium rate matrix and  $\mathcal{L}_{\Pi^*} = -\Pi^*$ , causing rapid diffusion of the risk cost from expensive to cheap regimes. (ii) When the risks are balanced ( $k_j \approx k_i$ ), the switching incentives vanish. The rates  $\mu^*$  drop, the Laplacian spectral gap closes, and the regimes effectively decouple, isolating the local risks to prevent unnecessary transition costs.

**Proposition 1** (Two-Scale Turnpike Property). *The system exhibits a dual Turnpike behavior characterized by two distinct time scales:*

1. *Established by the local control authority, the convergence rate is governed by the Hamiltonian Spectral Gap  $\rho_H := \min_i |\text{Re}(\lambda(\mathbf{H}_i))|$ , where  $\mathbf{H}_i = \begin{bmatrix} A_i & -\Sigma_i^{ctrl} \\ -Q_i & -A_i^\top \end{bmatrix} \in \mathbb{R}^{2n \times 2n}$  are the Riccati Hamiltonian.*
2. *Established by the strategic switching, the convergence rate is governed by the Laplacian spectral gap  $\lambda_2(\Pi^*(t)) := \min\{\Re(\lambda) > 0 : \lambda \in \sigma(\mathcal{L}_{\Pi(t)})\}$ .*

*The global equilibrium is reached only when both the risk generation (inner) and the risk distribution (outer) have settled to their algebraic limits.*

*Proof Sketch.* We analyze the convergence in backward time  $\tau = T - t$ . Let  $\tilde{P}_i(\tau) = P_i(\tau) - P_{i,ss}$  denote the deviation from the algebraic Riccati solution. Linearizing the Riccati flow around  $P_{i,ss}$  yields  $\frac{d}{d\tau} \tilde{P}_i = A_{cl,i}^\top \tilde{P}_i + \tilde{P}_i A_{cl,i}$ ,  $A_{cl,i} = A_i - \Sigma_i^{ctrl} P_{i,ss}$ . It is well known that the eigenvalues of  $A_{cl,i}$  coincide with the stable spectrum of the Hamiltonian matrix  $\mathbf{H}_i$ . Hence,  $\|\tilde{P}_i(\tau)\| \leq C e^{-2\rho_H \tau}$ , where  $\rho_H$  is the Hamiltonian spectral gap.

The homogeneous part of the outer value dynamics satisfies  $\frac{d}{d\tau} \mathbf{k}(\tau) = -\Pi^*(\tau) \mathbf{k}(\tau)$ . Interpreting  $-\Pi^*$  as the Laplacian  $\mathcal{L}_{\Pi^*}$  of the switching graph, standard spectral graph theory implies exponential contraction of the disagreement subspace at rate  $\lambda_2(\Pi^*)$ , i.e.,  $\|\mathbf{k}(\tau) - \mathbf{k}_{avg}\| \leq C e^{-\lambda_2(\Pi^*)\tau}$ . Since the full equilibrium dynamics are the Cartesian product of these two layers, the global turnpike behavior emerges only once both contraction mechanisms have taken effect, establishing the stated two-scale property.  $\square$

## 5 Application: Cross-layer Avellaneda-Stoikov Game

To demonstrate the efficacy of the games-in-games architecture, we apply the framework to a high-frequency market making problem under regime uncertainty. We extend the classical Avellaneda–Stoikov (AS) inventory management model [3] to a hierarchical hybrid-systems framework with two coupled decision layers. At the inner layer, a zero-sum differential game is played between a *Market Maker* (MM), who controls inventory and quoting decisions, and a *Strategic Predator* (SP), who acts adversarially by perturbing the short-term price drift.

At the outer layer, a separate strategic game governs the evolution of market regimes. The outer players, the *macro-attacker* and the *macro-stabilizer*, select discrete actions that parameterize the generator of a controlled Markov jump process over market regimes. To simplify the exposition, we assume binary actions for  $\{\text{off}, \text{stab}\}$  macro-stabilizer and  $\{\text{off}, \text{att}\}$  for macro-attacker. Hence the mixed strategies can be each parameterized by a single parameter within  $[0, 1]$ . The macro strategy pair  $(f_t, g_t)$  thus induces a transition-rate matrix governing switches between calm, volatile, and stressed market conditions. Through these rate controls, the outer game shapes the stochastic environment faced by the inner inventory game.

### 5.1 Market Dynamics & Game Formulation

Let the market operate in one of  $N$  regimes,  $I_t \in \mathcal{I}$ . The mid-price  $S_t \in \mathbb{R}_+$  follows a controlled diffusion process:

$$dS_t = w_t dt + \sigma(I_t) dW_t, \quad (21)$$

where  $\sigma(i)$  is the regime-dependent volatility, and  $w_t$  is the drift controlled by the inner adversary SP.

The MM holds inventory  $q_t \in \mathcal{Q} = \{-Q_{\max}, \dots, Q_{\max}\}$  and cash  $m_t \in \mathbb{R}$ . The inventory dynamics are pure jump processes driven by the execution of limit orders:

$$dq_t = dN_t^b - dN_t^a, \quad (22)$$

where  $N_t^b$  and  $N_t^a$  are Poisson processes with intensities  $\Lambda^b(u^b) = Ae^{-ku^b}$  and  $\Lambda^a(u^a) = Ae^{-ku^a}$ , controlled by the MM's spreads  $u^b, u^a \in \mathbb{R}_{\geq 0}$  and the market depth parameters  $A, k$ . Mapping to our general framework (Def. 1), we have the physical state:  $X_t = (S_t, q_t, m_t)$ ,  $x_0 = (S_0, q_0, m_0)$ . Note that  $S_t$  and  $m_t$  are continuous, while  $q$  is discrete. The MM is the *micro-player* who chooses spread  $u_t = (u_t^a, u_t^b)$ ; The predator is the *micro-adversary* chooses  $w_t$  (price drift). The jump rate is state-independent, with magnitude  $\rho(z) = 1$ .

At the outer layer, the regime transition rates  $\mu_{ij}(f, g)$  are controlled by a *Macro-Attacker* ( $f_t$ ) who seeks to maximize the MM's disutility (or induce a “Crisis” regime where  $\sigma$  is high); and a *Stabilizer*, ( $g_t$ ) who Seeks to maintain the “Calm” regime. The outer value function  $U_i(t, q)$  is computed by substituting the inner value  $v_{i,q}$  into the outer objective.

We formulate the inner layer as a zero-sum differential game. The MM maximizes the Constant Absolute Risk Aversion (CARA) utility of terminal wealth,  $U(x_0) = -\exp(-\gamma(m_T + q_T S_T))$ , where  $\gamma \geq 0$  is the risk aversion parameter. The SP observes the MM's inventory  $q_t$  and exerts price pressure  $w_t$  to minimize the MM's Certainty Equivalent, subject to a quadratic cost  $\frac{1}{2\xi} w_t^2$  representing the capital cost or risk of manipulation.

## 5.2 Hierarchical Solution: Matrix Exponential and Approximate Equilibrium

Using the Ansatz  $V_i(t, S, q, m) = -\exp(-\gamma(m + qS + \theta_i(t, q)))$ , the inner Hamiltonian  $\mathcal{H}_i$  decomposes additively due to the separation of drift (price) and jump (execution) controls. The SP minimizes the Hamiltonian component associated with the price drift:  $\min_w [w\partial_S V - V \frac{1}{2\xi} w^2]$ . Using  $\partial_S V = -\gamma qV$  and factoring out  $-V > 0$ , the optimization yields a closed-form structural reaction function:

$$w^*(t, q) = -\xi\gamma q. \quad (23)$$

This strategy reveals a *mean-averting* behavior: if the MM is long ( $q > 0$ ), the SP pushes the price down ( $w^* < 0$ ) to devalue the position; if short, the SP pushes the price up.

Simultaneously, the MM maximizes the trading component:

$$\max_{u^a, u^b \in \mathbb{R}_{\geq 0}} \sum_{side \in \{a, b\}} \Lambda^{side}(u^{side}) \left(1 - e^{-\gamma(u^{side} + \Delta\theta_{side})}\right), \quad (24)$$

where  $\Delta\theta_a = \theta(q-1) - \theta(q)$  and  $\Delta\theta_b = \theta(q+1) - \theta(q)$ . This recovers the standard AS spread formula adjusted for inventory shadow cost. Substituting the optimal strategies back into the HJB equation, the predatory term contributes a quadratic penalty scaled by inventory size:  $w^*(-\gamma q) - \frac{1}{2\xi}(w^*)^2 = \frac{1}{2}\xi\gamma^2 q^2$ . Consequently, the inventory value function  $\theta_i(t, q)$  satisfies the coupled system of ODEs:

$$\begin{aligned} -\dot{\theta}_i(t, q) = & \underbrace{\frac{1}{2}\gamma\sigma_i^2 q^2}_{\text{Volatility Risk}} + \underbrace{\frac{1}{2}\xi\gamma^2 q^2}_{\text{Predatory Risk}} + \sum_{side \in \{a, b\}} \frac{A}{\gamma} \left(1 + \frac{\gamma}{k}\right)^{-\frac{k}{\gamma}} e^{-\gamma\Delta\theta_{side}} \\ & + \sum_{j \neq i} \mu_{ij}(f, g) \frac{1}{\gamma} \left(1 - e^{-\gamma(\theta_j - \theta_i)}\right). \end{aligned} \quad (25)$$

Under CARA utility, we define the value function and its exponential transformation as:

$$V_i(t, S, q, m) = -\exp(-\gamma[m + qS + \theta_i(t, q)]), \quad v_{i,q}(t) := e^{-\gamma\theta_i(t, q)}.$$

Substituting optimal quotes into the HJB equation reduces the system to a linear ODE:

$$\dot{v}(t) = M(t)v(t), \quad v(T) = \mathbf{1}.$$

The generator  $M$  acts on the state space stacked by regimes  $i = 1, \dots, N$  and inventory  $q \in \{-Q_{\max}, \dots, Q_{\max}\}$ . It decomposes into micro-structure and macro-switching blocks:  $M = D + (Q \otimes I_{|\mathcal{Q}|})$ . Here,  $Q$  is the regime transition matrix ( $Q_{ij} = \mu_{ij}$  for  $i \neq j$ , row-sums zero). The block-diagonal matrix  $D = \text{diag}(A_1, \dots, A_N)$  captures the micro-dynamics. Each block  $A_i$  is tridiagonal in the inventory dimension  $q$ , containing the diagonal risk & outflow:  $\frac{1}{2}\gamma^2(\sigma_i^2 + \xi\gamma)q^2 - (\Lambda_i^a + \Lambda_i^b)$ , and off-diagonal entries  $\Lambda_i^{side} e^{-\gamma u_i^{*,side}}$  at  $(q, q \pm 1)$ .

For outer parameters piecewise-constant on  $[t, T]$ , the solution is explicit:

$$v(t) = \exp(-M\tau)\mathbf{1}, \quad \tau := T - t. \quad (26)$$

**Regime mixing and expected variance.** For small horizons  $\tau$ , we expand the matrix exponential to characterize how regime uncertainty affects pricing. Using the Feynman-Kac representation, the inventory cost  $\theta_i(t, q)$  is driven by the expected accumulated variance. Expanding the solution to second order in  $\tau$  yields (when  $q$  is at the boundaries, remove the coefficient 2 in the *monopoly rent* term):

$$\theta_i(t, q) = \frac{q^2}{2} \left( \gamma w_i(\tau) + \gamma^2 \xi \tau \right) - \frac{2A}{\gamma} \left( 1 + \frac{\gamma}{k} \right)^{-k/\gamma} \tau + \mathcal{O}(\tau^3). \quad (27)$$

Here,  $w_i(\tau)$  represents the *expected integrated variance* starting from regime  $i$ . By expanding the generator  $e^{Qu} \approx I + Qu$ , we explicitly capture the regime mixing effect. Letting  $s_i := \sigma_i^2$ :

$$\begin{aligned} w(\tau) &:= \int_0^\tau [e^{Qu} s]_i du \approx \int_0^\tau [(I + Qu)s]_i du \\ &= \sigma_i^2 \tau + \frac{1}{2} \sum_{j \neq i} \mu_{ij} (\sigma_j^2 - \sigma_i^2) \tau^2 + \mathcal{O}(\tau^3). \end{aligned} \quad (28)$$

The  $\mathcal{O}(\tau^2)$  term captures the *Regime Risk*: the probability-weighted drift into different volatility states. Thus, for very short horizons, the MM prices using the current regime's volatility. As  $\tau$  increases, the pricing formula "bends" to incorporate the volatilities of connected regimes.

**Remark 1** (Risk Isomorphism). *The Hamiltonian separability preserves the tractability of the solution while providing a key insight. The presence of a strategic predator ( $\xi > 0$ ) is mathematically isomorphic to an increase in market volatility. The MM perceives an effective volatility  $\sigma_{eff}^2(i) = w_i(\tau)^2 + \xi\gamma$ . This implies that in the presence of predatory order flow, the optimal policy is to widen spreads and liquidate inventory faster, exactly as one would in a high-volatility environment.*

**Optimal quotes and time-varying spreads.** Defining the effective risk factor  $C_i(\tau) := \gamma w_i(\tau) + \gamma^2 \xi \tau$ , the inventory indifference pricing implies:  $\Delta\theta_a \approx \left(\frac{1}{2} - q\right) C_i(\tau)$ ,  $\Delta\theta_b \approx \left(q + \frac{1}{2}\right) C_i(\tau)$ . The resulting optimal spreads  $u^*$  is:

$$u_i^*(t, q) = u_i^{*,a} + u_i^{*,b} = \frac{1}{\gamma} \ln \left( 1 + \frac{\gamma}{k} \right) + \frac{1}{2} C_i(\tau),$$

This demonstrates that spreads widen pre-emptively based on future expected volatility  $w_i(\tau)$ , pricing in the macro-attacker's threat before the regime shift occurs.

### 5.3 Macro Equilibrium (Outer HJI)

Let  $f_t$  (Attacker) and  $g_t$  (Stabilizer) control the regime generator  $\mu_{ij}(f, g)$  and micro-parameters. We posit that the macro-agents optimize against the *anticipated* inventory cost priced in by the Market Maker. Thus, the outer-layer running cost  $\varphi_i(q; f, g)$  is defined directly by the risk function  $\theta_i$  over the horizon  $\tau$ :

$$\varphi_i(q; f, g) := \theta_i(t, q) \approx \frac{q^2}{2} \left( \gamma w_i(\tau; f, g) + \gamma^2 \xi \tau \right) - \frac{2A}{\gamma} \left( 1 + \frac{\gamma}{k} \right)^{-k/\gamma} \tau.$$

This modeling choice implies a *sentiment-driven interaction*: the macro-attacker seeks to maximize the Market Maker's forward-looking risk assessment (which drives liquidity drying), rather than merely the instantaneous volatility.

The macro value function  $U_i(t, q)$ , representing the cumulative market stress, satisfies the Isaacs equation:

$$-\partial_t U_i(t, q) = \min_{g \in \Delta(\mathcal{A}_D)} \max_{f \in \Delta(\mathcal{A}_A)} \left\{ \varphi_i(q; f, g) + \sum_{j \neq i} \mu_{ij}(f, g) (U_j(t, q) - U_i(t, q)) \right\}. \quad (29)$$

**Remark 2** (Behavioral Interpretation). *By utilizing the integrated variance  $w_i(\tau)$  within the running cost, this formulation creates a feedback loop where the macro-agents are highly sensitive to future regime risks. The switching probability enters twice: once in the MM's pricing ( $w_i$ ) and again in the outer value dynamics ( $\sum \mu_{ij} \Delta U$ ). This creates a "Hyper-Alert" equilibrium where attackers preemptively strike as soon as the expectation of future volatility rises, mirroring the self-fulfilling nature of liquidity crises.*

Let  $\Delta_{ij}(t, q) := U_j(t, q) - U_i(t, q)$  be the stability gap (the cost impact of switching from  $i$  to  $j$ ).

1. *Affine Control*: If the transition rates  $\mu_{ij}$  are controllable within  $[\underline{\mu}_{ij}, \bar{\mu}_{ij}]$ ,  $\mu_{ij}(f, g) = \mu_{ij}^0 + f \cdot \lambda_{ij}^{\text{att}} - g \cdot \lambda_{ij}^{\text{stab}}$  the optimization decouples into pointwise Bang-Bang switches. The Attacker ( $f$ ) maximizes the drift toward higher cost regimes, while the Stabilizer ( $g$ ) minimizes it:

$$f^*(t) = \mathbb{I}_{\{\sum_{j \neq i} \lambda_{ij}^{\text{att}} \Delta_{ij} < 0\}}, \quad g^*(t) = \mathbb{I}_{\{\sum_{j \neq i} \lambda_{ij}^{\text{stab}} \Delta_{ij} < 0\}}.$$

This highlights the conflict: when a regime switch is dangerous ( $\Delta_{ij} > 0$ ), the Attacker pushes the accelerator ( $\bar{\mu}$ ) while the Stabilizer slams the brake ( $\underline{\mu}$ ).

2. *Quadratic Costs*: We relax the bounded control assumption and instead impose quadratic effort penalties  $\frac{\rho_f}{2} f^2$  and  $\frac{\rho_g}{2} g^2$  in the outer Hamiltonian. Assuming the transition rates remain affine in effort ( $\mu_{ij} = \mu_{ij}^0 + f \lambda_{ij}^{\text{att}} - g \lambda_{ij}^{\text{stab}}$ ), the control-dependent part of the Hamiltonian is:

$$\mathcal{H}(f, g) \propto f \left( \sum_{j \neq i} \lambda_{ij}^{\text{att}} \Delta_{ij} \right) - g \left( \sum_{j \neq i} \lambda_{ij}^{\text{stab}} \Delta_{ij} \right) - \frac{\rho_f}{2} f^2 - \frac{\rho_g}{2} g^2.$$

The first-order optimality conditions ( $\partial_f \mathcal{H} = 0, \partial_g \mathcal{H} = 0$ ) yield explicit proportional feedback rules:

$$f^*(t) = \frac{1}{\rho_f} \left[ \sum_{j \neq i} \lambda_{ij}^{\text{att}} (U_j(t, q) - U_i(t, q)) \right]^+,$$

$$g^*(t) = \frac{1}{\rho_g} \left[ \sum_{j \neq i} \lambda_{ij}^{\text{stab}} (U_i(t, q) - U_j(t, q)) \right]^+,$$

where  $[x]^+ = \max(0, x)$ . This result characterizes the macro-agents as *variable-gain controllers*: the intensity of their intervention scales linearly with the severity of the stability gap. For instance, the Attacker exerts minimal effort when the system is robust ( $\Delta_{ij} \approx 0$ ) but surges activity proportionally as the MM's inventory vulnerability increases ( $\Delta_{ij} \gg 0$ ).

## 5.4 Numerical Illustration

We demonstrate the equilibrium strategies and risk isomorphism principle using calibrated Bitcoin (BTC) market data. The experiment compares two market making strategies facing a strategic predatory trader: 1. *vanilla AS*: standard AS strategy, unaware of predatory drift; 2. *equilibrium AS*: modified strategy using effective volatility  $\sigma_{\text{eff}}^2 = w_i(\tau)^2 + \xi \gamma$  to account for predatory risk.

We calibrate regime-switching parameters from Kraken ticker “BTC-USD” 30-minute OHLCV data (December 7-14, 2025). Using rolling volatility with  $K$ -means clustering, we identify two distinct regimes: 1. *stable regime* (regime 0):  $\sigma_0 = 0.2253$  (22.53% annualized); 2. *volatile regime* (regime 1):  $\sigma_1 = 0.5305$  (53.05% annualized). The volatility ratio is  $\sigma_1/\sigma_0 = 2.35$ .

Empirical transition matrix estimation yields a base transition rate of  $\mu_0 = 30$  per day, corresponding to an average holding time of 48 minutes per regime. Figure 2 shows the calibrated regime evolution with price dynamics colored by regime state (green for stable, red for volatile).

### 5.4.1 Counterfactual Simulation Design

We simulate 12 hours of market making activity (December 12, 2025, 15:00–03:00) with the following setup: starting in stable regime ( $I_0 = 0$ ) with initial price  $S_0 = \$90,863.90$ , the market making process lasts for 2,880 total steps, with each step counting for  $\Delta t = 15$  seconds. We set the price drift to be 0 so that only the predator affects the drift, whose optimal drift control is  $w^*(q) = -\xi \gamma q$  with cost coefficient  $\xi = 10.0$ . MM's risk aversion parameter and inventory constraint are set to be  $\gamma = 0.02$ ,  $q \in [-10, 10]$ . We fit the order arrival to follow Poisson intensity  $\Lambda(u) = \lambda_0 e^{-ku}$  with market depth  $A = 250,000$  per year, spread sensitivity  $k = 10$ . We simulate 1000 Monte-Carlo paths for statistical significance.

Both strategies face the same strategic predator who observes their inventory in real-time and applies adversarial drift. The key difference is that *Vanilla AS* uses the actual volatility  $\sigma_i$  in the spread formula, while *Equilibrium AS* uses the effective volatility  $\sigma_{\text{eff},i} = \sqrt{w_i(\tau)^2 + \xi \gamma}$  derived from Remark 1 (Risk Isomorphism).

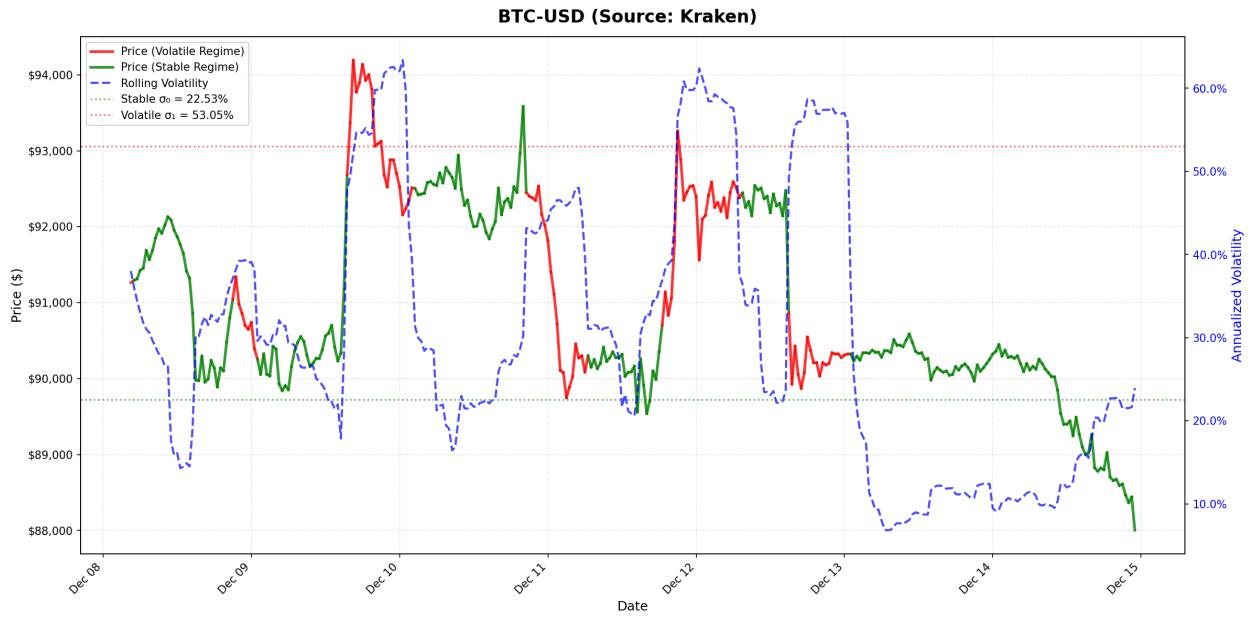


Figure 2:

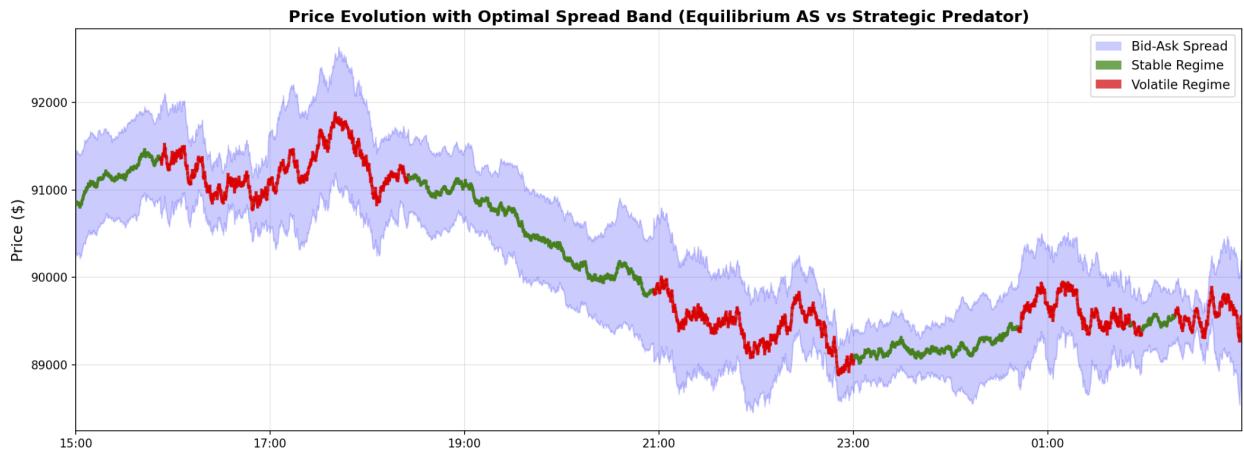


Figure 3: Sample BTC price evolution with equilibrium AS spread bands, (10 times the actual spread for clearer visualization.)

### 5.4.2 Results and Behavioral Analysis

Figure 3 presents one sample price evolution for the counterfactual simulation, with equilibrium AS spread bands, colored by regime state (green for stable, red for volatile). The simulation exhibits realistic regime dynamics with 6 regime switches over 12 hours, spending approximately equal time in each regime (54.5% volatile, 22.5% stable, calibrated from Figure 2).

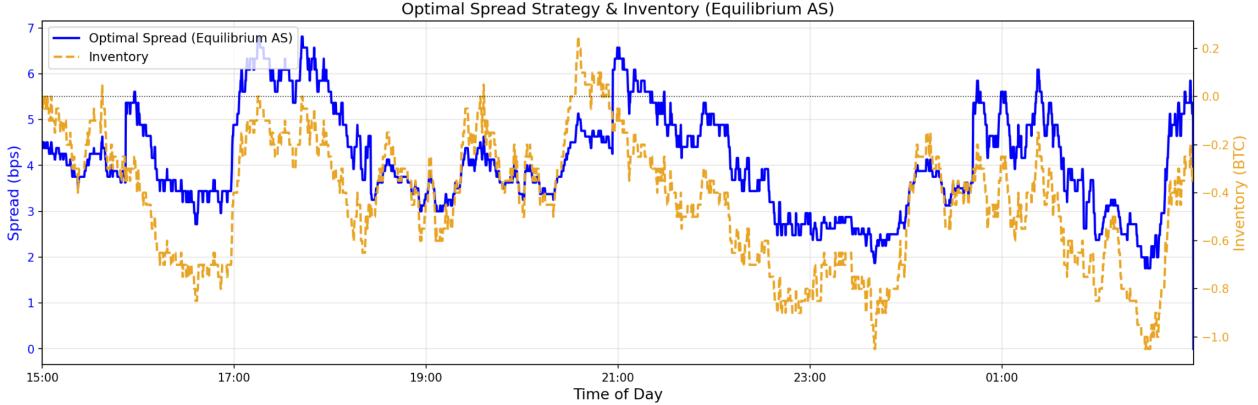


Figure 4: Sample optimal spread strategy evolution and inventory dynamics of MM under SP adverse selection.

Figure 4 shows the dynamic spread adjustment and inventory management. The equilibrium strategy adaptively widens spreads in volatile regimes, incorporating both the heightened market volatility  $\sigma_i$  and the predatory risk  $\xi\gamma$  into the reservation price calculation.

We plot the Profit & Loss (PnL) histogram across 1,000 Monte Carlo paths, from which we summarize: The equilibrium strategy achieves 111% higher mean PnL and 58% better Sharpe ratio despite facing a stronger predatory attack (16.4% higher average drift magnitude).

This seemingly counterintuitive result validates the risk isomorphism principle, in that wider spreads compensate for predatory risk, (equilibrium AS quotes 27% wider spreads against Vanilla AS), capturing the effective volatility increase from  $\sigma_{\text{eff}}^2 = w_i(\tau)^2 + \xi\gamma$ ; inventory accumulation is strategic according to equilibrium AS as it intentionally accumulates more inventory because the wider spreads not only provide adequate compensation for predatory drift exposure, and dominate fill rate loss: although wider spreads reduce order arrival rates by approximately 0.27% per basis point (via  $\Lambda(u) = \lambda_0 e^{-ku}$ ), the increased spread revenue more than compensates for reduced volume.

## 6 Conclusion

This paper presents a hierarchical “games-in-games” control framework for systems governed by *regime-switching jump-diffusions*. By decomposing the problem into a fast inner game and a strategic outer game, we derived a coupled system of Hamilton-Jacobi-Isaacs (HJI) equations via a unified Dynkin identity. A key theoretical contribution is the proof that for Exponential-Affine and Linear-Quadratic games, this hierarchy admits tractable *spectral solutions* via the matrix exponential, bridging the gap between hybrid modeling and computational feasibility.

The framework’s practical value is demonstrated through an adversarial market microstructure case study. The results reveal a *Risk Isomorphism* principle, where the hierarchical controller naturally interprets strategic predation as effective volatility, inducing a “hyper-alert” equilibrium that pre-emptively widens spreads. Future research will extend this architecture to partially observable regimes and integrate data-driven learning for empirical transition kernels.

## References

- [1] R. Aïd, M. Basei, G. Callegaro, L. Campi, and T. Vargioli. Nonzero-sum stochastic differential games with impulse controls: A verification theorem with applications. *Mathematics of Operations Research*, 45(1):205–232, 2020.
- [2] D. Applebaum. *Lévy processes and stochastic calculus*, volume 116. Cambridge university press, 2009.

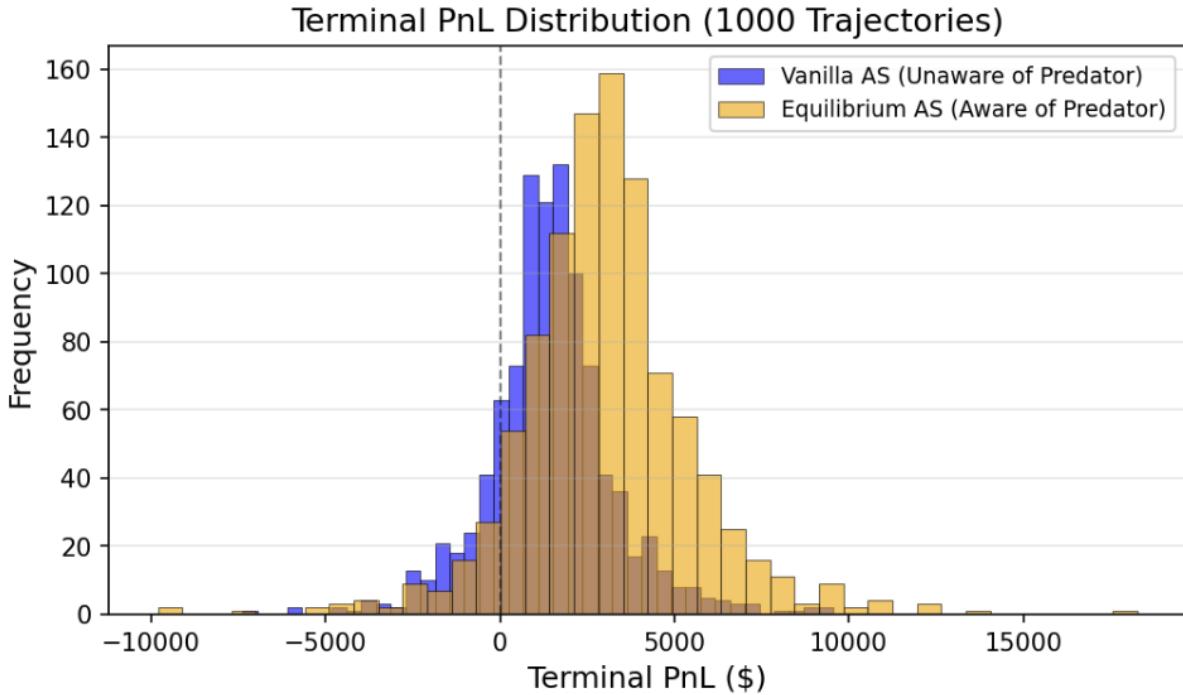


Figure 5: Distribution of terminal PnL over 1,000 simulated 12-hour market-making episodes (vanilla AS vs equilibrium AS with predator).

- [3] M. Avellaneda and S. Stoikov. High-frequency trading in a limit order book. *Quantitative Finance*, 8(3):217–224, 2008.
- [4] M. Bardi and I. Capuzzo-Dolcetta. *Optimal Control and Viscosity Solutions of Hamilton–Jacobi–Bellman Equations*. Birkhäuser, 2008.
- [5] G. Barles, R. Buckdahn, and E. Pardoux. Backward stochastic differential equations and integral-partial differential equations. *Stochastics: An International Journal of Probability and Stochastic Processes*, 60(1-2):57–83, 1997.
- [6] T. Başar and G. J. Olsder. *Dynamic Noncooperative Game Theory*. SIAM, Philadelphia, PA, 2nd edition, 1999.
- [7] R. Buckdahn and J. Li. Stochastic differential games and viscosity solutions of Hamilton-Jacobi-Bellman-Isaacs equations. *SIAM Journal on Control and Optimization*, 47(1):444–475, 2008.
- [8] Y. Chen and U. Vaidya. Deception attacks on networked control systems: A game-theoretic approach. *IEEE Transactions on Automatic Control*, 63(9):2905–2912, 2018.
- [9] O. L. V. Costa and F. Dufour. Stability of piecewise-deterministic Markov processes. *SIAM Journal on Control and Optimization*, 37(5):1483–1502, 1999.
- [10] O. L. V. Costa, M. D. Fragoso, and R. P. Marques. *Discrete-Time Markov Jump Linear Systems*. Springer, London, 2006.
- [11] M. H. A. Davis. Piecewise-deterministic Markov processes: A general class of non-diffusion stochastic models. *Journal of the Royal Statistical Society, Series B (Methodological)*, 46(3):353–388, 1984. With discussion.
- [12] B. Djehiche, S. Hamadène, and A. Popier. Optimal switching in finite horizon. *SIAM Journal on Control and Optimization*, 48(4):2663–2688, 2009. Foundational reference for optimal switching.
- [13] V. Dragan and T. Morozan. Game-theoretic coupled Riccati equations associated to controlled linear differential systems with jump Markov perturbations. *Stochastic Analysis and Applications*, 19(5):715–751, 2001.
- [14] X. Mao and C. Yuan. *Stochastic Differential Equations with Markovian Switching*. Imperial College Press, London, 2006.

- [15] B. Øksendal. *Stochastic differential equations: an introduction with applications*. Springer, 6 edition, 2005.
- [16] Y. Pan, T. Li, and Q. Zhu. Is stochastic mirror descent vulnerable to adversarial delay attacks? a traffic assignment resilience study. In *2023 IEEE 62nd Conference on Decision and Control (CDC)*, pages 8328–8333. IEEE, 2023.
- [17] Y. Pan, T. Li, and Q. Zhu. On the resilience of traffic networks under non-equilibrium learning. In *2023 American Control Conference (ACC)*, pages 3484–3489. IEEE, 2023.
- [18] Y. Pan and Q. Zhu. On poisoned wardrop equilibrium in congestion games. In *Decision and Game Theory for Security: 13th International Conference, GameSec 2022*, volume 13727 of *Lecture Notes in Computer Science*, pages 191–211, Cham, 2023. Springer.
- [19] F. Pasqualetti, F. Dörfler, and F. Bullo. Control-theoretic methods for cyberphysical security: Geometric principles for optimal cross-layer resilient control systems. *IEEE Control Systems Magazine*, 35(1):110–127, 2015.
- [20] H. Pham. *Continuous-time stochastic control and optimization with financial applications*, volume 61. Springer Science & Business Media, 2009.
- [21] G. Yin and C. Zhu. *Hybrid Switching Diffusions: Properties and Applications*, volume 63 of *Stochastic Modelling and Applied Probability*. Springer New York, 2010.
- [22] Q. Zhu and T. Başar. Game-theoretic methods for robustness, security, and resilience of cyberphysical control systems: Games-in-games principle for optimal cross-layer resilient control systems. *IEEE Control Systems Magazine*, 35(1):46–65, 2015.