Traded-Coordinate Diagnostics for Hedging Feedback

Mohammed AHNOUCH Erwan LESAOUT*

Initial version: November 6, 2025. Current version: March 3, 2026.

Abstract

When derivatives desks hedge in crowded markets, their collective trading flow can turn otherwise liquid hedging instruments into endogenous drivers of risk. This dealer flow distorts market prices and their covariation, causing desks to suffer outsized losses even when their pricing models appear locally sound. This paper provides a transparent, theorem-driven framework to make these feedback effects explicit, measurable, and manageable.

We separate the problem into two layers. First, we establish an exact hedging-error identity that directly links trading losses to the gap between anticipated and realized market covariation, rather than relying on standard Greeks. Second, we introduce a local market-clearing mechanism that shows exactly how dealer hedging scales and twists underlying fundamental risk into the observed market risk. This "covariation conjugation" yields a natural scalar metric (an amplification index) that flags when the market is destabilizing.

Crucially, our framework relies only on observable traded coordinates, bypassing the need for unobservable state assumptions. We demonstrate how to estimate directional feedback channels from standard market data, enabling desks to identify signature risks like correlation reversal or lead–lag asymmetry. We also provide practical risk-management tools, including pathwise loss bounds, early-warning indicators, and optimal filtering baselines. Through realistically parameterized simulations of autocallable equity notes, we validate these tools end-to-end, showing how structural feedback cascades can be both anticipated and quantified.

Acronyms and notation (quick reference). AI: amplification index $\mathsf{AI}_t = \|J_t\|$; AUC: area under the ROC curve; OU: Ornstein–Uhlenbeck; DIUO: down-and-in up-and-out (double barrier) put; P&L: profit and loss. $\widetilde{Y}$ denotes observed (feedback-distorted) traded coordinates, $\widetilde{Z}$ the counterfactual fundamental coordinates; c^Y, c^Z their covariation densities; B loop gain; $J = (I - B)^{-1}$ amplification matrix; $\mu_t := c_t^Y - \widehat{c}_t$ realized covariation distortion.

Keywords: hedging error; quadratic covariation; market impact; feedback; amplification; traded coordinates.

JEL: G13, C61. **MSC (2020):** 60H05, 91G20.

1 Introduction

When dealing with highly structured products or concentrated derivatives, the collective hedging activity of dealers can fundamentally alter the market they are trading in. In the Nikkei Uridashi episode of 2012, aggregate dealer losses on autocallable equity notes exceeded $500 million [1, 2]. As the Nikkei 225 index fell toward closely watched knock-in barriers, dealers holding similar short-gamma and short-vanna positions were forced to hedge simultaneously into a thinning market. Their resulting flow actively suppressed, and eventually reversed, the spot-surface correlations their pricing models had taken for granted. The same self-reinforcing dynamic recurred in the KOSPI autocallable crisis of 2018–2019, costing a single institution roughly EUR 260 million [3], and has resurfaced intraday in the modern boom of zero-days-to-expiry (0DTE) option trading [4, 5]. In each episode, hedging flow distorted the very covariation matrix

*Both authors: Université Paris 1 Panthéon–Sorbonne.

the desk's tools assumed was exogenous, and standard Greek-based P&L attribution failed to flag the danger.

The present paper contributes a tractable mathematical framework to analyze and manage these hedging-intensive feedback loops. Rather than relying on unobservable state variables, we represent the feedback directly within the observable, traded instruments of the desk. We model the market impact as a local amplification operator that actively twists the underlying, fundamental covariation into the realized market covariation that the desk actually faces. This practical framing provides a direct, exact decomposition of hedging losses, a clear mathematical link connecting fundamental to observable risks, and distinct market signatures that go beyond simply tracking spikes in market volatility.

We anchor this framework on two central theoretical results. First, we establish an exact *hedging-error identity* (Layer 1, theorem 2.7). For any derivative that can be priced via a basket of traded assets, its hedging P&L decomposes neatly into two distinct parts: a standard pricing error (reflecting model mismatch or discrete trading costs) and a feedback-induced penalty. The latter captures the exact cost of the market drifting away from the model's assumed covariance, heavily penalized by the structured product's curvature (its Hessian). Specifically, if we denote the structured product's value function by v, its Hessian by $H_t := D^2_{yy} v$, the desk's assumed covariation by $\widehat{c}$, and an ongoing pricing residual by $\mathfrak{r}_t$, the realized tracking error takes the straightforward form:

$$\widetilde{X}_T - v_T \;=\; \widetilde{X}_0 - v_0 \;-\; \int_0^T \mathfrak{r}_t \,\mathrm{d}t \;-\; \tfrac{1}{2}\int_0^T H_t : \mu_t \,\mathrm{d}t.$$

Crucially, $\mu_t := c_t^Y - \widehat{c}_t$ is the real-time gap between the market's true realized covariation c_t^Y and the desk's assumed covariation $\widehat{c}_t$. This identity holds regardless of the specific pricing model the desk employs, yielding an objective accounting truth.

Second, we formulate a *covariation conjugation* theorem that explains where this distortion μ_t comes from (Layer 2, theorem 3.11). By modeling the market as constantly clearing the pressure of dealer demand against background liquidity, we show that the observed market covariation c_t^Y is simply a conjugated version of the counterfactual fundamental risk c_t^Z:

$$c_t^Y = J_t \, c_t^Z \, J_t^\top, \qquad J_t = (I - B_t)^{-1}.$$

Here, B_t is the system's loop-gain matrix (capturing how aggressively hedging flow pushes back on asset prices), and the resulting matrix J_t acts as the feedback amplifier. The magnitude of this matrix, $\mathsf{AI}_t := \|J_t\|$, provides a scalar early-warning signal (theorem 3.14) indicating when the market is caught in a self-reinforcing cycle.

Building on these foundations, our key contributions are fourfold:

(i) **Exact Hedging Attribution**: We prove a model-independent accounting identity linking exact hedging errors to realized covariation distortion, sidestepping local Taylor truncations (theorem 2.7).

(ii) **Endogenous Risk Conjugation**: We prove that dealer flow mathematically subjects fundamental risk to a Jacobian conjugation yielding both an intuitive amplification index (AI_t) and an explicit formula for amplified portfolio curvature (theorem 3.11, theorem 3.13).

(iii) **Diagnostic Market Signatures**: We derive structural tells indicating the presence of hidden feedback, such as covariation attenuation, correlation sign reversals, local over-identifying moment restrictions, and lead–lag asymmetries (theorems 4.1, 4.7 and 4.8). We also provide practical equations to back out the amplification matrices directly from market data.

(iv) **Cumulative Risk Measurement and Early Warning**: We provide desks with four practical evaluation tools: a worst-case pathwise loss bound, a metric for detecting parametric resonance (Lyapunov exponents), an optimal filtering baseline linking precision to losses (Riccati bound), and an actionable early-warning signal based on ROC curve

analysis (section 5). We finally validate the entire system using a realistically initialized synthetic desk framework, demonstrating how structured feedback cascading across strikes can be flagged before outsized losses occur (fig. 12).

Section 2 develops the probabilistic setup and derives the hedging-error identity. Section 3 introduces the clearing hypothesis, proves the covariation conjugation, develops the amplified-curvature corollary and a two-dimensional benchmark, and applies the framework to barrier products motivated by the Uridashi episode. Section 4 shows how to identify feedback channels from lead–lag covariances and signature moments. Section 5 develops four cumulative risk tools: a pathwise loss bound, a Lyapunov exponent for parametric resonance, a Riccati filtering benchmark for the estimation-loss tradeoff, and a benchmark ranking result showing that under a location-shift model any strictly increasing transformation of the amplification index is Bayes-optimal for ranking tail-loss events. Section 7 discusses modelling scope, empirical status, and points to the full implementation procedure in section B. Section 8 concludes with a summary of contributions and natural extensions of the framework. Figure 1 provides a schematic overview of the logical dependencies among the main results.

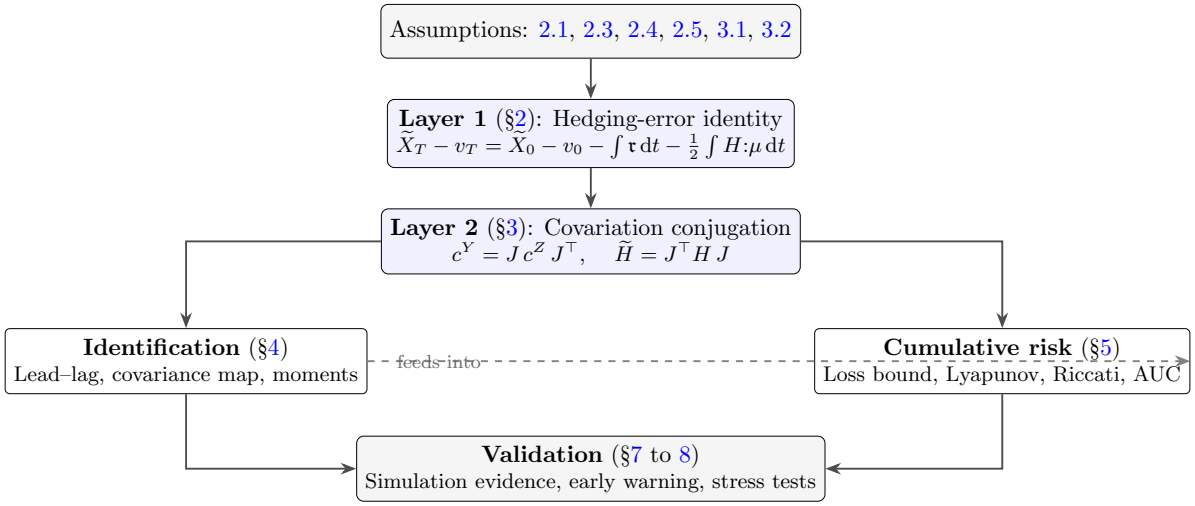

Figure 1: Logical architecture of the paper.

Figure 1 maps the dependency structure: Layer 1 derives the hedging-error identity from standard semimartingale assumptions; Layer 2 adds the clearing hypothesis to obtain covariation conjugation. The conjugation then feeds two toolkits (identification of feedback channels from observables (§4) and quantification of cumulative risk (§5)), which converge in the simulation-based validation.

Relative to the existing literature, the paper's value added lies in its object of analysis. Rather than studying feedback primarily through returns, volatility, or dealer-flow proxies, we represent it through a local amplification operator acting on covariation in traded coordinates. This shift in viewpoint brings together three strands that are usually treated separately: short-horizon hedging P&L attribution, endogenous distortion of covariance structure, and desk-level instability diagnostics. It also yields directional and topology-sensitive predictions (attenuation, reversal, over-identifying covariance relations, and lead–lag asymmetries) that are sharper than scalar volatility effects because they are tied directly to the structure of the local feedback map.

On the theoretical side, the hedging-feedback literature has progressed mainly through corrected pricing equations. Frey and Stremme [6] and Schönbucher and Wilmott [7] showed that feedback modifies the volatility entering the pricing equation; Loeper [8] established well-posedness; and Bouchard, Loeper, and Zou [9] developed almost-sure hedging under permanent impact. A parallel literature studies transient cross-impact and propagator models in optimal execution under exogenously specified kernels [10, 11, 12, 13, 14].

The present paper complements both programmes by moving the focus to the *realized covariation* faced by the desk. The conjugation $c^Y = Jc^Z J^\top$ is a statement about observable quadratic variation, and the hedging-error identity (theorem 2.7) holds for any desk pricing operator. Relative to the execution literature, the distinction is therefore substantive: rather than solving for an optimal trading policy under an exogenously specified impact kernel, we study how hedging-intensive dealer flow endogenously distorts observed second moments through a local amplification operator. The covariation distortion density μ_t is also the exact, path-level integrand minimized in expectation by the variance-optimal hedging tradition of Föllmer and Schweizer [15]; the framework makes its source (the amplification matrix J) structurally visible. In the same spirit, our conjugation provides a multivariate semimartingale-level counterpart to the exogenous/endogenous volatility decomposition of Wehrli and Sornette [16].

On the empirical side, Auh, Cho, and Foucault [17], Anderegg, Ulmann, and Sornette [18], and Egebjerg and Kokholm [19] document dealer-flow effects on volatility and market-maker hedging activity. Our framework complements that evidence with testable covariation signatures (lead–lag asymmetry, amplification-index spikes, and Lyapunov instability) that are designed to distinguish endogenous feedback from other sources of excess covariation.

The discriminative-power analysis (theorem A.2) adapts receiver operating characteristic methods to show that, under the stated location-shift assumptions, any strictly increasing transformation of the true amplification index is Bayes-optimal for ranking tail-loss events; it serves as a benchmark ranking result conditional on those assumptions.

At the cumulative-risk level, the analysis of section 5 draws on the multiplicative ergodic theorem of Oseledets [20] to establish the existence of Lyapunov exponents for the random propagator product, showing that time-varying feedback can produce exponential growth of the homogeneous propagator even when the system is instantaneously stable (parametric resonance), under a sufficient second-moment propagator bound. The Riccati-filter analysis adapts classical Kalman–Bucy theory to the feedback-amplified setting, where observation sensitivity scales with j^2; the resulting estimation-loss tradeoff is a structural consequence of the conjugation and serves as a canonical filtering benchmark that exhibits the locked scaling induced by amplification. The fast-relaxation limit that justifies instantaneous clearing (theorem 3.8) uses the singular perturbation theory of Khasminskii [21].

On the practitioner side, we show in section 7 that the standard decomposition of hedging P&L in terms of Greeks and implied-volatility parameters, codified by Bergomi [22], is a coordinate-dependent special case of the traded-coordinate identity. The mathematical toolkit draws on the functional Itô calculus of Cont and Fournié [23], which motivates future state-augmentation or functional extensions of theorem 2.7, and on the realized-covariation theory of Barndorff–Nielsen and Shephard [24], which underpins the estimation stage (section B).

2 Layer 1: Setup and Exact Hedging-Error Accounting

2.1 Setup and notation

Let $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\in[0,T]}, \mathbb{P})$ satisfy the usual conditions. Fix a strictly positive numéraire $(D_t)_{t\in[0,T]}$ of finite variation with $D_0 = 1$. Let

$$Y_t := (Y_t^0, Y_t^1, \dots, Y_t^n) \in (0,\infty)^{n+1}$$

denote the vector of traded hedging instruments (spot and liquid vanillas, or any chosen liquid coordinate set), and define the discounted traded coordinates

$$\widetilde{Y}_t := \frac{Y_t}{D_t} \in (0,\infty)^{n+1}.$$

Denote by $\mathbb{S}^{n+1}$ the space of real symmetric $(n+1)\times(n+1)$ matrices and by $\mathbb{S}_+^{n+1} \subset \mathbb{S}^{n+1}$ the positive-semidefinite cone.

Assumption 2.1 (Observed traded semimartingale with absolutely continuous covariation)**.** The discounted traded coordinates $\widetilde{Y}$ form a continuous semimartingale [25] whose quadratic covariation is absolutely continuous:

$$\mathrm{d}\langle \widetilde{Y} \rangle_t = c_t^Y \, \mathrm{d}t \quad \text{for a predictable process } c_t^Y \in \mathbb{S}_+^{n+1}.$$

Remark 2.2 (Why theorem 2.1 is natural under clearing). Under clearing, theorem 2.1 holds automatically: Itô's formula applied to the fixed-point map makes $\widetilde{Y}$ a continuous semimartingale, so absolute continuity of $\langle \widetilde{Y} \rangle$ is inherited from c^Z.

To accommodate structured products with monitored or state-augmented path dependence (autocallables, accumulators), we introduce a contract state variable. Let Ξ be a measurable state space. Let $\xi = (\xi_t)_{t\in[0,T]}$ be càdlàg, adapted, and piecewise constant with jumps only on a deterministic monitoring grid

$$0 = t_0 < t_1 < \cdots < t_m = T.$$

Let $\widetilde{C} = (\widetilde{C}_t)_{t\in[0,T]}$ be an adapted càdlàg finite-variation process representing cumulative discounted contractual cashflows, with jumps only at the monitoring times $\{t_k\}$; between monitoring dates, $\widetilde{C}$ is constant.

Assumption 2.3 (Finite-dimensional traded-coordinate representation and regularity)**.** There exists a function $v : [0,T] \times (0,\infty)^{n+1} \times \Xi \to \mathbb{R}$ such that the discounted product value satisfies

$$\widetilde{V}_t = v(t, \widetilde{Y}_t, \xi_t), \qquad t \in [0,T].$$

Thus Layer 1 is formulated for products that admit a finite-dimensional sufficient state in traded coordinates. This covers all monitored and state-augmented structured products, including autocallables, accumulators, and any path-dependent claim that can be embedded into such a finite-dimensional state. Moreover, for each $\bar{\xi} \in \Xi$, the map $(t,y) \mapsto v(t, y, \bar{\xi})$ is $C^{1,2}$ on $(t_k, t_{k+1}) \times U$ for some open neighborhood U containing the realized path of $\widetilde{Y}$ on (t_k, t_{k+1}) whenever $\xi \equiv \bar{\xi}$ on that interval. Finally, the local integrability condition

$$\int_0^T \|\nabla_y v(t, \widetilde{Y}_t, \xi_t)\|^2 \, \mathrm{d}\mathrm{Tr}\langle \widetilde{Y} \rangle_t < \infty \quad \text{a.s.}$$

holds, so that the stochastic integral $\int \nabla_y v^\top \, \mathrm{d}\widetilde{Y}$ is well defined.

Assumption 2.4 (Value matching at monitoring times)**.** At each monitoring time t_k and for all $y \in (0,\infty)^{n+1}$,

$$v(t_k^-, y, \xi_{t_k^-}) = v(t_k^+, y, \xi_{t_k^+}) + \Delta\widetilde{C}_{t_k}(y, \xi_{t_k^-}),$$

where $\Delta\widetilde{C}_{t_k} := \widetilde{C}_{t_k} - \widetilde{C}_{t_k^-}$ denotes the discounted cashflow paid at t_k (possibly state-dependent).

With the product representation in place, the final ingredient is the desk's view of covariation. Fix a predictable mapping

$$\widehat{c} : [0,T] \times (0,\infty)^{n+1} \times \Xi \times \Omega \to \mathbb{S}_+^{n+1}$$

interpreted as the desk-implied instantaneous covariation density in traded coordinates.

Assumption 2.5 (Desk pricing operator and residual)**.** On each interval on which the contract state ξ_t is constant, the value function v of theorem 2.3 satisfies

$$\partial_t v(t, \widetilde{Y}_t, \xi_{t-}) + \tfrac{1}{2} D_{yy}^2 v(t, \widetilde{Y}_t, \xi_{t-}) : \widehat{c}(t, \widetilde{Y}_t, \xi_{t-}) \;=\; \mathfrak{r}_t,$$

where $\widehat{c}(t, y, \xi) \in \mathbb{S}_+^{n+1}$ is a predictable desk-implied covariation operator chosen *ex ante* by the pricing/risk system, and $\mathfrak{r}_t$ is a predictable pricing residual capturing discretization error, model misspecification, and missing path-state effects. The operator $\widehat{c}$ is not defined from the realized quadratic-variation density c_t^Y of the observed process $\widetilde{Y}$; it is an independent input (e.g. from a local-volatility or stochastic-volatility calibration). Thus Layer 1 is an exact decomposition relative to a prescribed desk operator; it becomes economically informative precisely because $\widehat{c}$ is specified independently of the realized covariation of $\widetilde{Y}$. When the desk operator exactly solves the pricing PDE (the consistent Markovian case), $\mathfrak{r}_t \equiv 0$.

For $A, B \in \mathbb{S}^{n+1}$ define the Frobenius pairing $A : B := \mathrm{Tr}(AB)$. Define the realized feedback-induced covariation distortion density

$$\mu_t := c_t^Y - \widehat{c}(t, \widetilde{Y}_t, \xi_{t-}) \in \mathbb{S}^{n+1}.$$

Remark 2.6 (Three covariation objects). Three covariation processes appear throughout: c_t^Z (fundamental, unobserved), c_t^Y (observed, distorted by feedback), and $\widehat{c}(\cdot)$ (desk operator, chosen by the desk). The identity theorem 2.7 relates all three.

2.2 Exact hedging-error identity

The assumptions above yield the central identity of Layer 1. It is model-agnostic: no latent volatility model is assumed; only semimartingale prices and a desk operator $(v, \widehat{c})$ satisfying theorems 2.1 and 2.3 to 2.5.

Theorem 2.7 (Exact hedging-error identity in traded coordinates)**.** *Assume theorems 2.1 and 2.3 to 2.5. Define the predictable strategy*

$$\vartheta_t := \nabla_y v(t, \widetilde{Y}_t, \xi_{t-}) \in \mathbb{R}^{n+1}.$$

Let $\widetilde{X}$ be the discounted wealth process including discounted cashflows, defined by

$$\mathrm{d}\widetilde{X}_t = \vartheta_t^\top \, \mathrm{d}\widetilde{Y}_t - \mathrm{d}\widetilde{C}_t, \qquad \widetilde{X}_0 \in \mathbb{R}.$$

Then, on each monitoring interval (t_k, t_{k+1}) on which ξ is constant,

$$\widetilde{X}_{t_{k+1}^-} - v(t_{k+1}^-, \widetilde{Y}_{t_{k+1}}, \xi_{t_{k+1}^-}) = \widetilde{X}_{t_k^+} - v(t_k^+, \widetilde{Y}_{t_k}, \xi_{t_k^+}) - \int_{t_k}^{t_{k+1}} \mathfrak{r}_t \, \mathrm{d}t - \frac{1}{2} \int_{t_k}^{t_{k+1}} H_t : \mu_t \, \mathrm{d}t,$$

where

$$H_t := D_{yy}^2 v(t, \widetilde{Y}_t, \xi_{t-}), \qquad \mu_t := c_t^Y - \widehat{c}(t, \widetilde{Y}_t, \xi_{t-}).$$

Consequently,

$$\widetilde{X}_T - v(T, \widetilde{Y}_T, \xi_T) = \widetilde{X}_0 - v(0, \widetilde{Y}_0, \xi_0) - \int_0^T \mathfrak{r}_t \, \mathrm{d}t - \frac{1}{2} \int_0^T H_t : \mu_t \, \mathrm{d}t,$$

with the integral understood as the sum over monitoring subintervals. Thus the exact hedging error is decomposed into an operator-relative pricing residual and a curvature-weighted realized covariation distortion.

Remark 2.8 (Operator-relative interpretation and scope). The identity is exact once a desk operator $\widehat{c}$ and valuation function v have been fixed; $\mu_t = c_t^Y - \widehat{c}_t$ separates realized covariation from desk-implied covariation, and $\int \mathfrak{r}_t \, \mathrm{d}t$ is the residual relative to that operator. At each monitoring date t_k, the jump of the wealth process is $-\Delta\widetilde{C}_{t_k}$ and the jump of the value function is the same by theorem 2.4, so $\widetilde{X} - v$ is continuous across monitoring times and the interval identities telescope exactly. The leading minus sign means that excess realized covariation ($\mu_t > 0$) paired with positive curvature ($D_{yy}^2 v > 0$) produces a shortfall $\widetilde{X} - v < 0$, consistent with standard desk accounting: a short-gamma hedge portfolio loses when realized variance exceeds the model's.

Proof. Fix k and work on (t_k, t_{k+1}) where ξ is constant and $\mathrm{d}\widetilde{C}_t \equiv 0$. By Itô's formula [26] (using the absolute continuity of $\langle \widetilde{Y} \rangle$ from theorem 2.1),

$$\begin{aligned}\mathrm{d}v(t, \widetilde{Y}_t, \xi) &= \partial_t v(t, \widetilde{Y}_t, \xi)\ \mathrm{d}t + \nabla_y v(t, \widetilde{Y}_t, \xi)^\top\, \mathrm{d}\widetilde{Y}_t + \frac{1}{2}\, D^2_{yy} v(t, \widetilde{Y}_t, \xi) : \mathrm{d}\langle \widetilde{Y} \rangle_t \\ &= \partial_t v(t, \widetilde{Y}_t, \xi)\ \mathrm{d}t + \nabla_y v(t, \widetilde{Y}_t, \xi)^\top\, \mathrm{d}\widetilde{Y}_t + \frac{1}{2}\, D^2_{yy} v(t, \widetilde{Y}_t, \xi) : c^Y_t\ \mathrm{d}t.\end{aligned}$$

On (t_k, t_{k+1}), $\mathrm{d}\widetilde{X}_t = \vartheta_t^\top\, \mathrm{d}\widetilde{Y}_t$, and since $\widetilde{Y}$ is continuous, $\vartheta_t = \nabla_y v(t, \widetilde{Y}_t, \xi)$ for Lebesgue-a.e. t. Subtracting yields

$$\mathrm{d}(\widetilde{X}_t - v(t, \widetilde{Y}_t, \xi)) = -\partial_t v(t, \widetilde{Y}_t, \xi)\ \mathrm{d}t - \frac{1}{2}\, D^2_{yy} v(t, \widetilde{Y}_t, \xi) : c^Y_t\ \mathrm{d}t.$$

By theorem 2.5, $\partial_t v = \mathfrak{r}_t - \frac{1}{2}\, D^2_{yy} v : \widehat{c}$, hence

$$\mathrm{d}(\widetilde{X}_t - v(t, \widetilde{Y}_t, \xi)) = -\mathfrak{r}_t\ \mathrm{d}t - \frac{1}{2}\, D^2_{yy} v(t, \widetilde{Y}_t, \xi) : \mu_t\ \mathrm{d}t.$$

Integrating over (t_k, t_{k+1}) gives the interval identity.

At monitoring times t_k, the jump in $\widetilde{X}$ equals $-\Delta\widetilde{C}_{t_k}$ (the hedge portfolio pays out the contractual cashflow), while theorem 2.4 implies the jump in v is also $-\Delta\widetilde{C}_{t_k}$, so $\widetilde{X} - v$ has no jump at t_k. Summing over k yields the global-in-time statement (over $[0, T]$). □

Proposition 2.9 (Execution-cost add-on to the exact identity)**.** *Assume the setting of theorem 2.7 on a monitoring interval $[t_k, t_{k+1})$ on which ξ is constant. Let K be an adapted finite-variation process with $K_{t_k} = 0$, interpreted as cumulative discounted execution cost, and suppose the discounted wealth evolves according to*

$$\mathrm{d}\widetilde{X}_t = \vartheta_t^\top\ \mathrm{d}\widetilde{Y}_t - \mathrm{d}\widetilde{C}_t - \mathrm{d}K_t.$$

Then

$$\begin{aligned}\widetilde{X}_{t^-_{k+1}} - v(t^-_{k+1}, \widetilde{Y}_{t_{k+1}}, \xi_{t^-_{k+1}}) = \widetilde{X}_{t^+_k} - v(t^+_k, \widetilde{Y}_{t_k}, \xi_{t^+_k}) \\ - \int_{t_k}^{t_{k+1}} \mathfrak{r}_t\ \mathrm{d}t - \frac{1}{2}\int_{t_k}^{t_{k+1}} H_t : \mu_t\ \mathrm{d}t \\ - \big(K_{t^-_{k+1}} - K_{t^+_k}\big).\end{aligned}$$

Consequently, the global identity of theorem 2.7 acquires the additive term $-K_T$ when costs accumulate over $[0, T]$.

Proof. Repeat the proof of theorem 2.7 with the extra finite-variation term $-\mathrm{d}K_t$ in the wealth dynamics. The Itô cancellation of the first-order term is unchanged, so the only modification is the additional drift contribution $-\mathrm{d}K_t$, which integrates to the stated add-on. The monitoring-date telescoping is unaffected because K is already carried by the wealth process rather than by the value-matching condition. □

Remark 2.10 (Residual vs. distortion; technical extensions). Misspecification enters through $\mathfrak{r}_t$; if the desk PDE holds locally, $\mathfrak{r}_t \approx 0$ and the feedback term $-\frac{1}{2}\int H : \mu\, \mathrm{d}t$ dominates. The feedback mechanism enters through μ_t: under clearing, $c^Y = J c^Z J^\top$ gives μ_t a specific algebraic form driven by B, yielding falsifiable signatures (amplification spikes, lead–lag asymmetry, attenuation and reversal) that a generic misspecification narrative does not imply. Feedback losses concentrate where H_t and $\|J_t\|$ spike simultaneously (barrier-approach episodes), producing a tightly synchronised loss profile absent in diffuse model-error bleeds. With price jumps, Itô's formula acquires a second-order jump remainder (see [26, Ch. II, Thm 32]); the conjugation $c^Y_t = J_t c^Z_t J_t^\top$ applies only to the continuous quadratic-covariation density, and the discrete sum $\sum_u \Delta\widetilde{Y}_u\, \Delta\widetilde{Y}_u^\top$ must be handled separately.

Layer 2 endogenizes the distortion μ_t through the market-clearing mechanism.

3 Layer 2: Feedback Clearing, Conjugation, and Barrier Applications

3.1 Clearing hypothesis and amplification

The identity of theorem 2.7 shows that hedging error is driven by the covariation distortion μ_t, but is silent about its origin. This section introduces a structural equilibrium hypothesis that provides the missing mechanism. Let $\widetilde{Z}$ be a counterfactual "fundamental" traded-coordinate process in $\mathbb{R}^{n+1}$.

Assumption 3.1 (Fundamental no-arbitrage and absolutely continuous covariation)**.** There exists an equivalent measure $\mathbb{Q} \sim \mathbb{P}$ such that the discounted fundamental coordinates $\widetilde{Z}$ form a continuous $(\mathbb{Q}, \mathcal{F}_t)$-local martingale. Their quadratic covariation is absolutely continuous:

$$\mathrm{d}\langle \widetilde{Z} \rangle_t = c_t^Z \, \mathrm{d}t \quad \text{for a predictable process } c_t^Z \in \mathbb{S}_+^{n+1}.$$

Let I be an adapted continuous inventory/risk-limit state process taking values in an open set $\mathcal{I} \subset \mathbb{R}^r$, with pathwise finite variation on $[0,T]$ (so that $\langle I, I \rangle \equiv 0$ and $\langle \widetilde{Z}, I \rangle \equiv 0$). Let

$$\Phi : [0,T] \times (0,\infty)^{n+1} \times \mathcal{I} \to \mathbb{R}^m$$

be of class $C^{1,2,1}$ in (t, y, i), and let $\Lambda : \mathbb{R}^m \to \mathbb{R}^{n+1}$ be C^2.

Assumption 3.2 (Instantaneous clearing relation)**.** The observed traded coordinates satisfy

$$\widetilde{Y}_t = \widetilde{Z}_t + \Lambda(Q_t), \qquad Q_t := \Phi(t, \widetilde{Y}_t, I_t), \qquad t \in [0,T].$$

Moreover, along the realized path, the matrix

$$A_t := I_{n+1} - D_Q\Lambda(Q_t)\, D_y\Phi(t, \widetilde{Y}_t, I_t)$$

is invertible for all $t \in [0,T]$. Finally, the fixed point satisfies $\widetilde{Y}_t \in (0,\infty)^{n+1}$ for all $t \in [0,T]$, so that prices remain strictly positive and the pricing function v of theorem 2.3 is evaluated within its domain.

Why a fixed point. The self-referential structure reflects an economic reality: dealers choose hedge positions based on current traded prices (flow map Φ), executing those positions moves prices (impact map Λ), and observed prices must therefore be consistent with the flow they induce. Substituting $Q_t = \Phi(t, \widetilde{Y}_t, I_t)$ into $\widetilde{Y}_t = \widetilde{Z}_t + \Lambda(Q_t)$ closes the loop. The requirement is *local along the realized path* (invertibility of $I_{n+1} - B_t$ at each (t,ω)), not a global equilibrium over all price vectors. This is a reduced-form closure justified as the fast-relaxation limit of a well-posed dynamic adjustment process (theorem 3.8). The framework is agnostic to the specific microfoundation, relying only on the composite Jacobian $B_t = D_Q\Lambda(Q_t)\, D_y\Phi(t, \widetilde{Y}_t, I_t)$, which is what any impact model linearizes to [27, 28], making B and $J = (I - B)^{-1}$ the sufficient local summaries for second-order risk.

Define the clearing residual

$$F(t, y, z, i) := y - z - \Lambda(\Phi(t, y, i)).$$

Whenever $F(t, y, z, i) = 0$ we say (y, z, i) satisfy the clearing relation. Moreover,

$$\partial_y F(t, y, z, i) = I_{n+1} - D_Q\Lambda(\Phi(t, y, i))\, D_y\Phi(t, y, i).$$

Remark 3.3 (Existence, uniqueness, and the role of the implicit-function theorem). A sufficient condition for existence and uniqueness of $\widetilde{Y}_t$ for each (t,ω) is the Banach contraction condition

$\|D_Q\Lambda\| \cdot \mathrm{Lip}_y(\Phi) < 1$: the map $y \mapsto \widetilde{Z}_t + \Lambda(\Phi(t, y, I_t))$ is then a contraction in $\mathbb{R}^{n+1}$, and the Banach fixed-point theorem yields a unique solution.

Theorem 3.2 is weaker: it assumes only that the observed process $\widetilde{Y}$ satisfies the clearing relation and that A_t is nonsingular along the realized path. This is the condition needed for the local branch calculation and the covariation conjugation.

Accordingly, the implicit-function theorem is used only *locally along the realized path*: it provides, in a neighborhood of each realized point $(t, \widetilde{Y}_t, \widetilde{Z}_t, I_t)$, a differentiable branch and its Jacobian. The framework requires only a locally defined measurable selector $(t, z, i) \mapsto y^*(t, z, i)$ along the realized path, which is precisely what the implicit-function theorem delivers.

Lemma 3.4 (Measurable dependence under a global contraction strengthening)**.** *Assume, in addition to the regularity already imposed on Φ and Λ, that there exists $\kappa < 1$ such that*

$$\sup_{(t,y,i)} \left\| D_Q\Lambda(\Phi(t, y, i))\, D_y\Phi(t, y, i) \right\| \le \kappa,$$

and that for each fixed (y, i) the map $t \mapsto \Phi(t, y, i)$ is measurable. Then the fixed point equation

$$y = z + \Lambda(\Phi(t, y, i))$$

has a unique solution $y^*(t, z, i)$ for every (t, z, i), and the map $(t, z, i) \mapsto y^*(t, z, i)$ is jointly Borel measurable. Consequently, if $(\widetilde{Z}_t, I_t)$ is adapted, then $y^*(t, \widetilde{Z}_t, I_t)$ is adapted.

Proof. For fixed (t, z, i) define $T_{t,z,i}(y) := z + \Lambda(\Phi(t, y, i))$. The displayed bound makes $T_{t,z,i}$ a contraction with constant at most κ, so Banach's fixed-point theorem gives a unique fixed point. Starting from $y^{(0)}(t, z, i) := z$ and iterating $y^{(m+1)} := T_{t,z,i}(y^{(m)})$, each Picard iterate is jointly measurable because Φ and Λ are continuous in (y, i) and measurable in t. Uniform geometric convergence of the iterates to y^* implies joint measurability of the limit. Adaptedness follows by composition with the adapted pair $(\widetilde{Z}_t, I_t)$. □

Two derived objects will appear throughout the paper: the loop-gain matrix, which measures the marginal effect of one round of feedback, and the amplification matrix, which sums all feedback rounds. Define

$$B_t := D_Q\Lambda(Q_t)\, D_y\Phi(t, \widetilde{Y}_t, I_t), \qquad J_t := (I_{n+1} - B_t)^{-1}.^{1}$$

Remark 3.5 (Why J_t is the operative object). Linearizing the fixed-point equation yields the Jacobian $J_t = (I - B_t)^{-1}$, which makes feedback quantitative: it is the local multiplier $\mathrm{d}\widetilde{Y}_t \approx J_t\, \mathrm{d}\widetilde{Z}_t$, the operator conjugating fundamental into observed covariation ($c_t^Y = J_t c_t^Z J_t^\top$), and the sufficient statistic for all downstream identification and risk targets (lead–lag signatures, loss bound, Lyapunov exponent). Its operator norm $\mathsf{AI}_t = \|J_t\|$ measures distance to singularity of the clearing Jacobian, and the AI^2 scaling in the loss bound arises because one factor of $\|J\|$ comes from each side of $\mathrm{Tr}(J^\top H J\, c^Z)$.

Lemma 3.6 (Local clearing branch along the realized path)**.** *Assume theorem 3.2. Then for every point on the realized path there exists a neighborhood on which the equation*

$$F(t, y, z, i) = 0$$

defines a unique local branch

$$y = y^*(t, z, i),$$

[1] When $\rho(B_t) < 1$ (e.g. $\|B_t\| < 1$), the Neumann series $J_t = \sum_{k=0}^{\infty} B_t^k$ converges and the k-th term can be read as the k-th feedback round. This expansion provides intuition but is not needed: the inverse form $J_t = (I_{n+1} - B_t)^{-1}$ is the operative definition, and all results below hold on any time window where $A_t = I_{n+1} - B_t$ is nonsingular along the realized path.

with y^* of class $C^{1,2,1}$ in (t,z,i). Moreover, along the realized path,

$$J_t := \partial_z y^*(t,\widetilde{Z}_t,I_t) = \big(I_{n+1} - B_t\big)^{-1}, \qquad B_t := D_Q\Lambda(\Phi(t,\widetilde{Y}_t,I_t))\, D_y\Phi(t,\widetilde{Y}_t,I_t),$$

and the process J_t is predictable.

Proof. The statement is pathwise. Fix a point on the realized path and apply the classical implicit function theorem to $F(t,\cdot,\cdot,i)$ in a neighborhood of $(\widetilde{Y}_t,\widetilde{Z}_t,I_t)$. The regularity of Φ and Λ implies that F is $C^{1,2,1}$ in (t,y,z,i), while the invertibility of $\partial_y F$ yields a unique local branch $y^*(t,z,i)$ of the same regularity. Differentiating the identity $F(t,y^*(t,z,i),z,i)=0$ in z gives $(\partial_y F)\,\partial_z y^* + \partial_z F = 0$. Since $F(t,y,z,i) = y - z - \Lambda(\Phi(t,y,i))$, we have $\partial_z F = -I_{n+1}$ and $\partial_y F = I_{n+1} - D_Q\Lambda(\Phi(t,y,i))\, D_y\Phi(t,y,i) = I_{n+1} - B$. Therefore $\partial_z y^* = (I_{n+1}-B)^{-1}$, and evaluating at $(t,\widetilde{Z}_t,I_t)$ gives the stated J_t. Finally, B_t is predictable because it is obtained by composing predictable processes $(t,\omega)\mapsto(t,\widetilde{Y}_t(\omega),I_t(\omega))$ with deterministic C^1 maps; matrix inversion is continuous on the set of nonsingular matrices, hence $J_t = (I_{n+1}-B_t)^{-1}$ is predictable as well. □

The instantaneous clearing of theorem 3.2 is the fast-relaxation limit of a dissipative dynamics. Fix $\tau>0$. Consider the *relaxation* dynamics

$$\mathrm{d}\widetilde{Y}_t^\tau = \frac{1}{\tau}\Big(\widetilde{Z}_t + \Lambda(\Phi(t,\widetilde{Y}_t^\tau,I_t)) - \widetilde{Y}_t^\tau\Big)\,\mathrm{d}t \;+\; \Gamma_t\,\mathrm{d}W_t, \qquad \widetilde{Y}_0^\tau = y^*(0,\widetilde{Z}_0,I_0), \tag{1}$$

where W is a p-dimensional Brownian motion and Γ is predictable and locally bounded. This model is an extension layer: it does not alter Layer 1 (accounting) and reduces to instantaneous clearing as $\tau\downarrow 0$. We use the clearing residual F from section 3.

Assumption 3.7 (Regularity and diffusion setup for fast relaxation)**.** On a fixed finite horizon $[0,T]$, let the fundamental process admit the diffusion decomposition

$$\mathrm{d}\widetilde{Z}_t = b_t^Z\,\mathrm{d}t + \Sigma_t^Z\,\mathrm{d}W_t, \tag{2}$$

where W is a p-dimensional Brownian motion, b^Z is predictable, and Σ^Z is predictable. The inventory state I is continuous, adapted, and of finite variation on $[0,T]$.

There exists an open neighborhood $\mathcal{U}$ of the reference graph

$$\Gamma^* := \{(t,y^*(t,\widetilde{Z}_t,I_t),\widetilde{Z}_t,I_t) :\, t\in[0,T]\}$$

such that the clearing residual

$$F(t,y,z,i) := y - z - \Lambda(\Phi(t,y,i))$$

is C^1 in t and C^3 in (y,z) on $\mathcal{U}$, the Jacobian D_yF is uniformly invertible on $\mathcal{U}$, and there exists $\eta>0$ such that

$$\langle v, \partial_y F(t,y,z,i)\, v\rangle \ge \eta\|v\|^2 \qquad \text{for all } v\in\mathbb{R}^{n+1} \tag{3}$$

throughout $\mathcal{U}$.

Let $y^*(t,z,i)$ denote the unique local branch solving $F(t,y,z,i)=0$. Assume that $y^*(\cdot,\cdot,I_\cdot)\in C^{1,2}$ along the realized path and, moreover,

$$\mathbb{E}\Big[\sup_{t\in[0,T]}\big(\|b_t^Z\|^2 + \|\Sigma_t^Z\|_F^2 + \|\Gamma_t\|_F^2 + \|\partial_t y^*(t,\widetilde{Z}_t,I_t)\|^2 + \|\partial_z y^*(t,\widetilde{Z}_t,I_t)\|^2 + \|D_{zz}^2 y^*(t,\widetilde{Z}_t,I_t)\|^2\big)\Big] < \infty.$$

Finally, the relaxation dynamics are driven by the same Brownian motion:

$$\mathrm{d}\widetilde{Y}_t^\tau = -\frac{1}{\tau}F(t,\widetilde{Y}_t^\tau,\widetilde{Z}_t,I_t)\,\mathrm{d}t + \Gamma_t\,\mathrm{d}W_t, \qquad \widetilde{Y}_0^\tau = y^*(0,\widetilde{Z}_0,I_0), \tag{4}$$

and (4) has a unique strong solution for each $\tau>0$.

Theorem 3.8 (Fast relaxation limit yields instantaneous clearing)**.** *Assume theorem 3.7. Then there exists a constant $C_T < \infty$, independent of $\tau \in (0, 1]$, such that*

$$\sup_{t\in[0,T]} \mathbb{E}\Big[\|\widetilde{Y}_t^\tau - y^*(t, \widetilde{Z}_t, I_t)\|^2\Big] \le C_T\,\tau. \tag{5}$$

Hence

$$\widetilde{Y}^\tau \to y^*(\cdot, \widetilde{Z}_\cdot, I_\cdot) \quad \text{in } L^2(\Omega) \text{ uniformly on } [0, T].$$

If the initial condition is arbitrary, $\widetilde{Y}_0^\tau = y_0$, then

$$\mathbb{E}\|\widetilde{Y}_t^\tau - y^*(t, \widetilde{Z}_t, I_t)\|^2 \le \|y_0 - y^*(0, \widetilde{Z}_0, I_0)\|^2 e^{-2\eta t/\tau} + C_T\tau.$$

Proof. Set $e_t^\tau := \widetilde{Y}_t^\tau - y^*(t, \widetilde{Z}_t, I_t)$. Itô's formula applied to $y^*(t, \widetilde{Z}_t, I_t)$ gives $\mathrm{d}y^* = a_t\,\mathrm{d}t + G_t\,\mathrm{d}W_t$ with $G_t = \partial_z y^*\,\Sigma_t^Z$; the integrability assumption in theorem 3.7 provides a uniform bound $C < \infty$ on $\sup_t(\|a_t\|^2 + \|G_t\|_F^2)$. Subtracting dynamics and applying Itô's formula to $\|e_t^\tau\|^2$, then using the strong-monotonicity condition (3) and the Young inequality $2ab \le \eta\tau^{-1}a^2 + \tau\eta^{-1}b^2$, yields the differential inequality $u'(t) \le -(\eta/\tau)u(t) + C$ for $u(t) := \mathbb{E}\|e_t^\tau\|^2$. Since $e_0^\tau = 0$, Gronwall's lemma gives $u(t) \le C\tau/\eta$, proving (5); confinement of $\widetilde{Y}^\tau$ to $\mathcal{U}$ follows from the $O(\tau)$ bound itself. The general-initial-condition estimate uses the same inequality with $u(0) = \|y_0 - y_0^*\|^2$. □

Remark 3.9 (General initial conditions). If $\widetilde{Y}_0^\tau = y_0 \ne y^*(0, \widetilde{Z}_0, I_0)$, the same argument yields

$$\mathbb{E}[\|e_t^\tau\|^2] \le \|y_0 - y_0^*\|^2 e^{-2\eta t/\tau} + C_T\tau.$$

The first term is an initial-layer transient that decays on the fast relaxation scale τ; after that transient the $O(\tau)$ mean-square bound dominates.

Remark 3.10 (Scope of the dynamic justification). The strong-monotonicity condition (3) requires the symmetric part of $\partial_y F = I_{n+1} - B_t$ to be uniformly positive definite. In the two-dimensional benchmark of theorem 3.18, the symmetric part of $I - B$ has eigenvalues $1 \pm |\beta|/2$, so (3) fails when $|\beta| \ge 2$. This does not affect the paper's main results, which rely on theorem 3.2 (the static clearing hypothesis) rather than on its dynamic derivation: the conjugation, loss bound, identification, and risk tools all require only that the clearing fixed point exists and that $I_{n+1} - B_t$ is invertible. The dynamic regularization provides *one* sufficient mechanism for the clearing relation to arise; other mechanisms (e.g., tâtonnement with contractible impact, non-monotone dynamics with stability verified by other means) can produce clearing outside the strong-monotonicity regime.

With the clearing hypothesis established and justified, we derive its implications for covariation.

3.2 Covariation conjugation

Proposition 3.11 (Covariation conjugation under feedback)**.** *Assume theorems 2.1, 3.1, 3.2 and 3.6. Then the observed traded coordinates admit the local representation*

$$\widetilde{Y}_t = y^*(t, \widetilde{Z}_t, I_t),$$

and their local martingale part is

$$\mathrm{d}\widetilde{Y}_t^{(m)} = J_t\,\mathrm{d}\widetilde{Z}_t^{(m)}.$$

Consequently,

$$\mathrm{d}\langle\widetilde{Y}\rangle_t = J_t\,\mathrm{d}\langle\widetilde{Z}\rangle_t\,J_t^\top, \qquad \text{hence} \qquad c_t^Y = J_t\,c_t^Z\,J_t^\top \quad \text{for Lebesgue-a.e. } t \in [0, T].$$

Proof. By theorem 3.1, $\widetilde{Z}$ is a continuous $\mathbb{Q}$-local martingale, so $\mathrm{d}\widetilde{Z}_t = \mathrm{d}M_t$ with M a continuous local martingale. By theorem 3.6, along the realized path there is a local branch $\widetilde{Y}_t = y^*(t, \widetilde{Z}_t, I_t)$ with $\partial_z y^* = J_t$.

Because I is continuous and of finite variation, the multidimensional Itô formula applied locally to $y^*(t, \widetilde{Z}_t, I_t)$ yields

$$\mathrm{d}\widetilde{Y}_t = \partial_t y^* \, \mathrm{d}t + J_t \, \mathrm{d}\widetilde{Z}_t + \partial_i y^* \, \mathrm{d}I_t + \tfrac{1}{2} D^2_{zz} y^* : \mathrm{d}\langle \widetilde{Z} \rangle_t.$$

The $\mathrm{d}t$-, $\mathrm{d}I_t$-, and second-derivative terms are all finite variation. Therefore the continuous local martingale part of $\widetilde{Y}$ is

$$\mathrm{d}\widetilde{Y}^{(m)}_t = J_t \, \mathrm{d}\widetilde{Z}_t.$$

Since quadratic covariation depends only on continuous local martingale parts,

$$\mathrm{d}\langle \widetilde{Y} \rangle_t = J_t \, \mathrm{d}\langle \widetilde{Z} \rangle_t \, J_t^\top.$$

Using theorem 3.1 and theorem 2.1, both quadratic covariations are absolutely continuous with respect to Lebesgue measure, so $c^Y_t = J_t \, c^Z_t \, J_t^\top$ for Lebesgue-a.e. t. □

Remark 3.12 (Two distinct validity conditions). The framework involves two logically distinct steps with different validity requirements.

(A) *Instantaneous clearing* (the fixed point of theorem 3.2) is a fast-relaxation approximation: it holds when the market's price-adjustment timescale τ is small relative to the hedging horizon, with an $O(\tau)$ mean-square error bound uniformly on finite horizons under the diffusion assumptions of theorem 3.8. This is the structural hypothesis; its validity can be assessed empirically from the speed of lead–lag decay (theorem 4.1).

(B) *Covariation conjugation* ($c^Y = Jc^Z J^\top$, theorem 3.11) is *not* a further approximation. Once the fixed point holds and y^* is C^2 in z, the conjugation is an exact Itô consequence: quadratic covariation sees only the martingale part of $\mathrm{d}\widetilde{Y}_t = J_t \, \mathrm{d}M_t + (\text{finite variation})$; finite-variation terms drop out. The requirements are continuity of paths (theorem 2.1), absolute continuity of quadratic variation (theorem 3.1), C^2 regularity of the fixed-point map (theorem 3.6), and invertibility of $I_{n+1} - B_t$ along the realized path.

In particular, the loss bound, amplification index, and curvature amplification ($\widetilde{H} = J^\top H J$) are not "linearization approximations": they are exact consequences of the clearing relation once that structural hypothesis is imposed. The reduced-form and benchmark claims later in the paper require further assumptions and should not be read as consequences of Layer 1 alone. Near singularity ($\sigma_{\min}(I - B_t) \approx 0$), J_t blows up; the mapping remains the correct local object, but estimation becomes fragile and higher-order nonlinearities may matter.

3.3 Amplified curvature and two-dimensional benchmark

Combining the conjugation with the Layer 1 identity yields a representation that makes the role of feedback in hedging error explicit.

Corollary 3.13 (Amplified Hessian representation)**.** *Assume the hypotheses of theorems 2.7 and 3.11. Define*

$$H_t := D^2_{yy} v(t, \widetilde{Y}_t, \xi_t) \in \mathbb{S}^{n+1}, \qquad \widetilde{H}_t := J_t^\top H_t J_t.$$

Then

$$\widetilde{X}_T - v(T, \widetilde{Y}_T, \xi_T) = \widetilde{X}_0 - v(0, \widetilde{Y}_0, \xi_0) - \int_0^T \mathfrak{r}_t \, \mathrm{d}t - \frac{1}{2} \int_0^T \Big(H_t : (J_t c^Z_t J_t^\top - \widehat{c}(t, \widetilde{Y}_t, \xi_t)) \Big) \, \mathrm{d}t,$$

and the feedback term satisfies the trace-cycling identity

$$H_t : (J_t c^Z_t J_t^\top) = \widetilde{H}_t : c^Z_t.$$

Proof. Combine theorem 2.7 with $c_t^Y = J_t c_t^Z J_t^\top$ from theorem 3.11. Then $\mu_t = c_t^Y - \widehat{c}(t, \widetilde{Y}_t, \xi_t) = J_t c_t^Z J_t^\top - \widehat{c}(\cdot)$. The trace-cycling identity gives

$$H_t : (J_t c_t^Z J_t^\top) = \mathrm{Tr}(H_t J_t c_t^Z J_t^\top) = \mathrm{Tr}(J_t^\top H_t J_t c_t^Z) = \widetilde{H}_t : c_t^Z .$$

Substitute into the Layer 1 identity. □

This motivates the following definition.

Definition 3.14 (Amplification index)**.** Let $J_t = (I_{n+1} - B_t)^{-1}$ be the amplification matrix of theorem 3.11. The *amplification index* is

$$\mathsf{AI}_t \;:=\; \|J_t\|_2 \;=\; \frac{1}{\sigma_{\min}(I_{n+1} - B_t)},$$

where $\|\cdot\|_2$ is the spectral (operator-2) norm and $\sigma_{\min}$ the smallest singular value; we write $\|J\|$ for $\|J\|_2$ throughout.

Remark 3.15 (Singular values versus eigenvalues). No symmetry assumption is needed for the identity

$$\|J_t\|_2 = \|(I_{n+1} - B_t)^{-1}\|_2 = \frac{1}{\sigma_{\min}(I_{n+1} - B_t)}.$$

The amplification index is therefore well defined even when B_t is non-normal and has complex eigenstructure. It measures worst-case Euclidean amplification; other norms could be used for other risk geometries, but $\|\cdot\|_2$ is the natural choice for the matrix inequalities used in theorem 5.2 and the amplified-curvature bound.

The definition admits three mathematical readings.

Feedback resolvent. When $\|B_t\| < 1$ the Neumann series $J_t = \sum_{k=0}^{\infty} B_t^k$ converges absolutely, and $\mathsf{AI}_t = \|J_t\|$ measures the total gain of the infinite feedback loop: order 0 is the fundamental price, order 1 is direct feedback, order k is the k-th round of re-amplification. The Neumann series converges whenever $\rho(B_t) < 1$ and may diverge even when $I_{n+1} - B_t$ remains invertible (e.g. if B_t has eigenvalues of modulus ≥ 1 not equal to 1). The amplification index blows up as $\sigma_{\min}(I_{n+1} - B_t) \downarrow 0$, i.e. as an eigenvalue approaches 1, and diverges when $\det(I_{n+1} - B_t) = 0$. The operative definition $J_t = (I_{n+1} - B_t)^{-1}$ therefore extends beyond the contraction regime; the Neumann picture explains why amplification compounds through feedback rounds when contraction holds.

Congruence spectral bound. The conjugation $c^Y = J c^Z J^\top$ is a congruence transformation of the quadratic form c_t^Z. For any positive-semidefinite c_t^Z, the eigenvalue inflation satisfies

$$\frac{\lambda_{\max}(c_t^Y)}{\lambda_{\max}(c_t^Z)} \;\leq\; \|J_t\|_2^2 \;=\; \mathsf{AI}_t^2,$$

with equality when the leading eigenvector of c_t^Z aligns with the leading right singular vector of J_t. This is why AI enters the loss bound (theorem 5.2) at the *second* power: the quadratic form $\widetilde{H} : c^Z = \mathrm{Tr}(J^\top H J\, c^Z)$ inherits both singular-value factors.

Fold-bifurcation distance. $\sigma_{\min}(I_{n+1} - B_t)$ is the operator-norm distance from the Jacobian $\partial_y F = I_{n+1} - B_t$ of the clearing equation to the variety of singular matrices. As $\mathsf{AI}_t \to \infty$ this distance vanishes, the implicit-function parametrization becomes ill-conditioned and may fail; depending on higher-order nondegeneracy conditions, this can manifest as loss of branch uniqueness, loss of smoothness, or other local singular behavior, so small fundamental shocks can produce large or discontinuous changes in observed prices. The amplification index therefore doubles as a *robustness margin*: AI_t^{-1} is the smallest perturbation in B_t (measured in operator norm) that can drive the Jacobian to singularity and invalidate the local implicit-function description of clearing.

Remark 3.16 (Amplification bounds). The always-valid bound $\|\widetilde{H}_t\| \le \mathsf{Al}_t^2 \|H_t\|$ requires only that $I_{n+1} - B_t$ be nonsingular; it does not require the contraction condition $\|B_t\| < 1$. Under contraction, the Neumann-series bound $\|J_t\| \le (1 - \|B_t\|)^{-1}$ specializes this to $\|\widetilde{H}_t\| \le \|H_t\|/(1 - \|B_t\|)^2$. The $(1 - \|B_t\|)^{-2}$ rate is sharp in the scalar case but is generally conservative for non-normal matrices: equality need not hold, and the geometry of the singular vectors rather than eigenvalues alone determines how tight the bound is.

Remark 3.17 (Baseline sign benchmark). In a one-factor price-taking model with continuous paths and a deterministic numéraire, Itô's formula applied to $\widetilde{P}_t = p(t, \widetilde{S}_t)$ gives $\mathrm{d}\langle \widetilde{S}, \widetilde{P}\rangle_t = \partial_s p(t, \widetilde{S}_t)\, \mathrm{d}\langle \widetilde{S}\rangle_t$. When P is a put ($\partial_s p < 0$), spot-surface covariation is strictly negative. Any attenuation toward zero or sign reversal therefore requires either a second factor or an endogenous mechanism; the covariation conjugation of theorem 3.11 provides the latter. The following proposition makes this explicit.

Proposition 3.18 (Feedback-induced attenuation and sign reversal of observed correlation)**.** *Consider the two-channel traded system $(\widetilde{S}, \widetilde{U})$ with fundamental instantaneous covariance*

$$c_t^Z = \begin{pmatrix} \sigma_{S,t}^2 & \sigma_{SU,t} \\ \sigma_{SU,t} & \sigma_{U,t}^2 \end{pmatrix},$$

and assume that c_t^Z is positive definite, i.e.

$$\sigma_{S,t}^2 > 0, \qquad \sigma_{U,t}^2 > 0, \qquad \sigma_{S,t}^2 \sigma_{U,t}^2 - \sigma_{SU,t}^2 > 0.$$

Suppose the clearing feedback acts through the lower-triangular loop

$$B_t = \begin{pmatrix} 0 & 0 \\ \beta_t & 0 \end{pmatrix}, \qquad J_t = (I - B_t)^{-1} = \begin{pmatrix} 1 & 0 \\ \beta_t & 1 \end{pmatrix}.$$

Then the observed traded instantaneous covariance satisfies $c_t^Y = J_t c_t^Z J_t^\top$, and its off-diagonal entry is

$$(c_t^Y)_{SU} = \sigma_{SU,t} + \beta_t \sigma_{S,t}^2. \tag{6}$$

Assume $\sigma_{SU,t} < 0$. Define the critical feedback level

$$\beta_t^* := -\frac{\sigma_{SU,t}}{\sigma_{S,t}^2} > 0. \tag{7}$$

Then:

(i) if $0 < \beta_t < \beta_t^*$, then $\sigma_{SU,t} < (c_t^Y)_{SU} < 0$ (attenuation)*;*

(ii) if $\beta_t = \beta_t^*$, then $(c_t^Y)_{SU} = 0$ (neutralisation)*;*

(iii) if $\beta_t > \beta_t^*$, then $(c_t^Y)_{SU} > 0$ (sign reversal)*.*

Because $c_t^Y = J_t c_t^Z J_t^\top$ with J_t invertible and c_t^Z positive definite, the matrix c_t^Y is positive definite. Hence $(c_t^Y)_{SS} > 0$ and $(c_t^Y)_{UU} > 0$, the observed instantaneous correlation $\rho_{SU,t}^Y$ is well defined, and its sign equals the sign of $(c_t^Y)_{SU}$. Therefore the same trichotomy holds for correlation.

Proof. By theorem 3.11, observed traded covariation is the congruence transform $c_t^Y = J_t c_t^Z J_t^\top$. A direct multiplication gives

$$c_t^Y = \begin{pmatrix} \sigma_{S,t}^2 & \sigma_{SU,t} + \beta_t \sigma_{S,t}^2 \\ \sigma_{SU,t} + \beta_t \sigma_{S,t}^2 & \sigma_{U,t}^2 + 2\beta_t \sigma_{SU,t} + \beta_t^2 \sigma_{S,t}^2 \end{pmatrix},$$

proving (6). Now assume $\sigma_{SU,t} < 0$ and define β_t^* by (7). Then $(c_t^Y)_{SU} = \sigma_{S,t}^2(\beta_t - \beta_t^*)$, so the sign regimes follow. Since $\beta_t \sigma_{S,t}^2 > 0$, we have $(c_t^Y)_{SU} > \sigma_{SU,t}$ (strictly less negative), confirming attenuation in regime (i). Finally, because c_t^Y is obtained from the positive-definite c_t^Z by congruence with the invertible J_t, the matrix c_t^Y is positive definite. Therefore the denominator of $\rho_{SU,t}^Y$ is strictly positive and the sign of the correlation equals the sign of $(c_t^Y)_{SU}$. □

Corollary 3.19 (Correlation attenuation formula)**.** *Under the conditions of theorem 3.18 with constant coefficients, define the normalised feedback* $\gamma := \beta/\beta^* \in [0, \infty)$ *and the latent correlation* $\rho^Z := \sigma_{SU}/(\sigma_S \sigma_U) \in (-1, 1)$ *(strict inequalities hold because* c^Z *is positive definite). Then the observed instantaneous correlation satisfies*

$$\rho^Y_{SU}(\gamma) \;=\; \frac{\rho^Z\,(1-\gamma)}{\sqrt{(1-(\rho^Z)^2)\;+\;(\rho^Z)^2(1-\gamma)^2}}\,.$$

In particular:

(i) $\rho^Y_{SU}(0) = \rho^Z$ (no feedback recovers the latent correlation)*;*

(ii) $\rho^Y_{SU}(1) = 0$ (neutralisation, same threshold as covariation)*;*

(iii) $\rho^Y_{SU}(\gamma) \to -\operatorname{sgn}(\rho^Z)$ *as* $\gamma \to \infty$ (saturation at ± 1 under extreme feedback)*;*

(iv) $|\rho^Y_{SU}(\gamma)| < 1$ *for all finite* γ (Sylvester's law: positive definiteness is preserved).

Proof. From theorem 3.18, $c^Y_{SS} = \sigma_S^2$, $c^Y_{SU} = \sigma_{SU}(1-\gamma)$, and $c^Y_{UU} = \sigma_U^2[(1-(\rho^Z)^2)+(\rho^Z)^2(1-\gamma)^2]$. The formula follows from $\rho^Y_{SU} = c^Y_{SU}/\sqrt{c^Y_{SS}\,c^Y_{UU}}$. For (iii): $\rho^Y_{SU} \sim \rho^Z(-\gamma)/(|\rho^Z|\gamma) = -\operatorname{sgn}(\rho^Z)$. For (iv): $|\rho^Y_{SU}|^2 = (\rho^Z)^2(1-\gamma)^2/[(1-(\rho^Z)^2) + (\rho^Z)^2(1-\gamma)^2] < 1$ since $1-(\rho^Z)^2 > 0$. □

Remark 3.20 (Economic mechanism and hedging-loss connection). The observed covariation $(c^Y_t)_{SU} = \sigma_{SU,t} + \beta_t \sigma^2_{S,t}$ is a competition between two forces: the *fundamental negative* comovement $\sigma_{SU,t}$ (spot down ⇒ surface up) and the *feedback-induced positive* comovement $\beta_t \sigma^2_{S,t}$ (spot variance injected into the surface channel by hedging flow). Attenuation is when the second partially offsets the first; sign reversal is when it overwhelms it. The mechanism is structural: because Layer 1 pairs the desk Hessian with the covariation distortion $\mu_t = c^Y_t - \widehat{c}_t$, attenuation or reversal of $(c^Y_t)_{SU}$ directly enters the hedging P&L identity (theorem 2.7). When the desk prices off a covariance benchmark $\widehat{c}_t$ that does not incorporate feedback mixing, the sign flip creates hedging losses even when the local pricing model is otherwise internally consistent.

fig. 2 illustrates the three regimes for several values of σ_{SU}; the correlation-attenuation diagnostic cleanly separates feedback from no-feedback cases.

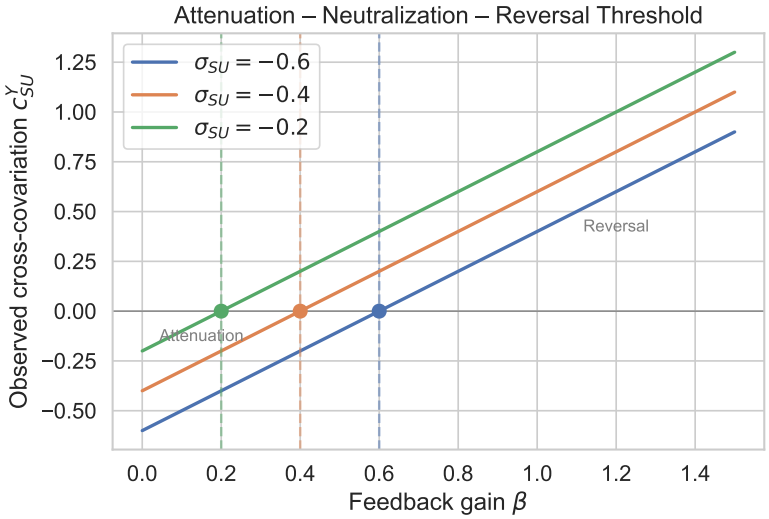

Figure 2: Attenuation, neutralization, and reversal of c^Y_{SU} vs. feedback gain β (theorem 3.18).

3.3.1 Two channels as the fundamental building block

The two-dimensional benchmark of theorem 3.18 is the *minimal* case that exhibits cross-channel feedback (off-diagonal entries of B_t). In the scalar case ($n+1=1$), the conjugation reduces to multiplication by $j^2=(1-b)^{-2}$: no cross-channel distortion, no lead–lag asymmetry, no vanna sign reversal. With two coordinates, $c^Z \mapsto Jc^ZJ^\top$ mixes rows and columns of c^Z through J while preserving the bilinear-form signature (Sylvester's law), creating directional distortion that scalar measures cannot capture: this is the mechanism behind the $\|J\|$-dependence of the loss bound and the Lyapunov instability from time-varying off-diagonal coupling. This is the *algebraic leap*: from scalar amplification to matrix-valued covariation distortion (fig. 4a → b).

Passing from two channels to many is a matter of composition: the Neumann series $J = I + B + B^2 + \cdots$ makes explicit that B^k captures k-step feedback chains of pairwise interactions, and all higher-dimensional effects are composed from the same two-channel algebra. The following proposition formalises this in dimension $n+1$.

Proposition 3.21 (Cross-channel mixing, attenuation, and sign reversal in dimension $n+1$)**.** *Let the observed traded instantaneous covariance be $c_t^Y = J_t\, c_t^Z\, J_t^\top$, $J_t=(I-B_t)^{-1}$, where $c_t^Z \in \mathbb{S}_+^{n+1}$ is fundamental and $J_t \in \mathbb{R}^{(n+1)\times(n+1)}$ is invertible. Fix two channels $i \neq j$. Then*

$$(c_t^Y)_{ij} = \sum_{k,\ell=1}^{n+1} (J_t)_{ik}\,(c_t^Z)_{k\ell}\,(J_t)_{j\ell}. \tag{8}$$

Separating the direct term gives

$$(c_t^Y)_{ij} = \underbrace{(J_t)_{ii}(J_t)_{jj}(c_t^Z)_{ij}}_{\text{direct}} + \underbrace{R_{ij,t}}_{\text{cross-channel mixing}}, \tag{9}$$

where $R_{ij,t} := \sum_{(k,\ell)\neq(i,j)} (J_t)_{ik}\,(c_t^Z)_{k\ell}\,(J_t)_{j\ell}$. Hence:

- *(i)* attenuation *occurs when $(c_t^Z)_{ij}$ and $(c_t^Y)_{ij}$ share sign but $|(c_t^Y)_{ij}| < |(J_t)_{ii}(J_t)_{jj}(c_t^Z)_{ij}|$;*
- *(ii)* neutralisation *occurs when $R_{ij,t} = -(J_t)_{ii}(J_t)_{jj}(c_t^Z)_{ij}$;*
- *(iii)* sign reversal *occurs when $R_{ij,t}$ and $(J_t)_{ii}(J_t)_{jj}(c_t^Z)_{ij}$ have opposite signs and the magnitude of the former exceeds that of the latter.*

If $(c_t^Y)_{ii} > 0$ and $(c_t^Y)_{jj} > 0$, the sign of $\rho_{ij,t}^Y$ equals the sign of $(c_t^Y)_{ij}$, so the same trichotomy holds for correlation.

Proof. Identity (8) is the (i,j) entry of $c_t^Y = J_t c_t^Z J_t^\top$. Separating the $(k,\ell)=(i,j)$ term gives (9). The sign and magnitude of $(c_t^Y)_{ij}$ are then determined by the competition between the direct fundamental contribution and the cross-channel remainder $R_{ij,t}$; the three regimes follow. □

In dimension two, sign reversal arises because spot variance is injected into the surface channel through a single feedback coefficient (theorem 3.18). In higher dimension, the mechanism is the same but richer: the remainder $R_{ij,t}$ collects all indirect paths by which shocks originating in other channels are transported into the (i,j) covariance slot. Attenuation occurs when those indirect contributions partially offset the direct fundamental covariance; sign reversal occurs when they dominate it. Hence apparently anomalous observed correlations need not reflect anomalous fundamentals: they can be generated endogenously by cross-channel amplification in the clearing map. Because Layer 1 pairs the desk Hessian with $\mu_t = c_t^Y - \widehat{c}_t$, theorem 3.21 shows that cross-channel mixing is a structural source of hedging error whenever the desk's covariance benchmark does not incorporate feedback.

What higher dimensions add is a second, distinct leap: *network topology* (fig. 4c to d). The Neumann series endows the resolvent with the structure of a *path algebra* on the directed

feedback graph G: entry $(B^k)_{ij}$ sums directed walks of length k from j to i, and $\mathrm{tr}(B^k)$ sums weights of length-k *closed walks* (in sparse graphs these are often dominated by simple k-cycles). Newton's identities connect these cycle counts $\{p_k = \mathrm{tr}(B^k)\}$ to the eigenvalues of B, so the *cycle topology of G governs spectral stability.* Cycles require at least one reciprocal pair of edges. In the two-dimensional benchmark of theorem 3.18 we deliberately take B strictly lower-triangular (spot→put only), so $B^2 = 0$ and no cycle exists. In general, however, a bilateral 2-cycle $i \leftrightarrow j$ already lives in a 2×2 subsystem with both off-diagonal entries nonzero, and is detected by $\mathrm{tr}(B^2) \neq 0$ (equivalently $(B^2)_{ii} \neq 0$ and $(B^2)_{jj} \neq 0$ in the corresponding principal submatrix; the sign depends on edge weights).

Three topological increments. The two-dimensional benchmark has three independent topological extensions, forming a nested sequence of increasing cycle richness.

Increment A (baseline): spot exogenous, no cycle. The first row of B is zero ($b_{0j} = 0$ for all j): hedging flow distorts vanilla prices but does not feed back into spot. $B^2 = 0$ and $J = I + B$; feedback does not compound across rounds. This is the two-coordinate benchmark of theorem 3.18, which already captures the full algebra of pairwise covariation distortion (attenuation, neutralization, reversal).

Increment B: spot endogenized ($b_{01} \neq 0$), first cycle in 2D. Allowing the spot coordinate to be impacted (for instance, because vanilla dealers delta-hedge the puts they absorb from the exotic desk, creating spot flow proportional to $N_{\text{put}}\,\delta_{\text{van}}$) adds the return edge $P \to S$. The loop-gain matrix becomes $B = \left(\begin{smallmatrix} 0 & b_{01} \\ b_{10} & 0 \end{smallmatrix}\right)$ with $B^2 = b_{01}b_{10}\,I_2 \neq 0$: a *spot-vanilla 2-cycle* $S \to P \to S$ in which hedging flow and counterparty delta-hedging compound across rounds. Economically, this is the "vicious circle" of barrier crises: spot falls toward the barrier, the exotic desk's vega hedge creates vanilla flow, the counterparty's delta hedge pushes spot further toward the barrier, amplifying the next round. The baseline sets $b_{01} = 0$; a parameter sweep over $b_{01} \in \{0, 0.05, \ldots, 0.2\}$ (fig. 3) is consistent with the prediction that the $S \leftrightarrow P$ cycle raises mean AI^* monotonically (1.08 → 1.20), with $\sigma_{\min}(I - B_t)$ (the 5th-percentile per-path minimum) decreasing from 0.030 to 0.023: the fixed point remains nonsingular but moves closer to singularity. The incremental AUC of $\widehat{\mathsf{AI}}$ *decreases* with b_{01} (+0.005 → 0; block-bootstrap 95% CI crosses zero at $b_{01} \approx 0.15$), but the mechanism is *redundancy*, not misspecification or estimator failure. Empirically, $\mathrm{corr}(\widehat{\mathsf{AI}}, \mathsf{AI}^*)$ stays near 0.11 for all $b_{01} > 0$ (fig. 3B), while the baseline predictor improves: baseline AUC rises from 0.756 to 0.798 (fig. 3C). The $S \leftrightarrow P$ cycle therefore makes barrier proximity a stronger proxy for tail losses, absorbing the feedback signal that $\widehat{\mathsf{AI}}$ would otherwise add. This delineates when $\widehat{\mathsf{AI}}$ adds value: it is most informative when the feedback topology is richer than what a scalar proximity variable can encode (Increment C), not when the dominant mechanism is a two-node $S \leftrightarrow P$ cycle (Increment B).

Increment C: cross-strike cycle ($\rho_{\text{surface}} > 0$), 3D. Adding a second vanilla coordinate P_2 (with or without Increment B) introduces a cycle *inside the vanilla block*: $P_1 \to P_2 \to P_1$, with B^2 cross-term $(B^2)_{12} = 2\lambda^2 c^2 \rho_{\text{surface}}$, provably absent in any single-put model (section 6.1). This cross-strike cascade is a distinct mechanism from the spot-vanilla cycle of Increment B: it operates even when spot is exogenous ($b_{0j} = 0$) and captures how hedging flow at one strike propagates to another through the shared impact surface.

Richer 3-cycles $i \to j \to k \to i$ require at least three feedback-active nodes and are detected by $\mathrm{tr}(B^3) \neq 0$. Hub structures (one asset's flow distorting many others) create correlations between otherwise independent pairs. These topological features affect the spectral properties of J (and hence Lyapunov exponents and stability boundaries) in ways that cannot be predicted from any single 2×2 block of B. However, the *algebraic machinery* is unchanged: the conjugation $c^Y = Jc^Z J^\top$, the amplified Hessian $\widetilde{H} = J^\top H J$, and all downstream results hold in any dimension without modification.

There is no fourth qualitative leap: covariation is bilinear (no higher-order brackets exist in continuous semimartingales), so congruence is the complete algebraic mechanism, and once the feedback graph contains a reciprocal pair (a 2-cycle, arising in Increment B from spot-vanilla

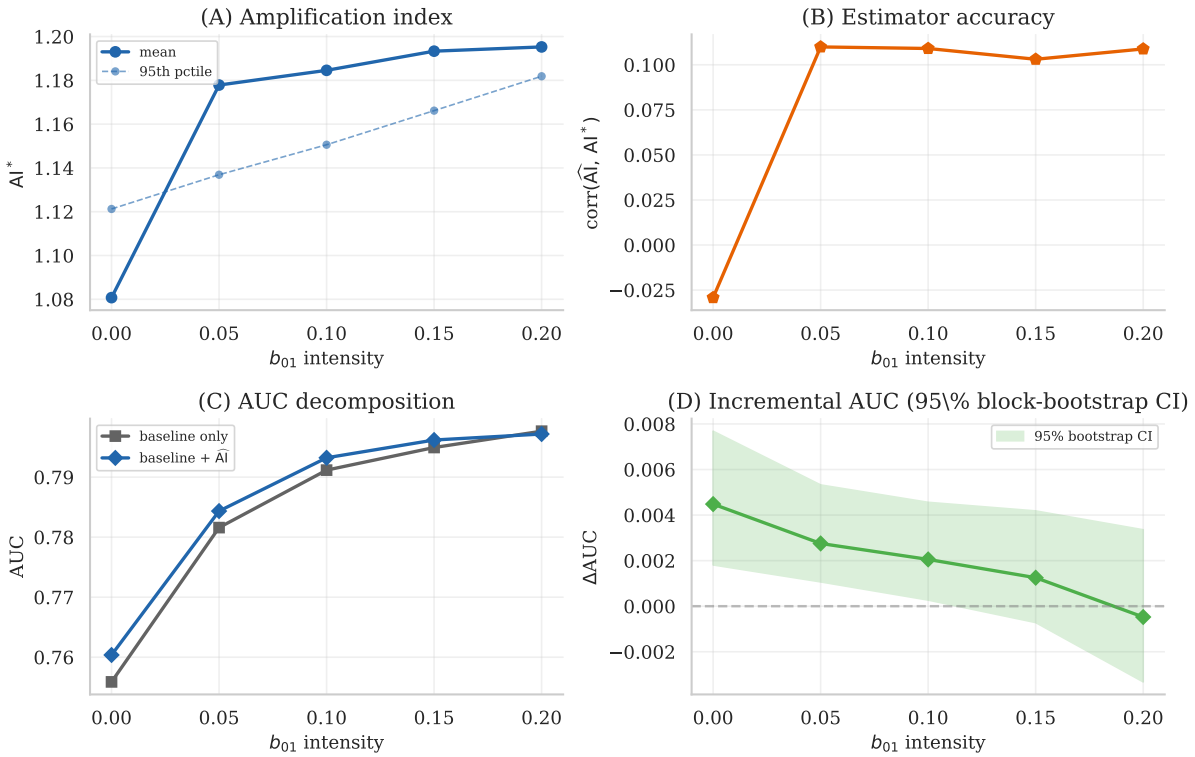

Figure 3: Increment B parameter sweep: spot endogenization via vanilla delta-hedging. The desk estimator assumes $b_{01} = 0$ throughout. (A) AI^* quantiles increase monotonically with b_{01}. (B) $\mathrm{corr}(\widehat{\mathsf{AI}}, \mathsf{AI}^*)$: estimator accuracy is roughly constant. (C) AUC decomposition: baseline AUC rises (barrier proximity absorbs the signal), combined AUC stays flat. (D) Incremental AUC with 95% block-bootstrap CI; the decline reflects redundancy, not estimator failure.

interaction or in Increment C from cross-strike coupling), adding coordinates enriches the topology but introduces no new algebraic structure. The two-channel case thus contains the complete algebra of pairwise covariation distortion (attenuation, neutralization, reversal, amplification, resonance), from which all higher-dimensional effects are composed (fig. 4a to b). Extending further to a (K, T) grid of vanillas (fig. 4d) would introduce richer cycle structures ($\mathrm{tr}(B^3) \neq 0$) and hub topologies, while requiring parsimonious coupling parameterizations (nearest-neighbor, exponential-decay, or factor structures across strikes and maturities).

Figure 4 summarizes the nested increments discussed in sections 3.3.1 and 6.1. Panel (a) (scalar) admits only amplification. Panel (b) (two channels) is the algebraic leap: off-diagonal loop gain distorts cross-covariation. Panel (c) (spot + two vanillas) is the minimal topological increment in our spot-exogenous sequence: cross-strike coupling creates a reciprocal pair and a B^2 feedback cycle $P_1 \to P_2 \to P_1$. Panel (d) sketches a (K, T) grid, where richer closed-walk/cycle structure (e.g. $\mathrm{tr}(B^3) \neq 0$) and hubs appear; the algebra $c^Y = Jc^Z J^\top$ with $J = (I - B)^{-1}$ remains unchanged.

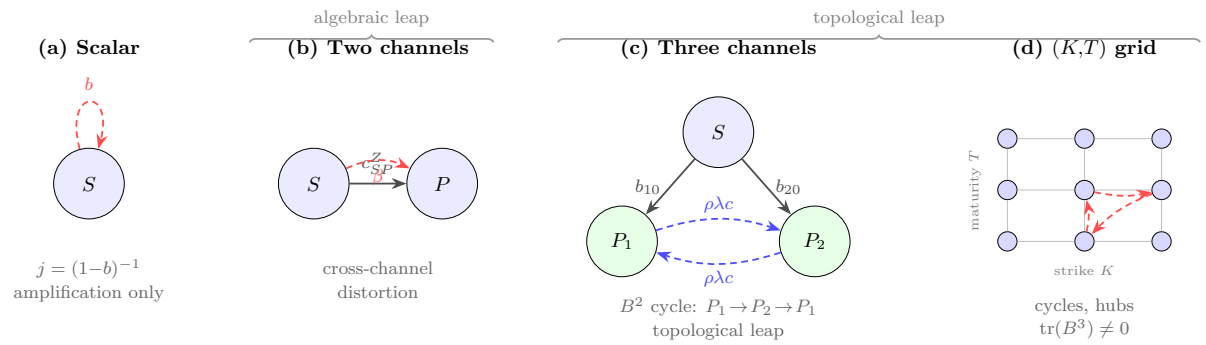

Figure 4: Algebraic vs. topological leaps as the number of traded coordinates grows.

3.4 Barrier products and the uridashi mechanism

The two-dimensional benchmark establishes the algebra of feedback-induced covariation distortion; we apply it next to a concrete product class that motivated the framework.

The following clarification of the simulation architecture is essential. In both simulation specifications below, the pricing function v in theorem 2.7 is the *vanilla Black–Scholes put*: the Hessian $H_t = D^2_{yy}v$ is the put's gamma, vanna, and volga, and the hedge $\vartheta_t = \nabla_y v$ is the put's delta and vega. The exotic product (autocallable or DIUO) does *not* appear as v; it enters only through the *flow map* Φ that determines aggregate hedging demand and hence the feedback coefficient. This separation mirrors actual desk practice: the exotic sits on the book, but hedging and P&L accounting are done in vanilla instruments whose prices are observable and whose sensitivities have closed-form expressions.

We use two simulation specifications. The baseline autocallable specification (section B.2) serves as a control with a smooth, essentially one-dimensional proximity-driven flow map. The uridashi-motivated specification is a barrier product: down-and-in, up-and-out, double-knockout, and their combinations. The *product* on the desk's book is the exotic, but *hedging* is done through liquid vanilla instruments (spot for delta, vanilla puts/calls for vega), so aggregate hedging demand flows through vanilla markets and distorts vanilla prices.

We therefore focus on a down-in up-out (DIUO) put with lower knock-in barrier B_- and upper knock-out barrier B_+, priced analytically via

$$V_{\mathrm{DIUO}}(S) \;=\; P_{\mathrm{UO}}(S;\, B_+) \;-\; P_{\mathrm{DKO}}(S;\, B_-, B_+),$$

where P_{UO} is the Rubinstein–Reiner [29] up-and-out put (method of images) and P_{DKO} is the double-knockout put via the Fourier sine spectral method:

$$P_{\mathrm{DKO}}(S) \;=\; S^{\alpha} \sum_{n=1}^{N} \bar{c}_n \, \sin\!\Big(\frac{n\pi \log(S/B_-)}{L}\Big)\, e^{(\beta - n^2\pi^2/L^2)\,\tau},$$

with $\tau = T - t$, $L = \log(B_+/B_-)$, $\alpha = \frac{1}{2} - r/\sigma^2$, $\beta = -\frac{1}{2}(r/\sigma - \sigma/2)^2$ (the eigenvalue constant from the drift-removal transform of the Black–Scholes PDE in log-coordinates), and $\bar{c}_n$ computed in closed form from the put payoff.

The desk hedges the DIUO using two instruments:

1. *Spot* (for delta): with net delta $\delta_{\mathrm{net}} = \delta_{\mathrm{DIUO}} - N_{\mathrm{put}}\,\delta_{\mathrm{van}}$, where N_{put} is the vega hedge ratio below.
2. *Vanilla puts* (for vega): the hedge ratio $N_{\mathrm{put}} = \nu_{\mathrm{DIUO}}/\nu_{\mathrm{van}}$ is the number of vanilla puts bought or sold to neutralize the DIUO's vega exposure.

The exotic's mark-to-market return decomposes as $\delta_{\mathrm{net}}\,\mathrm{d}S + N_{\mathrm{put}}\,\mathrm{d}P$, where $\mathrm{d}P$ is the change in the vanilla put price *inclusive of feedback distortion*. Note that, under a common implied-volatility coordinate σ (holding (S, t, r) fixed), the chain rule gives $\partial V_{\mathrm{DIUO}}/\partial P_{\mathrm{van}} = \nu_{\mathrm{DIUO}}/\nu_{\mathrm{van}} = N_{\mathrm{put}}$.

To neutralize this $+N_{\text{put}}\,\mathrm{d}P$ exposure, the hedge takes the opposite vanilla-put position: $-N_{\text{put}}$ puts per unit of exotic (a short position whenever $N_{\text{put}} > 0$), yielding a per-unit hedge P&L of $-N_{\text{put}}\,\mathrm{d}P$ from the vanilla leg.

The single-put coordinate P captures the dominant level factor of the implied-volatility surface, which accounts for 60–70% of surface variance; the multi-strike extension in section 6.1 adds the skew dimension in the $n{+}1$-dimensional framework of section 3.3.1 and quantifies the additional cross-strike propagation.

This is where the feedback enters: the vanilla put price P is the traded coordinate, and the clearing relation (theorem 3.2) applied to the put coordinate gives

$$\widetilde{P}_t = P_t^Z + \Lambda(Q_t), \qquad Q_t = \Phi(t, \widetilde{P}_t, I_t).$$

For a desk that is long n_t units of the exotic (the typical issuer position in uridashi notes), the aggregate short-put hedge generates a signed demand of

$$Q_t \;=\; -\,n_t\,N_{\text{put}} \;<\; 0 \qquad \text{when } N_{\text{put}} > 0,$$

so DIUO hedging creates sell pressure that depresses the vanilla put price below its fundamental value.

The DIUO's atypical Greeks (delta reversal near $\mathcal{B}_+$, negative gamma near barriers, vanna sign reversal; see fig. 5) drive large hedge ratios near barriers. In the linearized model, the feedback coefficient is $c = -\phi_{\text{pre}} \cdot N_{\text{put}}$, where $\phi_{\text{pre}} > 0$ scales with position size and market impact; because N_{put} inherits the nonlinear Greek structure of the DIUO, the amplification index $\mathsf{AI} = \|J\|$ captures signal that a monotone proximity function cannot replicate.

The barrier levels $B_- = 65\%$ and $B_+ = 110\%$ match the structure of real uridashi products [1]: the deep lower barrier produces massive hedge ratios ($N_{\text{put}} > 10$) when spot approaches B_-, while the tight upper barrier creates frequent knock-outs that force rapid position unwinds. Figure 6 shows three representative scenarios:

- *Spot falls toward* B_-: the hedge ratio N_{put} spikes as DIUO vega peaks, so the desk builds a large short-put position. Aggregate put-selling pressure depresses vanilla put prices. The put-leg P&L diverges from the spot leg as feedback distortion accumulates.
- *Spot rises toward* B_+: DIUO vega collapses and $N_{\text{put}} \to 0$. The desk must buy back the puts it previously sold, pushing prices higher in an illiquid environment. The *unwind* generates a second round of feedback losses.
- *Knock-in then reversal*: the most painful combination: a full build-up of the short-put hedge during decline, followed by a forced unwind as spot rallies toward B_+. This matches the 2012 uridashi episode described in [1]: "the dynamics of a double barrier cannot be hedged 100% through vanillas."

Running the full 9-step externally validated procedure with the DIUO product yields an incremental AUC of $\widehat{\mathsf{AI}}$ over the best single baseline of $+0.004$: in this simulation, the amplification index captures information (vanna sign reversals, delta flips, nonlinear vega-ratio dynamics) that no scalar proximity measure can replicate.

Remark 3.22 (Autocallable as control experiment). As a control, replacing the DIUO with an autocallable (whose flow map is a smooth, essentially monotone function of spot-to-barrier distance) yields negative incremental AUC (-0.005): barrier distance alone subsumes the amplification signal. This delineates *when* $\widehat{\mathsf{AI}}$ adds value: its marginal contribution scales with the structural complexity of the product-to-hedge separation. When the feedback coefficient reduces to a single scalar (proximity), simpler predictors suffice; the DIUO specification captures the structurally richer case that arises in practice.

Figure 5: (A) price; (B) delta reversal near B_+; (C) negative gamma near both barriers; (D) vanna sign reversal. DIUO is everywhere below vanilla and vanishes at both barriers.

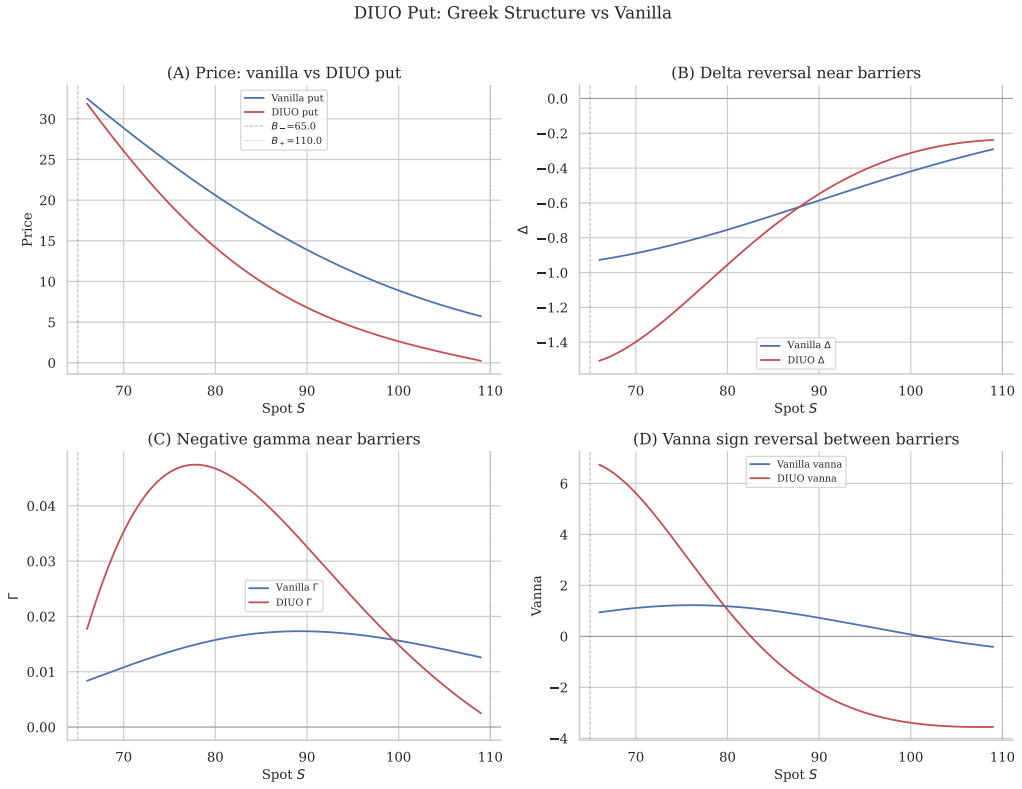

Figure 5: DIUO vs. vanilla put Greeks (representative parameters).

Figure 6: left, spot path with KI/KO markers; right, cumulative P&L by leg (spot green, put purple). Row 1: approach to B_-, N_{put} spike. Row 2: approach to B_+, knock-out and forced unwind. Row 3: knock-in then reversal ("vicious circle").

The conjugation identifies *what* feedback does to covariation; the next section addresses *how* to detect and measure it from observable data.

3.4.1 Scope map: pointwise, local, and global statements

The framework combines three types of statements with different quantifiers.[2] The most important distinction is between *exact* results (Layer 1 identity on $[0, T]$; covariation conjugation $c^Y = Jc^Z J^\top$ given the clearing fixed point) and *approximate* ones (the OU-window lead–lag recovery; the instantaneous clearing itself as a fast-relaxation limit with $O(\tau)$ error). Consequences that use J_t (loss bounds, amplified Hessian, identification formulas) are local-in-state statements along the realized trajectory under theorem 3.2, not global existence results; identification tools additionally require local stationarity or conditional-moment assumptions that are separate from the exact accounting identities.

[2]To apply Itô's formula to $y^*(t, \widetilde{Z}_t, I_t)$, it suffices that $z \mapsto y^*(t, z, i)$ is C^2 for each (t, i), which theorem 3.6 guarantees. Since $\widetilde{Z}$ is a $\mathbb{Q}$-local martingale (theorem 3.1), the continuous local martingale part of $\widetilde{Y}$ is $\mathrm{d}\widetilde{Y}_t^{(m)} = J_t\, \mathrm{d}\widetilde{Z}_t$; the remaining terms ($\partial_t y^*$, $\partial_i y^*\, \mathrm{d}I$, and the second-order correction) are finite variation and do not contribute to $\langle \widetilde{Y} \rangle$.

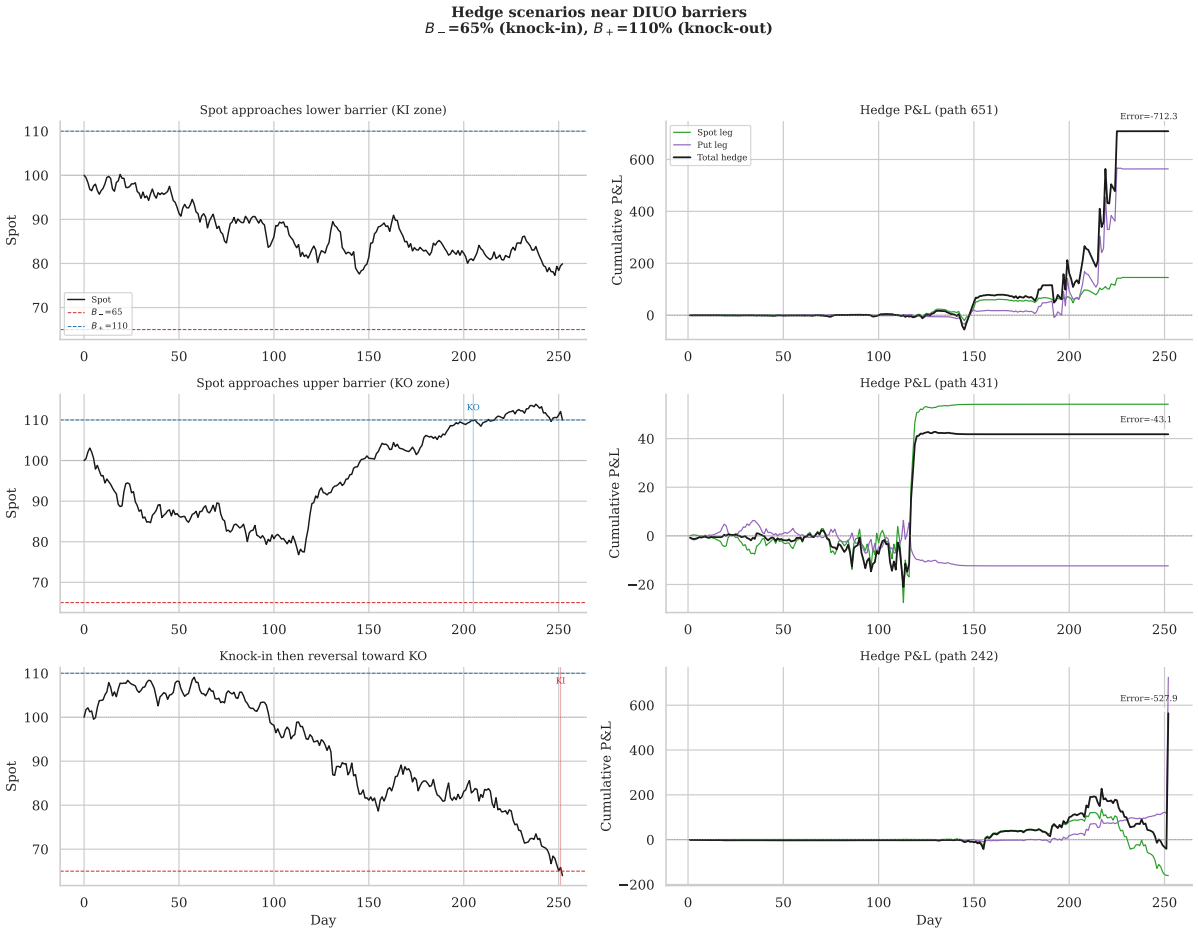

Figure 6: Hedge P&L decomposition for three barrier scenarios.

4 Identification and Estimation from Observables

The preceding section establishes the conjugation $c^Y = Jc^Z J^\top$ and its consequences for hedging error, barrier Greeks, and covariation distortion. A natural question follows: given observed data, how can one identify the feedback channels (their direction, magnitude, and stability) from market observables alone? This section develops two complementary identification routes, plus a built-in specification diagnostic:

(i) **Lead–lag identification** (OU window): recovers a local generator $\mathcal{A}$ and directional channel structure from lagged covariances (section 4.1). This route is exact under the OU approximation and degrades gracefully as a diagnostic outside it.

(ii) **Moment identification** (covariance map + signatures): uses conditional covariance moments to recover individual gains $b_{m,t}$ and provides over-identifying restrictions as a misspecification check (section 4.3).

Both routes are formulated in terms of observables; their modelling assumptions are separate from the exact identities of Layers 1 and 2 (cf. section 3.4.1).

4.1 Lead–lag identification

Contemporaneous covariation identifies only the symmetric stretch $JJ^\top$, not the full (generally non-symmetric) amplification matrix J. Lead–lag covariances resolve this.

Consider a locally stationary linearization of (1) on a window $[t_0, t_1]$:

$$\mathrm{d}X_t = \mathcal{A}\, X_t\, \mathrm{d}t + \Gamma\, \mathrm{d}W_t, \qquad \mathcal{A} = \tfrac{1}{\tau}(B - I_{n+1}), \tag{10}$$

with $\mathcal{A}$ Hurwitz ($\max_i \Re(\lambda_i(\mathcal{A})) < 0$), stationary covariance Σ, and lead–lag covariance $C(s) := \mathbb{E}[X_{t+s}X_t^\top]$ for $s \geq 0$.

Proposition 4.1 (Lead–lag identification of the local generator on the principal branch)**.** *Under* (10) *with* $\mathcal{A}$ *Hurwitz and* Σ *invertible,*

$$C(s) = e^{\mathcal{A}s}\,\Sigma, \qquad s \geq 0. \tag{11}$$

Assume, moreover, that $e^{\mathcal{A}s}$ *has spectrum avoiding* $\mathbb{R}_{\leq 0}$*, so that the principal matrix logarithm is well defined. If* $\mathcal{A}$ *belongs to the principal-log branch at lag* s *(equivalently, all eigenvalues of* $\mathcal{A}$ *have imaginary parts in* $(-\pi/s, \pi/s)$*), then for any* $s > 0$*,*

$$\mathcal{A} = \frac{1}{s}\,\mathrm{Log}\big(C(s)\Sigma^{-1}\big), \tag{12}$$

where Log *denotes the principal matrix logarithm. The result is exact within the local OU benchmark and on the chosen logarithm branch; applied beyond that benchmark, it provides a local reduced-form identification of directional feedback channels from observable lead–lag covariances.*[3]

Proof. The OU solution gives $X_{t+s} = e^{\mathcal{A}s}X_t + \int_0^s e^{\mathcal{A}(s-u)}\Gamma\, \mathrm{d}W_{t+u}$. Independence of increments yields $C(s) = e^{\mathcal{A}s}\,\Sigma$. Hence $e^{\mathcal{A}s} = C(s)\Sigma^{-1}$. Under the stated branch condition, the principal logarithm inverts the matrix exponential on this spectral strip, yielding (12). □

Corollary 4.2 (Directional structure from the identified generator)**.** *If* $\mathcal{A}$ *is identified via* (12)*, then the loop-gain matrix satisfies*

$$B = I_{n+1} + \tau\mathcal{A}.$$

[3] Without the principal-branch condition, a single lag identifies only $e^{\mathcal{A}s} = C(s)\Sigma^{-1}$, not $\mathcal{A}$ uniquely; this is the continuous-time analogue of frequency aliasing. In practice one resolves the ambiguity by checking consistency across multiple lags or imposing stability restrictions on the spectrum.

The directional channel structure is encoded in the off-diagonal entries, sparsity pattern, and impulse-response asymmetry of $\mathcal{A}$, and the stability margin

$$m := -\max_i \Re(\lambda_i(\mathcal{A}))$$

is likewise τ-free once $\mathcal{A}$ is identified from lead–lag covariances. By contrast, the eigenvalue magnitudes of B depend on the unknown relaxation scale τ through

$$\lambda_i(B) = 1 + \tau\lambda_i(\mathcal{A}).$$

The eigenvectors of $\mathcal{A}$ characterize the dominant amplification/stability modes, not causality by themselves.

Proof. The identity $B = I_{n+1} + \tau\mathcal{A}$ follows from (10). Once $\mathcal{A}$ is identified, its off-diagonal structure and impulse responses are directly observable, as is the spectral margin m. The final statement follows because B is an affine function of $\mathcal{A}$. □

Remark 4.3 (Early warning via generator spectrum). The amplification index $\mathsf{AI}_t = \|J_t\|$ measures distance to singularity of the static clearing map, but dynamic stability is governed by the generator margin $m := -\max_i \Re(\lambda_i(\mathcal{A}))$: as $m \downarrow 0$, stationary covariance diverges as $O(m^{-1})$, providing an early-warning criterion for covariance blow-up. On the synthetic OU benchmark, the lead–lag asymmetry reaches $|C_{21}(s)/C_{12}(s)| = 11{:}1$ at $B_{21} = 0.6$, and the matrix-logarithm recovery attains element-wise errors below 10^{-13}; the exact identities of sections 2 and 3 remain valid across all local-stationarity regimes; the OU diagnostic degrades gracefully as a local approximation when stationarity fails, and the lead–lag asymmetry still provides directional channel information. The lead–lag asymmetry and generator recovery are illustrated in fig. 7.

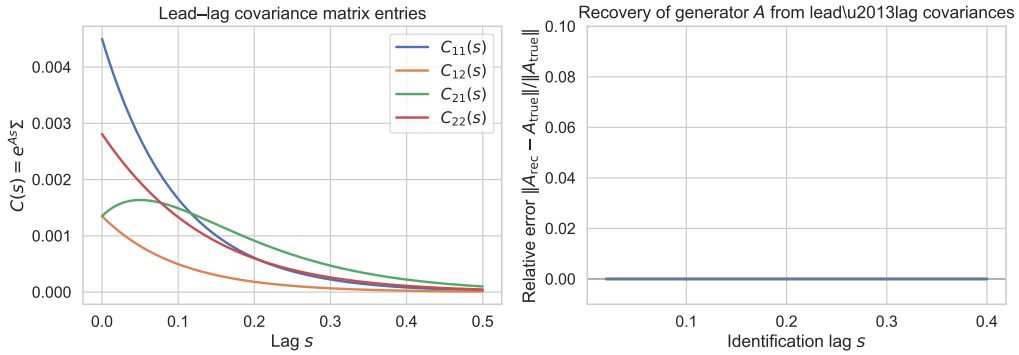

Figure 7: Lead–lag asymmetry and generator recovery in a two-channel OU linearization.

4.2 Covariance map and moment identification

The lead–lag approach recovers the full generator $\mathcal{A}$ but requires a locally stationary window and an OU approximation. A complementary strategy works directly with conditional covariance moments under an *affine reduced-form parametrization* $J(b_t) = I_{n+1} + \sum_{m=1}^{M} b_{m,t}K_m$, a first-order approximation to the exact resolvent that yields sparse, estimable feedback representations; theorems 4.4, 4.7 and 4.8 develop this reduced-form identification layer, which complements the exact conjugation of theorem 3.11 with practically computable moment conditions.

Proposition 4.4 (Covariance map under increment model)**.** *Assume the increment model for observed returns*

$$\Delta Y_t = J(b_t)\,\Delta Z_t + \varepsilon_t, \qquad J(b_t) = I_{n+1} + \sum_{m=1}^{M} b_{m,t}\,K_m, \tag{13}$$

where ΔZ_t are fundamental returns with $\mathbb{E}[\Delta Z_t \mid \mathcal{F}_{t-1}] = 0$ and covariance $\Sigma_t^Z := \mathrm{Cov}(\Delta Z_t \mid \mathcal{F}_{t-1})$, and ε_t is conditionally mean-zero noise independent of ΔZ_t given $\mathcal{F}_{t-1}$, with covariance $\Omega_t := \mathrm{Cov}(\varepsilon_t \mid \mathcal{F}_{t-1})$. Then the observed covariance is

$$\Sigma_t^Y := \mathrm{Cov}(\Delta Y_t \mid \mathcal{F}_{t-1}) = J(b_t)\,\Sigma_t^Z\,J(b_t)^\top + \Omega_t. \tag{14}$$

Proof. Conditioning on $\mathcal{F}_{t-1}$, since $\mathbb{E}[\Delta Z_t \mid \mathcal{F}_{t-1}] = 0$ and $\mathbb{E}[\varepsilon_t \mid \mathcal{F}_{t-1}] = 0$, we have $\mathbb{E}[\Delta Y_t \mid \mathcal{F}_{t-1}] = 0$. Write $J := J(b_t)$. Then

$$\Sigma_t^Y = \mathbb{E}[\Delta Y_t\,\Delta Y_t^\top \mid \mathcal{F}_{t-1}] = \mathbb{E}[(J\,\Delta Z_t + \varepsilon_t)(J\,\Delta Z_t + \varepsilon_t)^\top \mid \mathcal{F}_{t-1}].$$

Expanding and using conditional independence, $\mathbb{E}[J\,\Delta Z_t\,\varepsilon_t^\top \mid \mathcal{F}_{t-1}] = J\,\mathbb{E}[\Delta Z_t \mid \mathcal{F}_{t-1}]\,\mathbb{E}[\varepsilon_t^\top \mid \mathcal{F}_{t-1}] = 0$, and similarly for the transpose term. Thus $\Sigma_t^Y = J\,\mathbb{E}[\Delta Z_t\,\Delta Z_t^\top \mid \mathcal{F}_{t-1}]\,J^\top + \mathbb{E}[\varepsilon_t\,\varepsilon_t^\top \mid \mathcal{F}_{t-1}] = J\,\Sigma_t^Z\,J^\top + \Omega_t$. □

4.3 Identification by signature moments

Within the reduced-form parametrization just introduced, the covariance map of theorem 4.4 expresses the observed covariance as a quadratic function of the gain parameters. To extract individual gains, additional structure is needed. We specialize to two channels ($M = 2$) for clarity; the argument extends to more channels with disjoint signatures.

Assumption 4.5 (Rank-one channels and exact signature isolation)**.** Let

$$K_1 = e_{i_1} e_{j_1}^\top, \qquad K_2 = e_{i_2} e_{j_2}^\top,$$

and let $J(b_t) = I_{n+1} + b_{1,t}K_1 + b_{2,t}K_2$. Fix signature indices k_1, k_2.

Assume:

(i) The fundamental covariance entries $\Sigma^Z_{j_1j_1}$, $\Sigma^Z_{i_1j_1}$, $\Sigma^Z_{j_1k_1}$, $\Sigma^Z_{i_1k_1}$, $\Sigma^Z_{j_2j_2}$, $\Sigma^Z_{i_2j_2}$, $\Sigma^Z_{j_2k_2}$, $\Sigma^Z_{i_2k_2}$ are known (or time-invariant on the estimation window).

(ii) On the signature coordinates for channel 1, channel 2 is exactly absent:

$$\Delta Y_{j_1} = \Delta Z_{j_1}, \qquad \Delta Y_{k_1} = \Delta Z_{k_1}, \qquad \Delta Y_{i_1} = \Delta Z_{i_1} + b_{1,t}\Delta Z_{j_1}.$$

(iii) On the signature coordinates for channel 2, channel 1 is exactly absent:

$$\Delta Y_{j_2} = \Delta Z_{j_2}, \qquad \Delta Y_{k_2} = \Delta Z_{k_2}, \qquad \Delta Y_{i_2} = \Delta Z_{i_2} + b_{2,t}\Delta Z_{j_2}.$$

(iv) Any residual noise term Ω_t is either zero or known and subtracted off before applying the formulas below.

These conditions hold, for example, under the support restrictions of theorem A.7, where the signature rows/columns are untouched by the other channel and the only active feedback contribution on the signature block is the designated rank-one term.

Remark 4.6 (Practical status of the signature assumption). Theorem 4.5 is deliberately strong: it isolates two channels by placing identifying moment conditions on covariance entries that are (approximately) stable and uncontaminated by other channels. Its role is to provide a transparent, closed-form identification argument; in data one should treat it as an *approximation* and check it. Operationally, the over-ID discrepancy $d_{m,t} := b^{(a)}_{m,t} - b^{(b)}_{m,t}$ (and its aggregation $\mathcal{T}_m$ in theorem 4.8) serves as a built-in diagnostic: if the chosen signature entries are contaminated or drifting, $d_{m,t}$ will exhibit systematic bias or persistence rather than behaving like estimation noise. The appendix conditions of section A.3 give algebraic sufficient conditions for exact isolation; when these fail, one can still use the signature construction as a heuristic by choosing entries that are empirically stable and minimally cross-exposed, and by reporting the over-ID residual as a specification check.

Proposition 4.7 (Two identifying equations per gain)**.** *Under theorem 4.5, assume moreover that*

$$\Sigma^Z_{j_1 j_1} \neq 0, \qquad \Sigma^Z_{j_1 k_1} \neq 0, \qquad \Sigma^Z_{j_2 j_2} \neq 0, \qquad \Sigma^Z_{j_2 k_2} \neq 0.$$

Then each gain satisfies:

$$b^{(a)}_{1,t} = \frac{\Sigma^Y_{i_1 j_1}(t) - \Sigma^Z_{i_1 j_1}}{\Sigma^Z_{j_1 j_1}}, \qquad b^{(b)}_{1,t} = \frac{\Sigma^Y_{i_1 k_1}(t) - \Sigma^Z_{i_1 k_1}}{\Sigma^Z_{j_1 k_1}}, \tag{15}$$

and similarly

$$b^{(a)}_{2,t} = \frac{\Sigma^Y_{i_2 j_2}(t) - \Sigma^Z_{i_2 j_2}}{\Sigma^Z_{j_2 j_2}}, \qquad b^{(b)}_{2,t} = \frac{\Sigma^Y_{i_2 k_2}(t) - \Sigma^Z_{i_2 k_2}}{\Sigma^Z_{j_2 k_2}}. \tag{16}$$

Moreover, in the correctly specified model $b^{(a)}_{m,t} = b^{(b)}_{m,t}$, yielding over-identifying restrictions. This matches the 2D benchmark of theorem 3.18 where $\beta = (c^Y_{SU} - \sigma_{SU})/\sigma^2_S$.

Proof. By theorem 4.5(ii), on the channel-1 signature coordinates,

$$\Delta Y_{i_1} = \Delta Z_{i_1} + b_{1,t}\Delta Z_{j_1}, \qquad \Delta Y_{j_1} = \Delta Z_{j_1}, \qquad \Delta Y_{k_1} = \Delta Z_{k_1}.$$

Taking conditional covariances gives

$$\Sigma^Y_{i_1 j_1}(t) = \mathrm{Cov}(\Delta Z_{i_1} + b_{1,t}\Delta Z_{j_1},\, \Delta Z_{j_1}) = \Sigma^Z_{i_1 j_1} + b_{1,t}\,\Sigma^Z_{j_1 j_1},$$

hence $b_{1,t} = (\Sigma^Y_{i_1 j_1}(t) - \Sigma^Z_{i_1 j_1})/\Sigma^Z_{j_1 j_1}$. Likewise,

$$\Sigma^Y_{i_1 k_1}(t) = \mathrm{Cov}(\Delta Z_{i_1} + b_{1,t}\Delta Z_{j_1},\, \Delta Z_{k_1}) = \Sigma^Z_{i_1 k_1} + b_{1,t}\,\Sigma^Z_{j_1 k_1},$$

which yields the second formula in (15). The argument for channel 2 is identical using theorem 4.5(iii). □

Because each gain is identified from two independent moment conditions, the over-identifying restrictions in theorem 4.7 yield a specification test.

Proposition 4.8 (Over-ID residual is zero in population)**.** *Under theorem 4.5, define*

$$\mathcal{T}_m := \frac{1}{T}\sum_{t=1}^{T}\big(b^{(a)}_{m,t} - b^{(b)}_{m,t}\big)^2. \tag{17}$$

Then $\mathcal{T}_m = 0$ in population (i.e. for the true conditional covariances Σ^Y_t).

Proof. By theorem 4.7, both formulas equal the same true $b_{m,t}$ for each t. Hence each summand is zero. □

Remark 4.9 (Diagnostic use of the over-ID statistic). In practice $\mathcal{T}_m$ is computed using realized covariance estimators $\widehat{\Sigma}^Y_t$; then $\mathcal{T}_m$ becomes a diagnostic for (i) signature invariance violations, (ii) estimator bias/leakage, or (iii) model misspecification. Because $(\widehat{\Sigma}^Y_t)$ and the implied $(b^{(a)}_{m,t}, b^{(b)}_{m,t})$ are serially dependent on rolling windows, finite-sample calibration of "large" versus "small" $\mathcal{T}_m$ should use a dependence-robust device (e.g. a block bootstrap over time, or HAC standard errors for the mean discrepancy $T^{-1}\sum_t d_{m,t}$ with $d_{m,t} := b^{(a)}_{m,t} - b^{(b)}_{m,t}$). In the externally validated simulation procedure, we treat $\mathcal{T}_m$ primarily as a *model diagnostic* rather than as an asymptotic χ^2 test, and report it alongside other fit checks (loss-bound R^2, correlation between $\widehat{\mathsf{AI}}$ and oracle AI^*).

Sufficient algebraic conditions for uncontaminated identification of individual feedback channels are given in section A.3 (theorem A.7 and theorem A.8).

The identification results above determine *which* feedback channels are active and *how strongly*; the next section quantifies the cumulative risk consequences.

5 Cumulative Feedback Risk

The preceding sections establish instantaneous diagnostics: the amplification index AI_t, the stability margin m, and the covariation conjugation. This section develops four cumulative risk tools: a pathwise loss bound (section 5.1), a Lyapunov exponent characterizing long-run amplification (section 5.2), a feedback-aware Riccati benchmark (section 5.3), and a benchmark ranking result for the amplification index (section 5.4). The development is motivated by the Nikkei Uridashi episode of 2012 to 2013, in which dealer hedging of autocallable equity notes generated feedback-driven covariation distortion and aggregate losses exceeding $500M.

Remark 5.1 (Hierarchy of claims). For clarity, the paper's claims should be read in the following order.

(i) theorem 2.7 is an exact semimartingale identity relative to a chosen desk operator $\widehat{c}$.

(ii) theorems 3.11 and 3.13 are exact consequences of the instantaneous-clearing hypothesis theorem 3.2 and the local branch regularity assumptions.

(iii) theorems A.2, 4.1, 4.7, 5.4, 5.5 and 5.9 are reduced-form identification or diagnostic results that require additional model assumptions and should not be interpreted as exact consequences of the accounting identity alone.

5.1 Pathwise loss bound and feedback VaR multiplier

Proposition 5.2 (Feedback-induced loss bound)**.** *Assume the hypotheses of theorem 3.13 and suppose the desk calibrates to fundamental covariation: $\widehat{c}_t = c_t^Z$. Then the feedback-induced hedging loss satisfies*

$$|L_{\mathrm{fb}}| := \frac{1}{2}\left|\int_0^T H_t : \left(J_t c_t^Z J_t^\top - c_t^Z\right) \mathrm{d}t\right| \le \frac{1}{2}\int_0^T (\mathsf{AI}_t^2 + 1)\,\|H_t\|_2\,\mathrm{Tr}(c_t^Z)\,\mathrm{d}t, \tag{18}$$

where AI_t is the amplification index (theorem 3.14).

Proof. By trace cycling, $H_t : (J_t c_t^Z J_t^\top) = \widetilde{H}_t : c_t^Z$, where $\widetilde{H}_t = J_t^\top H_t J_t$. For any symmetric A and PSD C, $|A : C| \le \|A\|_2 \mathrm{Tr}(C)$. By the triangle inequality, $|\widetilde{H}_t : c_t^Z - H_t : c_t^Z| \le |\widetilde{H}_t : c_t^Z| + |H_t : c_t^Z| \le (\|\widetilde{H}_t\|_2 + \|H_t\|_2)\,\mathrm{Tr}(c_t^Z) \le (\mathsf{AI}_t^2 + 1)\,\|H_t\|_2\,\mathrm{Tr}(c_t^Z)$, where the last step uses $\|\widetilde{H}_t\|_2 = \|J_t^\top H_t J_t\|_2 \le \|J_t\|_2^2 \|H_t\|_2 = \mathsf{AI}_t^2 \|H_t\|_2$. □

Remark 5.3 (Conservatism at zero feedback). When $B_t = 0$ the loss L_{fb} vanishes exactly, yet (18) yields $\int \|H_t\|\mathrm{Tr}(c_t^Z)\,\mathrm{d}t > 0$ because the triangle inequality discards the cancellation between $\widetilde{H} : c^Z$ and $H : c^Z$. A tighter form that respects this cancellation uses $\Delta J_t := J_t - I_{n+1} = B_t J_t$: the factorization $\widetilde{H}_t - H_t = \Delta J_t^\top H_t + H_t \Delta J_t + \Delta J_t^\top H_t \Delta J_t$ gives $\|\widetilde{H}_t - H_t\|_2 \le (2\|\Delta J_t\| + \|\Delta J_t\|^2)\,\|H_t\|_2$. Since $\|\Delta J_t\| \le \|B_t\|\,\mathsf{AI}_t$, this bound vanishes when $B_t = 0$. In the scalar case $\|\Delta J_t\| = \mathsf{AI}_t - 1$, and the factored bound reduces to $(\mathsf{AI}_t^2 - 1)\|H_t\|\mathrm{Tr}(c_t^Z)$, which is sharp. For the dominant regime $\mathsf{AI}_t \gg 1$ the two forms coincide at leading order.

Corollary 5.4 (Feedback VaR multiplier)**.** *Condition on the predictable paths of J_t and H_t, and write*

$$f_t := (\mathsf{AI}_t^2 + 1)\|H_t\|_2.$$

Suppose that, conditionally on these paths, the random variable

$$\int_0^T f_t\,\mathrm{Tr}(c_t^Z)\,\mathrm{d}t$$

is sub-Gaussian with variance proxy at most

$$\sigma_c^2 \int_0^T f_t^2 \, \mathrm{d}t.$$

Then, conditionally on $(J_t, H_t)_{t\in[0,T]}$*, the feedback-induced loss satisfies*

$$\mathbb{P}\big(|L_{\mathrm{fb}}| > x \mid (J_t), (H_t)\big) \le 2\exp\!\left(-\frac{x^2}{2\sigma_{\mathrm{fb}}^2}\right),$$

with

$$\sigma_{\mathrm{fb}}^2 = \frac{1}{4}\,\sigma_c^2 \int_0^T f_t^2 \, \mathrm{d}t \le \frac{T}{4}\,\sigma_c^2 \sup_{t\in[0,T]} f_t^2.$$

In particular, conditional Value-at-Risk scales at most linearly in the factor (AI^2+1) *and as* $\sqrt{T}$ *in horizon. If, in addition, there exists a deterministic constant* $\bar\sigma^2$ *such that*

$$\frac{1}{4}\,\sigma_c^2 \int_0^T f_t^2 \, \mathrm{d}t \le \bar\sigma^2 \qquad a.s.,$$

then the same sub-Gaussian bound holds unconditionally with variance proxy $\bar\sigma^2$*.*

Proof. By theorem 5.2,

$$|L_{\mathrm{fb}}| \le \frac{1}{2}\int_0^T f_t \,\mathrm{Tr}(c_t^Z) \, \mathrm{d}t.$$

By the conditional sub-Gaussian assumption,

$$\mathbb{P}\left(\frac{1}{2}\int_0^T f_t \,\mathrm{Tr}(c_t^Z) \, \mathrm{d}t > x \;\middle|\; (J_t), (H_t)\right) \le 2\exp\!\left(-\frac{x^2}{2\cdot\frac{1}{4}\sigma_c^2\int_0^T f_t^2 \, \mathrm{d}t}\right).$$

This gives the conditional bound. The estimate using $\sup_t f_t^2$ follows from

$$\int_0^T f_t^2 \, \mathrm{d}t \le T \sup_t f_t^2.$$

The unconditional statement follows only under the additional deterministic a.s. bound

$$\frac{1}{4}\,\sigma_c^2 \int_0^T f_t^2 \, \mathrm{d}t \le \bar\sigma^2.$$

□

The simulations are consistent with the AI^2+1 scaling predicted by theorem 5.4: at $\mathsf{AI}=6$, the pathwise loss bound reaches 37 times the baseline curvature-variance product, matching the $\mathsf{AI}^2+1=37$ benchmark scaling; at $\mathsf{AI}=2$ (lower end of the simulated range), the multiplier is 5. Even modest feedback ($\Lambda=0.05$) produces a measurable attenuation jump from 0.013 to 0.165, indicating that the bound is practically relevant well below the extreme regime.

In Figure 8, the left panel plots the pathwise bound $|L_{\mathrm{fb}}| \le \frac{1}{2}\int(\mathsf{AI}^2+1)\|H\|\mathrm{Tr}(c^Z)\,\mathrm{d}t$ as a function of AI for several curvature levels (theorem 5.2), while the right panel shows the induced 99% VaR scaling (theorem 5.4). The shaded band marks the simulated Uridashi range $\mathsf{AI}\approx 2$ to 5.

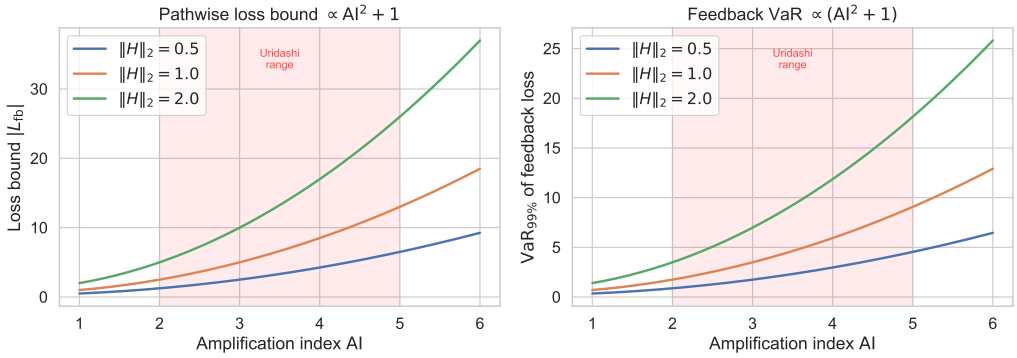

Figure 8: Loss bound and tail-risk scaling with the amplification index.

5.2 Lyapunov exponents and parametric resonance

The loss bound of theorem 5.2 treats the amplification index as a static quantity. In practice, however, J_t varies over time (e.g., as spot traverses barrier regions), the cumulative feedback effect is governed by the propagator of the linearized dynamics. Let $\mathcal{A}_t = \tau^{-1}(B_t - I_{n+1})$ be piecewise constant on a partition $\{t_k\}_{k=0}^N$ of $[0, T]$ with step Δt. Define the propagator

$$\Psi(t_n, 0) := \prod_{k=1}^{n} e^{\mathcal{A}_{t_k}\Delta t} = e^{\mathcal{A}_{t_n}\Delta t} \cdots e^{\mathcal{A}_{t_1}\Delta t}.$$

Theorem 5.5 (Pathwise propagator growth; sufficient condition for second-moment growth bounds)**.** *Consider the linear random system*

$$\mathrm{d}X_t = \mathcal{A}_t X_t \, \mathrm{d}t + \Gamma_t \, \mathrm{d}W_t, \tag{19}$$

where $(\mathcal{A}_t, \Gamma_t)$ is stationary and ergodic, and

$$\mathbb{E}\log^+ \|\mathcal{A}_0\| < \infty, \qquad \mathbb{E}\log^+ \|\Gamma_0\| < \infty.$$

Let $\Psi(t, s)$ denote the homogeneous propagator generated by $\mathcal{A}$, so that

$$\partial_t \Psi(t, s) = \mathcal{A}_t \Psi(t, s), \qquad \Psi(s, s) = I.$$

Then the top Lyapunov exponent

$$\lambda_1 := \lim_{t\to\infty} \frac{1}{t} \log \|\Psi(t, 0)\| \tag{20}$$

exists almost surely and is deterministic.

(i) ***(Pathwise Lyapunov exponent).*** *The limit (20) exists almost surely and is deterministic.*

(ii) ***(Covariance representation).*** *If $X_0 \in L^2$ is independent of $(\mathcal{A}_s, \Gamma_s, W_s - W_0)_{s\geq 0}$, and if $\int_0^t \|\Psi(t, s)\Gamma_s\|_F^2 \, \mathrm{d}s < \infty$ a.s. for every t, then*

$$X_t = \Psi(t, 0)X_0 + \int_0^t \Psi(t, s)\Gamma_s \, \mathrm{d}W_s,$$

and

$$\Sigma(t) := \mathbb{E}[X_t X_t^\top] = \mathbb{E}\big[\Psi(t, 0)X_0 X_0^\top \Psi(t, 0)^\top\big] + \mathbb{E}\left[\int_0^t \Psi(t, s)\Gamma_s\Gamma_s^\top \Psi(t, s)^\top \, \mathrm{d}s\right].$$

(iii) ***(Sufficient condition for an upper second-moment growth bound).*** *Assume there exists $\beta \in \mathbb{R}$ and $C < \infty$ such that*

$$\mathbb{E}\|\Psi(t,s)\|^2 \le Ce^{2\beta(t-s)} \qquad \text{for all } 0 \le s \le t,$$

and

$$\sup_t \mathbb{E}\|\Gamma_t\|_F^2 < \infty.$$

Then

$$\|\Sigma(t)\| \le C_1 e^{2\beta t} + C_2 \int_0^t e^{2\beta(t-s)}\, \mathrm{d}s,$$

for finite constants C_1, C_2 depending on $\mathbb{E}\|X_0\|^2$ and $\sup_t \mathbb{E}\|\Gamma_t\|_F^2$. In particular,

$$\limsup_{t\to\infty} \frac{1}{t} \log \|\Sigma(t)\| \le 2\beta.$$

(iv) ***(Diagnostic interpretation).*** *If, in a given model class, one can verify a second-moment bound with exponent $\beta > 0$, then persistent positive propagator growth implies exponential amplification of second moments on that scale.*

Proof. Part (i) follows from the subadditivity of $t \mapsto \log\|\Psi(t,0)\|$ (cocycle identity + stationarity) and the Furstenberg–Kesten theorem [30] (Oseledets [20]), given the integrability condition. Part (ii) uses the mild solution $X_t = \Psi(t,0)X_0 + \int_0^t \Psi(t,s)\Gamma_s\, \mathrm{d}W_s$, obtained by the integrating-factor change $Y_s := \Psi(s,0)^{-1}X_s$; the cross term in $\mathbb{E}[X_t X_t^\top]$ vanishes when $\mathbb{E}[X_0] = 0$ by independence and the Itô isometry. Part (iii) applies the triangle inequality and the second-moment propagator bound together with Fubini's theorem; the growth rate bound follows by elementary calculus. □

Remark 5.6 (Stationarity and finite-horizon diagnostics). The stationarity hypothesis in theorem 5.5 ensures that the Furstenberg–Kesten limit exists. In practice, $\mathcal{A}_t = \tau^{-1}(B_t - I)$ depends on the state of the market (spot relative to barriers, inventory, etc.) and is generically nonstationary during the stress episodes motivating this analysis. The rolling estimator $\widehat{\lambda}_1 = \frac{1}{N\Delta t}\log\|\prod_{k=1}^N e^{\widehat{\mathcal{A}}_{t_k}\Delta t}\|$ used in the pseudo-algorithm is well-defined without stationarity: it is a finite-horizon growth rate, not a limit. Theorem 5.5 identifies the pathwise propagator exponent exactly; turning that into deterministic covariance growth bounds requires additional second-moment control such as the sufficient condition stated there. In nonstationary regimes, $\widehat{\lambda}_1$ should be interpreted as a *local* indicator of cumulative amplification over the estimation window, not as a global exponent.

Remark 5.7 (Transient growth bound via numerical abscissa). Lyapunov exponents are asymptotic and can miss large *finite-time* transients, especially when generators are non-normal. A complementary (conservative) finite-horizon diagnostic comes from the numerical abscissa (logarithmic norm) for the operator-2 norm:

$$\omega(\mathcal{A}) \;:=\; \lambda_{\max}\Big(\tfrac{1}{2}(\mathcal{A} + \mathcal{A}^\top)\Big).$$

It bounds the growth of the propagator: $\|e^{\mathcal{A}t}\| \le e^{\omega(\mathcal{A})t}$ for $t \ge 0$. For piecewise-constant switching as in (20),

$$\|\Psi(t_n,0)\| \;=\; \Big\|\prod_{k=1}^n e^{\mathcal{A}_{t_k}\Delta t}\Big\| \;\le\; \exp\Big(\sum_{k=1}^n \omega(\mathcal{A}_{t_k})\,\Delta t\Big).$$

Thus $\sum_k \omega(\widehat{\mathcal{A}}_{t_k})\Delta t$ provides an easily computed upper bound on finite-window amplification, even when commutation fails and eigenvalue margins are misleading.

Corollary 5.8 (Parametric resonance: local stability $\not\Rightarrow$ global stability)**.** *It is possible that* $\max_i \Re(\lambda_i(\mathcal{A}_t)) < 0$ *for every* t (each instantaneous generator is Hurwitz) *while* $\lambda_1 > 0$ (the long-run system is exponentially unstable). *The instantaneous stability margin* $m_t > 0$ *(which is* τ*-free and directly estimable from lead–lag covariances, theorem 4.2) is therefore necessary but not sufficient for long-run stability; the Lyapunov exponent* λ_1 *is the correct global criterion [21].*

Proof. By explicit construction in $\mathbb{R}^2$. Let $\Delta t = 1$ and define two Hurwitz generators

$$\mathcal{A}_1 = \begin{pmatrix} -\epsilon & M \\ 0 & -\epsilon \end{pmatrix}, \qquad \mathcal{A}_2 = \begin{pmatrix} -\epsilon & 0 \\ M & -\epsilon \end{pmatrix},$$

with $M > 0$ and $\epsilon > 0$. Both have eigenvalues $-\epsilon < 0$. Since $(\mathcal{A}_i + \epsilon I)^2 = 0$, we get $e^{\mathcal{A}_i} = e^{-\epsilon}(I + \mathcal{A}_i + \epsilon I)$. A direct computation gives

$$e^{\mathcal{A}_1} e^{\mathcal{A}_2} = e^{-2\epsilon} \begin{pmatrix} 1 + M^2 & M \\ M & 1 \end{pmatrix},$$

whose larger eigenvalue is $e^{-2\epsilon}\frac{1}{2}\big(2 + M^2 + M\sqrt{M^2+4}\big)$. For $M = 2$, this equals $e^{-2\epsilon}(3 + 2\sqrt{2}) \approx 5.83\, e^{-2\epsilon}$, which exceeds 1 whenever $\epsilon < \frac{1}{2}\ln(3 + 2\sqrt{2}) \approx 0.88$. Hence the two-step propagator has $\lambda_1 = \frac{1}{2}\log \rho(e^{\mathcal{A}_1} e^{\mathcal{A}_2}) > 0$ despite $m_t = \epsilon > 0$ at every t. □

Numerical evaluation yields a critical threshold at $\epsilon^* = \frac{1}{2}\ln(3 + 2\sqrt{2}) \approx 0.881$: for $\epsilon < \epsilon^*$, the system is exponentially unstable ($\lambda_1 > 0$) despite each instantaneous generator being Hurwitz. In the synthetic switching experiment reported here, the sign-reversal diagnostic, which flags episodes where the rolling $\hat{\lambda}_1$ turns positive while m_t remains strictly positive, substantially outperforms the simpler attenuation-based screen; this is consistent with parametric resonance being detectable from covariation data in that benchmark setting (fig. 9).

The three-dimensional stability surface (fig. 13) reveals the geometry of this instability region across the full (ϵ, M) plane. The critical boundary $\epsilon^*(M)$ grows monotonically with coupling strength: at $M = 0.5$ (weak coupling), $\epsilon^* \approx 0.24$ and even modest stability margins suffice; at $M = 2$ (the two-state benchmark), $\epsilon^* \approx 0.88$ as the analytic formula predicts; but at $M = 4$ (strong coupling), ϵ^* exceeds 1.4, meaning even substantial local margins fail. The maximum Lyapunov exponent reaches $\lambda_1 \approx 1.39$ at minimal margin and maximal coupling, corresponding to a doubling time of 0.5 periods. For the Uridashi application, the off-diagonal coupling M maps to the ratio of cross-asset hedging flow to own-market liquidity; the implication is that desks hedging a diversified book (large M) face a wider instability region than those with concentrated exposures. The externally validated procedure is consistent with these patterns: fig. 18(A) shows the rolling $\hat{\lambda}_1$ trajectory from desk-observable data, while panels (C,D) are numerically consistent with the $P^* \propto j^{-2}$ and $\mathrm{Var}(l) \propto j^2$ scalings of theorem 5.10.

5.3 Feedback-aware monitoring via Riccati filtering

We now introduce a *continuous-time filtering benchmark* that makes the estimation-loss coupling analytically transparent. This reduced-form benchmark complements the exact accounting identities of theorems 2.7 and 3.11 by isolating the structural tradeoff between estimation precision and hedging-loss variance in a tractable scalar model. The benchmark models a latent fundamental variance state observed through a noisy continuous-time covariation channel whose sensitivity is amplified by feedback. [4]

[4]The scalar j_t is the one-dimensional analogue of the amplification matrix $J_t = (I - B_t)^{-1}$: it is the local gain mapping a fundamental variance scale to a traded variance scale in a single observed direction. In full multivariate models, J_t is a matrix and amplification is direction-dependent; $\mathsf{AI}_t = \|J_t\|_2$ is a worst-case summary and can replace j_t in bounds, but the scalar benchmark is sufficient to exhibit the locked estimation-loss coupling.

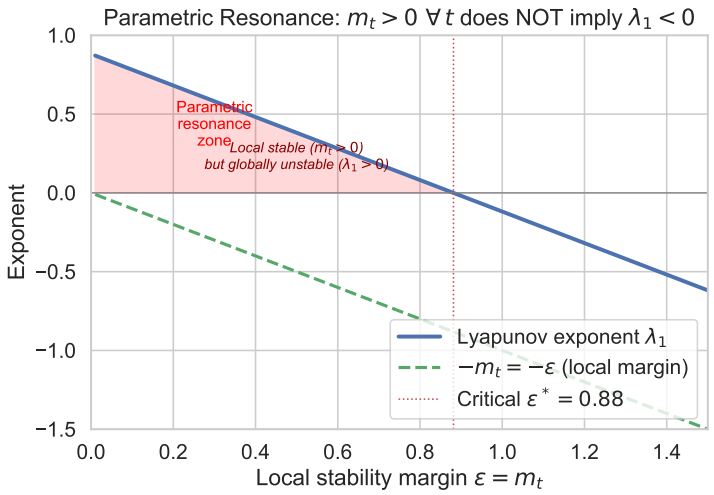

Figure 9: Parametric resonance: local stability does not imply global stability.

Proposition 5.9 (Continuous-time filtering benchmark under amplified observation)**.** *Let the latent fundamental variance state follow the scalar OU diffusion*

$$\mathrm{d}x_t = -\kappa x_t\ \mathrm{d}t + \nu\ \mathrm{d}W_t^{(x)}, \qquad \kappa, \nu > 0, \tag{21}$$

and let the desk observe the continuous-time process

$$\mathrm{d}O_t = j(t)^2\, x_t\ \mathrm{d}t + \sqrt{R}\ \mathrm{d}V_t, \qquad R > 0, \tag{22}$$

where $j : [0,T] \to (0,\infty)$ is deterministic, bounded, and measurable, and V is a Brownian motion independent of $W^{(x)}$. Then the Kalman–Bucy filter $\widehat{x}_t := \mathbb{E}[x_t \mid \mathcal{F}_t^O]$ has conditional estimation-error variance

$$P_t := \mathbb{E}[(x_t - \widehat{x}_t)^2 \mid \mathcal{F}_t^O],$$

which is deterministic and satisfies the Riccati ODE

$$\dot{P}_t = -2\kappa P_t + \nu^2 - \frac{j(t)^4}{R} P_t^2. \tag{23}$$

If $j(t) \equiv j$ is constant, the unique positive steady state is

$$P^*(j) = \frac{R}{j^4}\left(\sqrt{\kappa^2 + \nu^2 j^4/R} - \kappa\right). \tag{24}$$

In particular,

$$P^*(j) \sim \frac{\nu\sqrt{R}}{j^2} \qquad \text{as } j \to \infty.$$

Proof. This is the scalar continuous-time Kalman–Bucy filter with deterministic observation coefficient $H_t = j(t)^2$, state matrix $A = -\kappa$, process-noise variance $Q = \nu^2$, and observation-noise variance R. Because H_t is deterministic, the Riccati ODE for P_t is a deterministic ODE. Standard filtering theory gives

$$\dot{P}_t = AP_t + P_t A^\top + Q - P_t H_t^\top R^{-1} H_t P_t,$$

which here becomes (23).

For constant j, the steady state solves

$$0 = -2\kappa P + \nu^2 - \frac{j^4}{R} P^2.$$

The positive root is (24). Expanding for large j gives

$$P^*(j) = \frac{R}{j^4}\left(\frac{\nu j^2}{\sqrt{R}} + O(1)\right) = \frac{\nu\sqrt{R}}{j^2} + O(j^{-4}).$$

□

Corollary 5.10 (Benchmark estimation-loss scaling under amplified observation)**.** *In the setting of theorem 5.9, suppose the desk's instantaneous hedging-loss proxy is proportional to amplified estimation error:*

$$\ell_t = j(t)^2\,(x_t - \widehat{x}_t). \tag{25}$$

Then

$$\mathrm{Var}(\ell_t \mid \mathcal{F}_t^O) = j(t)^4 P_t.$$

If $j(t) \equiv j$ is constant and the filter is at steady state, then

$$\mathrm{Var}(\ell_t) = j^4 P^*(j) = R\Big(\sqrt{\kappa^2 + \nu^2 j^4/R} - \kappa\Big), \tag{26}$$

so that

$$\mathrm{Var}(\ell_t) \sim \nu\sqrt{R}\,j^2 \qquad \text{as } j \to \infty.$$

Thus, in this benchmark, stronger amplification improves estimation of the latent state

$$P^*(j) = O(j^{-2}),$$

but worsens the variance of amplified hedging loss

$$\mathrm{Var}(\ell_t) = O(j^2).$$

Proof. The conditional variance identity is immediate from (25). Substituting (24) into $j^4 P^*(j)$ yields (26), and the large-j asymptotic follows from the previous proposition. □

Quantitatively, as j increases from 1 to 8, the steady-state estimation error P^* drops by a factor of 38 (from 0.057 to 0.0015), while the hedging-loss variance $\mathrm{Var}(l_t)$ increases by a factor of 102 (from 0.015 to 1.505). The tradeoff is lopsided: the j^4 sensitivity amplification dwarfs the j^{-2} estimation improvement by two orders of magnitude at the upper end of the simulated range.

Remark 5.11 (Interpretive scope of the Riccati benchmark). When j_t is stochastic but $\mathcal{F}_t^O$-adapted, the Riccati equation (23) holds pathwise; the deterministic benchmark is sufficient to exhibit the locked estimation-loss coupling ($P^* \sim j^{-2}$, $\mathrm{Var}(\ell) \sim j^2$), with all qualitative features preserved. The hedge is fixed (the gradient $\nabla_y v$), not a control variable, so only the filter Riccati appears, not a full LQG pair. The control-theoretic precedents (Bode's integral, risk-sensitive filtering, dual control) are recorded in section A.2; in the Uridashi-parameterized simulation ($\mathsf{AI} \approx 2$–5 near barriers), barrier-induced regime switching can produce $\lambda_1 > 0$ via parametric resonance (theorem 5.8), and the Riccati gain $K_t = j_t^2 P_t/R$ spikes accordingly near barriers.

5.4 Discriminative power of the amplification index

The preceding subsections bound, detect, and monitor feedback-induced risk. A natural question is whether the amplification index AI_t is the *right* summary statistic for these purposes. Under the stylized location-shift assumptions collected in theorem A.1 of Appendix A, the answer is affirmative in a benchmark sense: under the model of theorem A.2, any strictly increasing transformation of $\mathsf{AI}_t^{\mathrm{true}}$ is Bayes-optimal for ranking tail-loss events, and in the high-separation regime described there the corresponding AUC tends to 1. An estimated $\widehat{\mathsf{AI}}_t$ that is sufficiently

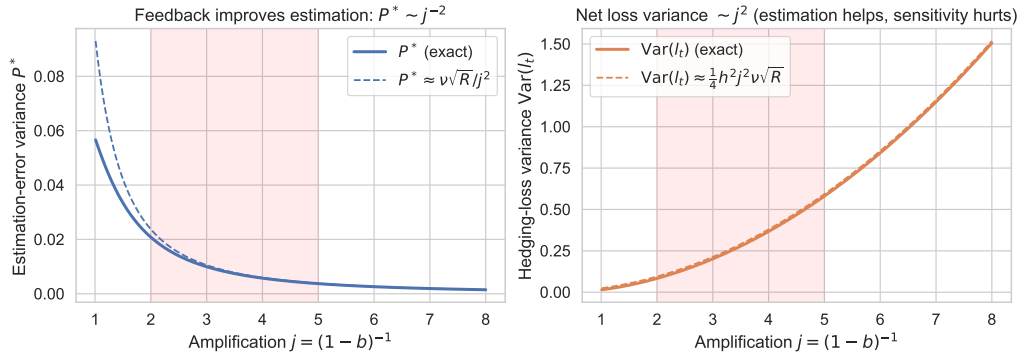

Figure 10: Riccati tradeoff: estimation improves while hedging-loss variance rises with amplification.

accurate preserves this ranking up to infrequent reversals (theorem A.5). The proofs are given in section A.

How the horse-race AUC is computed. Each candidate predictor produces a scalar score S_t; the AUC equals the concordance probability $\mathbb{P}(S_{t^+} > S_{t^-})$ (with standard tie correction) where t^+ and t^- are randomly drawn event and non-event days. AUC is monotone-invariant, threshold-free, and nonparametric, making it the natural metric for ranking continuous risk signals; full computational details (event-indicator construction, logistic-regression combination, block-bootstrap inference) are given in section B.

In the externally validated horse-race simulation (section B; ROC curves in fig. 17), realized volatility dominates single-predictor AUC (0.756), with barrier distance close behind (0.729), vol-of-vol at 0.632, and the estimated amplification index at 0.553. The incremental AUC of $\widehat{\mathsf{AI}}$ over realized volatility is +0.004, modest but positive; the margin reflects the structural complexity of the DIUO's Greek profile (vanna sign reversals, delta flips) relative to isotropic volatility measures. That realized volatility ranks first is unsurprising: it subsumes much of the feedback signal because feedback *inflates* realized volatility via the conjugation $c^Y = Jc^Z J^\top$. The amplification index adds explanatory power beyond this inflation because it captures the *directional* component of the distortion (lead–lag asymmetry, vanna sign reversals) that isotropic volatility measures cannot detect. The theoretical guarantee of theorem A.2 establishes oracle-level optimality; in practice, the amplification index adds directional and structural information (lead–lag asymmetry, vanna sign reversals) that complements standard risk indicators such as realized volatility. The reported AUCs are point estimates from the synthetic-desk procedure; incremental AUCs are accompanied by uncertainty bands from a circular block bootstrap (Politis and Romano, 1992) with block length $b = w = 20$ days matching the rolling-window size, capturing the autocorrelation induced by overlapping windows (theorem A.6).

The theoretical results of sections 2 to 4 (the identity, the conjugation, the loss bound, the Lyapunov criterion, the Riccati filter) depend on the traded-coordinate Hessian H_t and the conjugation $c^Y = Jc^Z J^\top$, not on which exotic product generates the flow map; the DIUO specification provides one numerical stress test of those mechanisms. The discriminative-power question, by contrast, *is* product-dependent (theorem 3.22): the amplification index's marginal contribution scales with the structural complexity of the feedback channel. We next present a minimal multi-coordinate extension that adds one additional vanilla strike (cross-strike propagation), then turn to modelling scope, empirical status, and an implementation roadmap.

6 Extensions: Multi-Coordinate Cross-Strike Feedback

The framework is formulated for any finite set of traded coordinates and any loop-gain matrix B. This section presents the first nontrivial *topological* increment beyond the two-coordinate

benchmark: adding one additional vanilla strike creates cross-strike propagation and a genuine feedback cycle inside the vanilla block.

6.1 Multi-coordinate extension: cross-strike feedback

Defensibility of the incremental sequence. The two-coordinate model (S, P) projects the vanilla surface onto a single liquid summary coordinate (the level factor, which accounts for 60 to 70% of surface variance). This rank-1 projection suffices to isolate the core structural claim: even with one vanilla, the mechanism $c^Y = Jc^Z J^\top$ already produces attenuation, neutralization, and reversal of covariation. But it cannot represent *cross-strike propagation*: with one vanilla there is no "other strike" through which a feedback cycle can pass.

The minimal dimensional extension that introduces a qualitatively new propagation mechanism is to add one vanilla put P_2 struck near the lower knock-in barrier ($K_2 = 70\%$ of S_0, approximately a 10Δ put in the simulation's vol regime, capturing the skew/wing factor), alongside the existing ATM put P_1 ($K_1 = S_0$, level factor). This produces a rank-2 projection of the surface (matching the empirical observation that two PCs capture 80 to 85% of smile variance) and introduces a single new parameter $\rho_{\text{surface}} \in [0, 1]$ controlling cross-strike feedback propagation. An increment is defensible if it is *nested* (reduces to the previous model at a parameter limit), introduces the *smallest new mechanism* the previous model cannot represent, stays within the *same axioms* ($B = D_Q \Lambda\, D_y \Phi$, $J = (I - B)^{-1}$), and yields *testable restrictions* rather than extra knobs. We now verify all four.

Model. The loop-gain matrix for the vanilla block becomes

$$B_{22} = \lambda\, c \begin{pmatrix} 1 & \rho_{\text{surface}} \\ \rho_{\text{surface}} & 1 \end{pmatrix},$$

where c is the scalar flow coefficient from the two-coordinate model. The first-column entries b_{10}, b_{20} (spot $\to$ put feedback) are determined by barrier Greeks (specifically the finite-difference delta and vega of the DIUO at each strike) and carry no free parameters. The off-diagonal entries of B_{22} are not free: $(B_{22})_{12} = \rho \cdot (B_{22})_{11}$ and $(B_{22})_{21} = \rho \cdot (B_{22})_{22}$, so ρ_{surface} is the *sole* new parameter, interpretable as "fraction of feedback that propagates across strikes" and, in principle, estimable from the ratio of off-diagonal to diagonal entries of $\hat{B}$.

Estimating ρ_{surface} from data. In practice, one can estimate the vanilla-block loop-gain $\widehat{B}_{22}(t)$ on rolling windows using the covariance-map identification of sections 4.1 and 4.3 (or any other estimator of B), then exploit the parametric restriction $\big(\widehat{B}_{22}\big)_{12} \approx \rho_{\text{surface}} \big(\widehat{B}_{22}\big)_{11}$ and $\big(\widehat{B}_{22}\big)_{21} \approx \rho_{\text{surface}} \big(\widehat{B}_{22}\big)_{22}$. A simple moment estimator is

$$\widehat{\rho}_{\text{surface}} \;:=\; \frac{1}{2}\left(\frac{(\widehat{B}_{22})_{12}}{(\widehat{B}_{22})_{11}} \;+\; \frac{(\widehat{B}_{22})_{21}}{(\widehat{B}_{22})_{22}} \right),$$

with trimming or ridge regularization when diagonal entries are small. The two ratios provide an internal consistency check (they should agree if the parametric coupling is correct); large discrepancies indicate either asymmetric cross-strike propagation or failure of the symmetric coupling specification. At $\rho_{\text{surface}} = 0$ the two puts decouple, recovering the single-put baseline exactly (*nested*). At $\rho_{\text{surface}} = 1$ every unit of hedging flow propagates identically to both strikes (common-channel limit); the two puts become a single factor and the model collapses back to rank-1. The coupled fixed point is a closed-form 2×2 matrix inversion per path-time: $\vec{P}_Y = (I_2 - \lambda c\, R)^{-1} \vec{P}_Z$, where $R = \left(\begin{smallmatrix} 1 & \rho \\ \rho & 1 \end{smallmatrix}\right)$, valid whenever $\det(I_2 - \lambda c\, R) = (1 - \lambda c)^2 - (\lambda c\, \rho)^2 \neq 0$, i.e. $\lambda c \neq 1/(1 \pm \rho)$. All Layer-2 machinery (conjugation, loss bound, identification) carries through unchanged (*same axioms*).

New phenomenon: cross-strike feedback cycle. The Neumann expansion $(I - B)^{-1} = I + B + B^2 + \cdots$ contains a B^2 cross-term

$$(B^2)_{12} = 2\lambda^2 c^2\, \rho.$$

This represents a two-step feedback cycle $P_1 \to P_2 \to P_1$: hedging-flow pressure on P_1 propagates to P_2 via the cross-coupling ρ, and returns to P_1 at the next round. This cycle is *provably absent* in any single-put model (*smallest new mechanism*). The eigenvalues of $I_2 - B_{22}$ are $1 - \lambda c(1 \pm \rho_{\text{surface}})$, so the amplification index becomes

$$\mathsf{AI}^{\text{3D}} = \frac{1}{\min\big(\big|1 - \lambda c(1{+}\rho_{\text{surface}})\big|,\ \big|1 - \lambda c(1{-}\rho_{\text{surface}})\big|\big)}.$$

Cross-coupling *lowers the critical feedback threshold* from $\lambda c = 1$ (single-put) to $\lambda c = \frac{1}{1+\rho_{\text{surface}}}$, making the system more fragile. At $\rho_{\text{surface}} = 1$ the perpendicular eigenvalue equals 1 (the skew eigenmode passes through with unit gain), so the two puts become a single effective coordinate and the three-coordinate model provides no differential information beyond the level factor (common-channel collapse). Both the threshold reduction and the $\rho = 1$ collapse are *testable predictions* (*empirically falsifiable*).

Backbone propagation across strikes. Here "backbone" and "skew" denote different objects and should not be conflated. Backbone is a *spot-direction* response, for example the co-movement $\langle dS,\, d\sigma_{\text{ATM}}\rangle_t$ or the sensitivity of ATM volatility to spot. Skew is a *strike-direction* object, namely the slope $\partial\sigma_{\text{imp}}/\partial K$ (or with respect to log-moneyness). The claim of the model is therefore not that backbone *is* skew. The claim is narrower and structural: a backbone-driven fundamental shock can acquire a cross-strike component in price space, and after the price-to-implied-volatility map is applied this can appear empirically as a skew move.

Under Black–Scholes with scalar volatility σ_t, the instantaneous put return at strike K_i is

$$dP_i^Z \;=\; \delta_i\, dS \;+\; \nu_i\, d\sigma \;+\; \tfrac{1}{2}\,\Gamma_i\,(dS)^2 \;+\; \cdots,$$

where $d\sigma$ is the *same* scalar shock at every strike. Even under this common-volatility specification, the two strikes need not move identically in price space because the Greeks differ:

$$dP_1^Z - dP_2^Z = (\delta_1{-}\delta_2)\, dS + (\nu_1{-}\nu_2)\, d\sigma + \cdots.$$

Thus a backbone shock need not be a purely common-mode shock for the two vanillas; heterogeneous Greeks generally give it both a common component and a differential component.

To make this precise, decompose the fundamental two-put shock into the eigenbasis of $R = \left(\begin{smallmatrix} 1 & \rho \\ \rho & 1 \end{smallmatrix}\right)$:

$$d\vec{P}_Z \;=\; a\, e_+ {+} b\, e_-, \qquad e_+ = \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \quad e_- = \begin{pmatrix} 1 \\ -1 \end{pmatrix}, \qquad a = \tfrac{1}{2}(dP_1^Z {+} dP_2^Z), \quad b = \tfrac{1}{2}(dP_1^Z {-} dP_2^Z).$$

The feedback fixed point then gives

$$d\vec{P}_Y \;=\; V\ \mathrm{diag}(\alpha_+, \alpha_-)\, V^\top d\vec{P}_Z \;=\; \alpha_+ a\, e_+ + \alpha_- b\, e_-,$$

with

$$\alpha_+ = \frac{1}{1 - \lambda c(1{+}\rho)}, \qquad \alpha_- = \frac{1}{1 - \lambda c(1{-}\rho)}, \qquad V = \tfrac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}.$$

This is the exact model statement. The common mode e_+ is the backbone channel, while e_- is the differential cross-strike channel. If $b = 0$, feedback preserves symmetry in price space: a purely common shock remains common. Hence feedback alone does not manufacture differential strike behavior from a perfectly symmetric price shock. What ρ controls is the separation between the common and differential amplifications through the eigenmode gap

$$\alpha_+ - \alpha_- = \frac{2\lambda c\rho}{(1 - \lambda c)^2 - (\lambda c\rho)^2}.$$

At $\rho = 0$ the two modes amplify equally; as ρ rises, the common and differential channels are increasingly separated.

This ordering is important for interpretation. The framework does *not* begin by postulating vegas or an implied-volatility surface dynamics. It begins with traded coordinates and a clearing relation in *price space*: the exotic enters through the hedge-demand map Φ (via quantities such as $N_{\text{put}} = \nu_{\text{DIUO}}/\nu_{\text{van}}$ in the uridashi application), and the exact output of the model is a statement about distorted vanilla prices and covariations. Vega enters only later, and only conditionally on a chosen volatility parametrization, as a Jacobian for translating those price perturbations into implied-volatility language.

The link to skew appears only after one passes from price space to implied-volatility space. To first order,

$$\Delta\sigma_{\text{imp}}(K_i) \;\approx\; \frac{P_i^Y - P_i^Z}{\nu_i},$$

where $\nu_i = \partial P_i/\partial\sigma$ is the Black–Scholes vega at K_i.[5] For a two-strike skew proxy,

$$\Delta\text{Skew}_{12} \;:=\; \frac{\Delta\sigma_{\text{imp}}(K_2) - \Delta\sigma_{\text{imp}}(K_1)}{K_2 - K_1} \;\approx\; \frac{1}{K_2 - K_1}\left(\frac{\Delta P_2^Y}{\nu_2} - \frac{\Delta P_1^Y}{\nu_1}\right).$$

Substituting the eigenmode representation shows that the observed skew response is generated by two distinct ingredients: a differential price component b and a strike-dependent inverse-vega map. In particular, if $b = 0$ and $\nu_1 = \nu_2$, then no skew is induced. If $b = 0$ but $\nu_1 \neq \nu_2$, a common price move can still appear as skew after implied-vol inversion; this is a Jacobian effect, not a cross-strike-propagation effect. If $b \neq 0$, then feedback matters through the differential treatment of the two eigenmodes, governed by $\alpha_+ - \alpha_-$.

Using Black–Scholes is therefore a conservative reading of the mechanism. Within the model, the rigorous result is about how feedback reshapes the common and differential *price* components across strikes. The statement about skew is an interpretive corollary obtained by mapping those price perturbations into implied-volatility space; genuinely strike-specific volatility dynamics would add a separate exogenous source of skew variation on top.

Simulation results. Figure 11 reports a ρ_{surface} sweep over $\{0, 0.25, 0.50, 0.75, 1.00\}$ with 200 paths and 252 daily steps. The incremental AUC of $\widehat{\text{AI}}$ beyond the best single-predictor baseline more than doubles from +0.026 (single-put, $\rho_{\text{surface}} = 0$) to +0.057 at $\rho_{\text{surface}} = 0.25$, peaking at $\rho_{\text{surface}} \approx 0.25to0.50$ (fig. 11C). The standalone AUC of $\widehat{\text{AI}}$ rises from 0.746 to 0.804, surpassing both realized volatility (0.762) and barrier distance (0.747) to become the strongest single predictor of tail hedging losses (fig. 12A). At $\rho_{\text{surface}} = 1$ the incremental AUC collapses to +0.002, consistent with the theoretical prediction.

The desk-side estimator solves the linear relation $(I_2 - B_{22})\vec{P}_Y = \vec{P}_Z$ for the scalar $b = \lambda c$ using both put channels:

$$\hat{b}_1 = \frac{P_1^{\text{mkt}} - P_1^{\text{model}}}{P_1^{\text{mkt}} + \rho\, P_2^{\text{mkt}}}, \qquad \hat{b}_2 = \frac{P_2^{\text{mkt}} - P_2^{\text{model}}}{P_2^{\text{mkt}} + \rho\, P_1^{\text{mkt}}},$$

averaged for robustness; at $\rho_{\text{surface}} = 0$ this reduces to the scalar price-ratio estimator. Figure 12B shows the three-coordinate model (+0.055) improves on the two-coordinate baseline (+0.026) by a factor of 2.1.

Scope and further increments. As anticipated by section 3.3.1, the two- and three-coordinate models form the first two steps in a *nested sequence of finite-dimensional projections* of

[5] This is the vanilla's own price-to-vol sensitivity, used here solely as an interpretive Jacobian from price space to implied-vol space. It should not be confused with the exotic's *traded-coordinate sensitivity* $\partial v/\partial\widetilde{P}_i = \nu_{\text{DIUO}}/\nu_i$, a component of $\nabla_y v$ that determines hedge demand and drives the feedback mechanism (section 3.4). The core framework operates entirely in traded-coordinate space and never requires this vol interpretation; the latter is auxiliary and is included only for readers accustomed to thinking in implied-vol terms (cf. the "no assumption about what vega means" remark following theorem 2.7).

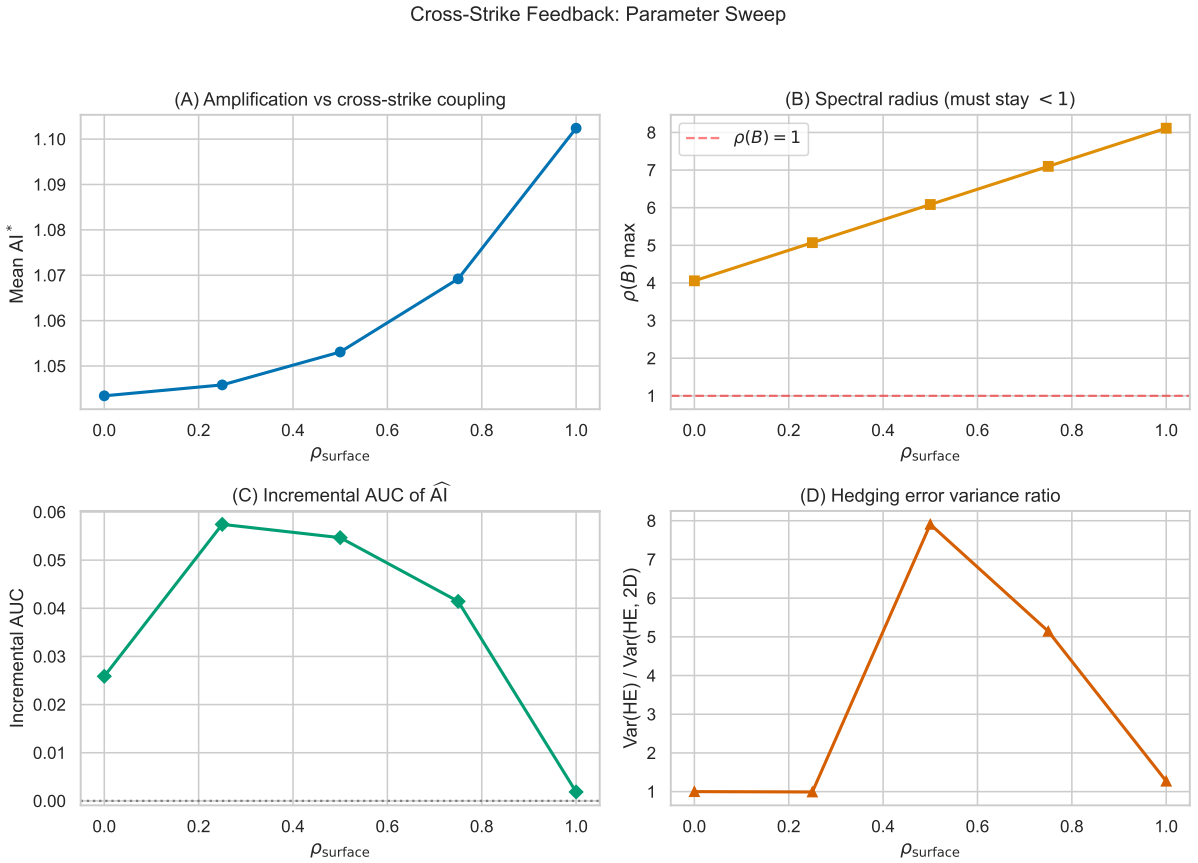

Figure 11: Cross-strike feedback parameter sweep over ρ_{surface}.

the volatility surface: rank-1 (level) $\rightarrow$ rank-2 (level + skew) $\rightarrow$ rank-p (richer surface structure), corresponding to the algebraic and topological leaps of fig. 4. Each step adds either one traded coordinate or one coupling mode, and each has a clear parameter reduction back to the previous one. The full-surface limit (fig. 4d) would let B_{vanillas} be an arbitrary matrix on a (K, T) grid; parsimony can be maintained by imposing nearest-neighbor, exponential-decay, or factor structure on the coupling. The present three-coordinate extension is the smallest step that introduces cross-strike propagation while preserving closed-form $J = (I - B)^{-1}$.

Figure 11 summarizes how cross-strike propagation changes instability and predictability: mean oracle amplification rises with ρ_{surface}, the fixed point becomes fragile as eigenvalues approach 1, and the incremental AUC of $\widehat{\text{AI}}$ peaks at intermediate coupling ($\rho_{\text{surface}} \approx 0.25 to 0.50$), where a genuine "skew" mode is present but not collapsed into the level factor.

Figure 12 shows the four-way comparison. Increment C (cross-strike coupling) makes $\widehat{\text{AI}}$ materially more informative: incremental AUC rises from +0.026 (A) to +0.055 (C). Increment B alone ($b_{01} = 0.1$, $\rho = 0$) raises mean AI^* to 1.06 but compresses incremental AUC to +0.008, because the realized-volatility baseline itself jumps from 0.762 to 0.795 (the redundancy mechanism of fig. 3). The combined B+C specification produces the highest mean AI^* (1.07) but its incremental AUC (+0.030) falls between B and C: the $S \leftrightarrow P$ cycle partially erodes the cross-strike informational gain, though the cross-strike signal is strong enough to survive. This is consistent with the claim that the *topology* of the feedback graph (not just the level of amplification) determines whether $\widehat{\text{AI}}$ is informatively distinct from simpler predictors.

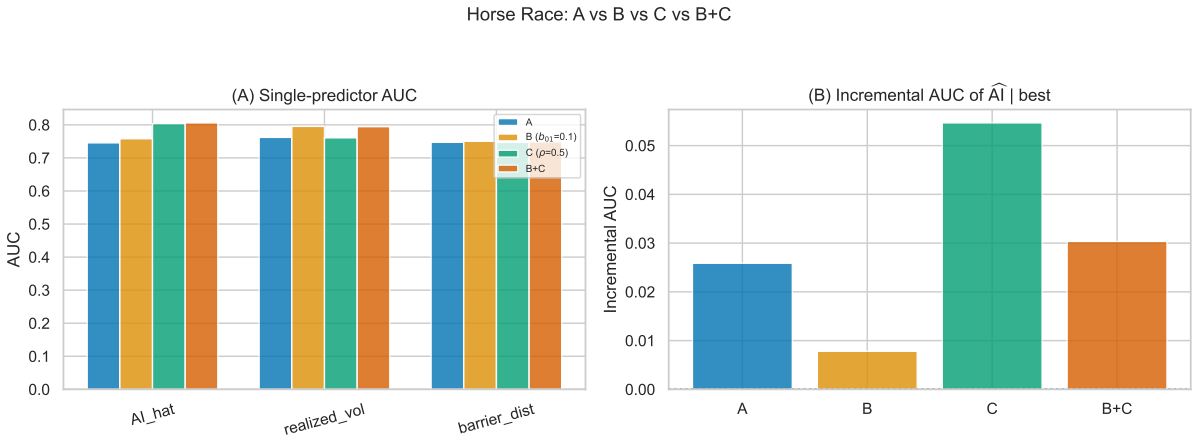

Figure 12: Horse race: Increments A (baseline), B ($b_{01} = 0.1$, $\rho = 0$), C ($\rho = 0.5$, $b_{01} = 0$), and B+C ($\rho = 0.5$, $b_{01} = 0.1$).

7 Discussion, Empirical Status, and Implementation Roadmap

This section addresses modelling scope, empirical predictions, and deployment of the framework.

7.1 Simulation architecture and econometric pipeline (overview)

The simulation and validation follow a strictly sequential pipeline with an explicit information firewall: fundamental quantities $(B^*, J^*, c^{Z,*})$ are simulated in an oracle world, the desk estimator sees only observed traded prices $\widetilde{Y}$, and oracle objects are consulted only after desk estimates are committed. Desk-side statistics comprise realized covariation (rolling returns), lead–lag covariances with matrix-logarithm generator recovery, over-identified moment ratios with a specification test, and block-bootstrapped AUC/ROC statistics. Full details (including the nine-step synthetic desk procedure, filtering stage, and daily monitoring pseudo-algorithm) are given in section B.

7.2 What this framework is, and what it is not

The framework operates one level above microstructure: it takes traded prices $\widetilde{Y}$ as given continuous semimartingales (theorem 2.1) and asks what can be said exactly about hedging error. This yields Layer 1 (theorem 2.7), an accounting identity that holds under any market organization. Layer 2 (theorem 3.2) adds a reduced-form equilibrium condition (analogous to a Walrasian clearing relation) specifying the outcome of aggregate hedging flow without modelling the adjustment mechanism. The clearing relation is kept underspecified beyond smoothness and invertibility, so the conjugation $c^Y = Jc^Z J^\top$ and its downstream consequences apply across a broad class of impact/flow mechanisms.

7.3 Modelling choices

We address here the specific modelling decisions that shape its generality.

The standard practitioner decomposition of hedging P&L writes losses in terms of Greeks (delta, gamma, vega, vanna, volga) and the changes in their underlying parameters (spot, implied vol, skew). This conflates at least three logically distinct operations: a change of coordinates from traded prices to implied parameters, an identification of hedge ratios with first-order sensitivities in those parameters, and an implicit assumption that the resulting strategy is self-financing. By working directly in traded coordinates, i.e. the prices at which hedges are actually executed, the present framework avoids all three conflations. The hedging-error identity (theorem 2.7) is

exact relative to the chosen desk operator: no assumption about what "vega" means, and no self-financing fiction for a strategy defined in non-traded coordinates.

A second key choice concerns the clearing assumption. The clearing relation (theorem 3.2) postulates that the market clears instantaneously: $\widetilde{Y}_t$ satisfies the fixed-point equation at each (t, ω). This structural hypothesis is economically grounded and is justified by two arguments. First, theorem 3.8 shows that instantaneous clearing is the $\tau \downarrow 0$ limit of a well-posed dynamic adjustment process; the approximation error is controlled by an $O(\tau)$ mean-square bound uniformly on finite horizons under the diffusion assumptions of that theorem. This reduces the instantaneous-clearing hypothesis to the assertion that the relaxation timescale τ is small relative to the hedging horizon, a qualitative claim that is plausible for liquid markets (seconds to minutes for major equity index options) and that can be assessed empirically from lead–lag covariances (theorem 4.1). Second, the instantaneous-clearing formulation allows the covariation conjugation to be stated as a clean algebraic identity, which is essential for the downstream loss bound, Lyapunov analysis, and Riccati filter. A fully dynamic model would require specifying the noise structure Γ and the relaxation timescale τ, neither of which is identified from contemporaneous covariation alone; this is precisely the identification gap that section 4.1 addresses.

Relatedly, the impact function $\Lambda : \mathbb{R}^m \to \mathbb{R}^{n+1}$ is not given an explicit functional form (e.g., linear, square-root, or power-law). Richer transient-impact models (Almgren–Chriss [28], Obizhaeva–Wang, Bouchaud–Farmer–Lillo) operate at a different granularity (individual order flow, temporary vs. permanent impact, cross-impact kernels) and require data on executed trades, which is not available to most desks in real time. The present framework requires only the *aggregate consequence* of impact: the Jacobian $D_Q\Lambda \cdot D_y\Phi = B$ (the loop gain), which determines the amplification matrix $J = (I - B)^{-1}$. This is a weaker data requirement, and it is sufficient for the covariation signatures that the framework predicts.

Finally, the framework is agnostic about the stochastic process driving fundamental covariation c^Z. No Heston, no SABR, no rough volatility, no local volatility surface is assumed. The only requirement is that $\widetilde{Z}$ (and hence $\widetilde{Y}$) be a continuous semimartingale with absolutely continuous quadratic variation (theorems 2.1 and 3.1). Every standard volatility model satisfies this condition, confirming it as broadly compatible with existing practice. The benefit is that the results are *robust* to model misspecification in the volatility dynamics: the identity of theorem 2.7 holds whether the desk uses Black–Scholes, a stochastic volatility model, or a data-driven machine learning surface.

Why Black–Scholes in simulation. Four considerations justify using Black–Scholes as the desk pricing function: (i) *identification*: a flat vol surface makes any price-model discrepancy attributable solely to feedback; (ii) *conservatism*: heterogeneous Greeks ($\nu(K_1) \neq \nu(K_2)$) already generate differential amplification without requiring strike-dependent vol dynamics (richer models would compound this); (iii) *algebraic invariance*: the conjugation, loss bound, and Lyapunov structure are model-agnostic; and (iv) *desk realism*: exotic desks typically price vanillas with BS. The fundamental vanilla price coordinate $P_t^Z = p_{\mathrm{BS}}(S_t, K, u_t, \tau_t)$ is generated from a latent OU surface-level factor u_t; since $\partial_u p_{\mathrm{BS}} > 0$, working in (S, P) coordinates is a smooth change of basis, and P^Z is a continuous semimartingale (the only structural input Layer 2 requires). Table 1 separates the three levels.

Level	**Definition**
Fundamental drivers	(S_t, u_t) (simulation choice; latent surface-level factor)
Fundamental traded coordinates	$\widetilde{Z}_t = (S_t,\, P_t^Z(K_1), \ldots)$ via BS map
Observed traded coordinates	$\widetilde{Y}_t = \widetilde{Z}_t + \Lambda\big(\Phi(t, \widetilde{Y}_t, I_t)\big)$ (clearing)

Table 1: World versus desk: the BS map connects the first two levels; the clearing map produces the third.

Two consequences merit emphasis. *First*, the residual $\mathfrak{r}_t$ and the covariation distortion μ_t play distinct roles in Layer 1: $\mathfrak{r}_t$ absorbs model/PDE misspecification and missing state variables (e.g.

the vol-of-vol premium that BS omits), while $\mu_t = c_t^Y - \widehat{c}_t$ isolates the gap between realized traded covariation and desk-implied covariation. If "BS is wrong for stochastic vol," this manifests primarily in $\mathfrak{r}_t$, not automatically in μ_t. *Second*, a scalar u_t implies a rank-1 cross-sectional volatility factor (all strikes share one instantaneous surface level), so any cross-strike shape emerging in the *observed* implied-volatility surface (e.g. endogenous skew between P_1 and P_2) cannot be attributed to exogenous skew dynamics inside the fundamental model; it must come from feedback via J. A full (K, T) surface can be accommodated either by selecting a finite set of liquid vanillas as traded coordinates, or by introducing a low-dimensional factor model for the surface (level/skew/curvature) and treating the resulting factor-linked vanilla prices as coordinates; the theory remains finite-dimensional and identical.

Worked example: separating $\mathfrak{r}$ from μ. Consider a scalar "curvature-only" product $v(t, y) = \frac{\alpha}{2}y^2 - \frac{\alpha}{2}\int_0^t \widehat{c}_s \, \mathrm{d}s$ with constant Hessian $H = \alpha$ and $\mathfrak{r}_t \equiv 0$ by construction. *Case A (pure pricing residual, $\mu = 0$):* if the desk forgets the Itô correction, it bleeds $-\frac{1}{2}\alpha\sigma_{\text{true}}^2 T$ deterministically despite correct covariation: a diffuse, fixable PDE error. *Case B (pure covariation distortion, $\mathfrak{r} = 0$):* if feedback amplifies realized covariation by j^2 ($c^Y = j^2\sigma_{\text{fund}}^2$, $\widehat{c} = \sigma_{\text{fund}}^2$), the loss $-\frac{1}{2}\alpha(j^2 - 1)\sigma_{\text{fund}}^2 T$ is structural and episode-concentrated; it cannot be eliminated by improving the pricing PDE.

In the 2D spot-put setting with $\beta_t = 0.8$ and $c_{SP}^Z = -0.6$, $\sigma_S^2 = 1.0$ ($\times 10^{-4}$), the conjugation gives $c_{SP}^Y = -0.6 + 0.8 = +0.2$ (**sign reversal**): the put moves *with* spot, a qualitative signature no PDE refinement can produce. Over a 10-day barrier-approach episode, the cumulative feedback loss is $\approx 6.4 \times 10^{-5}$ per unit notional squared, all attributable to μ with $\mathfrak{r}_t \approx 0$. All the relevant monitoring quantities (rolling covariance $\widehat{\Sigma}^Y$, $\hat{\beta}$ from the price-ratio estimator, fundamental stripping $\widehat{\Sigma}^Z = (I - \hat{B})\widehat{\Sigma}^Y(I - \hat{B})^\top$, the loss bound, rolling $\hat{\lambda}_1$, and Riccati gain) are computable from the desk's own book.

7.4 Empirical status and testable predictions

theorem 2.7 is a mathematical identity: given any product and pricing function satisfying theorems 2.1 and 2.5, the hedging error equals $-\frac{1}{2}\int H : \mu \, \mathrm{d}t + \int \mathfrak{r}_t \, \mathrm{d}t$ exactly.

By contrast, Layer 2 (the clearing hypothesis) *is* an empirical claim. It produces four testable covariation signatures:

(a) Lead–lag covariances are asymmetric: $C_{21}(s) \neq C_{12}(s)$ (theorem 4.1).
(b) AI_t spikes near barriers (concentrated hedging flow).
(c) Cumulative losses correlate with $\int \mathsf{AI}_t^2 \|H_t\| \mathrm{Tr}(c_t^Z) \, \mathrm{d}t$ (theorem 5.2).
(d) The rolling Lyapunov exponent $\hat{\lambda}_1$ turns positive during stress episodes (theorem 5.5).

All four predictions are testable from observable prices and product sensitivities without proprietary order-level data.

The externally validated simulation procedure provides an internal consistency check across all theoretical results; table 2 consolidates the main quantitative outcomes (and section B gives full details).

Figure 16 and fig. 19 show the full dashboard and robustness; see section B for details.

These simulation results demonstrate that the framework's predictions are internally consistent and numerically sharp, with the externally validated procedure providing falsifiable evidence under controlled ground truth. (As a basic check: across 5,000 simulated paths, the regression of realized hedging loss $(v_T - \widetilde{X}_T)$ on $\frac{1}{2}\int H : \mu \, \mathrm{d}t$ yields slope $+0.999$ and $R^2 = 0.985$, consistent with the exact accounting identity of theorem 2.7.) The framework improves on the standard Greek decomposition in three ways: it is coordinate-free, self-financing by construction, and provides a structural account of covariation distortion via $c^Y = Jc^Z J^\top$. In particular, the heuristic vanna-carry correction proposed by Salon [3] for autocallable hedging (an add-on payoff compensating the local-volatility model's failure to capture spot–surface covariation) is a special case of the $\frac{1}{2}H_t : \mu_t$ leakage term restricted to the off-diagonal (S, σ) block; the sign-reversal

Table 2: Simulation evidence for each theoretical result.

Result	Prediction	Simulation outcome
Hedging-loss identity (2.7)	slope $= +1$, $R^2 = 1$	slope $= +0.999$, $R^2 = 0.985$
Attenuation threshold (3.18)	ρ_Y reversal for $\beta > \beta^*$	detection rate 85 to 94%
Barrier Greeks (3.4)	vanna sign reversal	DIUO vanna: $+6.74$ to -3.56
Lead–lag ident. (4.1)	$A = s^{-1}\log(C\Sigma^{-1})$	numerical recovery error $< 10^{-13}$ in the OU benchmark
Loss bound (5.2)	$\lvert L_{\text{fb}}\rvert \propto \mathsf{AI}^2$	$37\times$ at $\mathsf{AI} = 6$
Parametric resonance (5.8)	$\lambda_1 > 0$ despite $m_t > 0$	$\epsilon^* \approx 0.881$; empirical AUC ≈ 1 in simulation
Riccati tradeoff (5.10)	$P^* \sim j^{-2}$, $\mathrm{Var}(l) \sim j^2$	P^* drops $38\times$; $\mathrm{Var}(l)$ rises $102\times$
AUC optimality (A.2)	increasing score in true AI is Bayes-optimal	incr. AUC $= +0.004$ over vol
Early warning (B)	pre-crisis alert	96.9% hit rate, 60-day lead
Conjugation test (3.11)	$\widehat{C}^Y_{UU} > 1$ under feedback	$1.0 \to 1.65$; $p < 10^{-23}$

mechanism of theorem 3.18 provides the structural explanation for the negative vanna carry that Salon documents empirically. The precise degeneration is stated in theorem B.1.

Empirical calibration on public equity-index data during known feedback episodes would populate the identity with estimated H_t and μ_t and test the predicted covariation signatures against realized data; we view this as a natural next step.

7.5 Implementation roadmap

For practitioners seeking to deploy the framework, the theoretical results translate into a four-stage procedure:

Stage 1. Traded-coordinate selection (section 2.1). Identify the liquid hedging instruments $Y = (Y^0, \ldots, Y^n)$ and compute discounted coordinates $\widetilde{Y}$.

Stage 2. Covariation estimation and feedback detection (sections 3 and 4.1). Estimate c^Y, c^Z (or proxies), and the lead–lag covariance $C(s)$ to identify the generator A and stability margin m.

Stage 3. Instantaneous diagnostics (section 3.3). Compute AI_t, the amplified Hessian $\widetilde{H}_t$, and the feedback-induced covariation distortion μ_t.

Stage 4. Cumulative risk assessment (section 5). Evaluate the pathwise loss bound, estimate the Lyapunov exponent $\hat{\lambda}_1$, and run the Riccati filter.

The complete implementation, including a nine-step synthetic desk procedure, a daily monitoring pseudo-algorithm, and externally validated results, is given in section B.

8 Conclusion

We develop a traded-coordinate framework for analyzing, identifying, and monitoring hedging-induced feedback risk in structured-product markets. The four main contributions, stated in the Introduction (section 1), are: the hedging-error identity (theorem 2.7), the covariation conjugation and amplification index (theorems 3.11 and 3.14), the identification layer built on lead–lag and signature moments (section 4), and the cumulative-risk toolkit benchmarked by the procedure in section B. In the synthetic desk environment studied here, the multi-strike extension (section 6.1) adds a third coordinate and raises the incremental predictive contribution of $\widehat{\mathsf{AI}}$ (AUC: $+0.026 \to +0.057$).

The framework is formulated for continuous semimartingale prices; the second-order jump remainder of theorem 2.10 extends the identity to settings with price jumps. The clearing hypothesis is a reduced-form equilibrium condition that captures the aggregate consequence of hedging flow through the Jacobian B and amplification matrix $J = (I - B)^{-1}$, without requiring

an explicit functional form for Λ; this generality allows the conjugation and downstream results to apply across a broad class of impact mechanisms. The lead–lag identification rests on an Ornstein–Uhlenbeck linearization that provides a tractable benchmark; the exact identities of Layers 1 and 2 remain valid under general semimartingale dynamics and the OU diagnostic degrades gracefully as a local approximation when stationarity fails. The single-product clearing framework extends naturally to multi-product settings through a vector-valued clearing equation with a block-structured loop-gain matrix, as the three-coordinate extension of section 6.1 already illustrates. The externally validated synthetic-desk procedure provides controlled, falsifiable evidence for the framework's predictions; empirical calibration on public equity-index data during known feedback episodes is a natural next validation step.

These considerations point to several natural extensions of the framework. Empirical calibration on public equity-index data during known feedback episodes (Nikkei 2012, KOSPI 2018 to 2019, 0DTE options) would populate the identity with estimated H_t and μ_t and test the predicted covariation signatures against realized data. Multi-product clearing, in which several exotic books hedge through a common set of vanilla instruments, would require a vector-valued clearing equation and a block-structured loop-gain matrix; the three-coordinate extension of section 6.1 is a first step in that direction and shows that even one additional coordinate can matter materially. Adding discrete hedging, transaction costs, and market-impact decay would connect the framework more directly to the Almgren–Chriss execution literature and to the limit-order-book granularity that the present approach deliberately abstracts away from. The jump-diffusion extension, which adds a second-order jump remainder to the identity, would broaden applicability to markets with discontinuous price dynamics, such as credit derivatives and event-driven equity products. On the theoretical side, the present work provides an instantaneous norm ($\mathsf{AI}_t = \|J_t\|_2$) and an asymptotic growth rate (λ_1) as feedback-risk measures. Extending these to *finite-time transient amplification* driven by the non-normality of J and the *non-commutativity of successive generators* in the propagator $\Psi = \prod e^{\mathcal{A}_k \Delta t}$ is a natural theoretical extension. Developing risk indices based on resolvent norms or propagator-level bounds, and structural resonance conditions linked to generator non-commutation, would sharpen the connection between the hedging-loss integral and the finite-time growth of Ψ.

Acknowledgements

Omitted for anonymous submission.

A Technical Supplements

A.1 AUC justification for amplification-based early warning

The AUC claim rests on a clean separation of roles: events are defined from forward P&L (latent), and the amplification index enters only as a predictor, never as part of the event definition.

Assumption A.1 (Monotone drift and independent noise)**.** Let b_t be latent and define true instability $\mathsf{AI}_t^{\text{true}} := \psi\big(\max(b_{1,t}, b_{2,t})\big)$, where ψ is strictly increasing (e.g. $\psi(x) = 1/(1-x)$). Consider a forward-horizon loss score $L_t := -\sum_{s=t}^{t+H-1} \Delta\Pi_s$, where the P&L increments satisfy

$$\Delta\Pi_s = -\kappa\, \phi(\mathsf{AI}_s^{\text{true}})\, \Delta t + \sigma \sqrt{\Delta t}\, \xi_s, \tag{27}$$

with $\kappa > 0$, $\sigma > 0$, ϕ strictly increasing, and ξ_s i.i.d. with a continuous distribution independent of $\mathcal{F}_{s-1}$. Define the event indicator $E_t := \mathbf{1}\{L_t \geq q_\alpha\}$, where q_α is the α-quantile of L_t (e.g. $\alpha = 0.99$ for worst 1% losses).

Proposition A.2 (Monotone optimal ranking under the location-shift model)**.** *Under theorem A.1, assume* $\mathsf{AI}_s^{\text{true}}$ *is approximately constant over the horizon* H*, so that*

$$L_t = \kappa H \Delta t\, \phi(\mathsf{AI}_t^{\text{true}}) + \sigma\sqrt{H\Delta t}\,\eta_t, \tag{28}$$

where η_t *has a continuous distribution independent of* $\mathsf{AI}_t^{\text{true}}$*. Then:*

(i) The conditional law of L_t *given* $\mathsf{AI}_t^{\text{true}} = a$ *is a location shift whose center is strictly increasing in* a*. Consequently, for every threshold* q*,*

$$a \mapsto \mathbb{P}(L_t \geq q \mid \mathsf{AI}_t^{\text{true}} = a)$$

is increasing. In particular, for the event

$$E_t = \mathbf{1}\{L_t \geq q_\alpha\},$$

the posterior event probability

$$\mathbb{P}(E_t = 1 \mid \mathsf{AI}_t^{\text{true}} = a)$$

is increasing in a*. Therefore any strictly increasing function of* $\mathsf{AI}_t^{\text{true}}$ *is Bayes-optimal for ranking events.*

(ii) Consider a sequence of models indexed by n *such that*

$$L_t^{(n)} = \kappa_n H_n \Delta t_n\, \phi(\mathsf{AI}_t^{\text{true}}) + \sigma_n\sqrt{H_n\Delta t_n}\,\eta_t.$$

If

$$\frac{\kappa_n\sqrt{H_n\Delta t_n}}{\sigma_n}\,\mathrm{sd}\big(\phi(\mathsf{AI}_t^{\text{true}})\big) \longrightarrow \infty,$$

then the pairwise ranking error probability of any strictly increasing score based on $\mathsf{AI}_t^{\text{true}}$ *tends to zero, and its AUC tends to* 1*.*

Proof. Under (28), conditioning on $\mathsf{AI}_t^{\text{true}} = a$ yields

$$L_t = m(a) + \sigma\sqrt{H\Delta t}\,\eta_t, \qquad m(a) := \kappa H \Delta t\, \phi(a),$$

with m strictly increasing because ϕ is. Hence for every threshold q,

$$\mathbb{P}(L_t \geq q \mid \mathsf{AI}_t^{\text{true}} = a) = \mathbb{P}\left(\eta_t \geq \frac{q - m(a)}{\sigma\sqrt{H\Delta t}}\right),$$

which is increasing in a because the distribution function of η_t is continuous and $m(a)$ is increasing. Therefore the posterior event probability is increasing in a, and any strictly increasing transformation of $\mathsf{AI}_t^{\text{true}}$ yields the same optimal ranking.

For part (ii), take two independent draws (a, η) and (a', η') with $a = \mathsf{AI}_t^{\text{true}}$, $a' = \mathsf{AI}_{t'}^{\text{true}}$. The pairwise ranking of the score $S = \phi(a)$ is correct whenever $L > L'$ iff $\phi(a) > \phi(a')$, i.e.

$$\kappa_n H_n \Delta t_n\big(\phi(a) - \phi(a')\big) > \sigma_n\sqrt{H_n\Delta t_n}\,(\eta' - \eta).$$

Dividing by $\sigma_n\sqrt{H_n\Delta t_n}$ and setting $\Delta\phi := \phi(a) - \phi(a')$, the reversal probability conditional on $\Delta\phi > 0$ is

$$\mathbb{P}\Big(\eta' - \eta > \tfrac{\kappa_n\sqrt{H_n\Delta t_n}}{\sigma_n}\,\Delta\phi\Big).$$

Since $\eta' - \eta$ has a fixed continuous distribution, this probability tends to zero pointwise in $\Delta\phi > 0$ as $\kappa_n\sqrt{H_n\Delta t_n}/\sigma_n \to \infty$. By dominated convergence (the reversal probability is bounded by 1), the unconditional pairwise reversal probability tends to zero. Hence $\mathrm{AUC} = \mathbb{P}(L > L' \mid \phi(a) > \phi(a')) \to 1$. □

A.2 Riccati benchmark context

Remark A.3 (Structural precedents for the estimation-loss coupling). The Riccati ODE (23) is a standard Kalman–Bucy estimator Riccati [31]; the j^2 net scaling of $\mathrm{Var}(l_t)$ follows from elementary algebra once the steady-state asymptotics are substituted. What is structurally distinctive is that the observation gain and the loss sensitivity are *locked to the same parameter*: in the observation model $\widehat{c}_t^Y = j_t^2\sigma_t^2 + \varepsilon_t$ the gain is j_t^2, and in the loss rate $l_t = -\frac{1}{2}h_t j_t^2(\sigma_t^2 - \widehat{\sigma}_t^2)$ the sensitivity is also j_t^2. In a generic Kalman–Bucy or LQG problem these are independent design parameters; here, both arise from the same amplification matrix $J = (I - B)^{-1}$ because the desk observes traded covariation and incurs losses in traded coordinates.

Three precedents in the control-theory literature illuminate different facets of this coupling. (i) *Bode's sensitivity integral* [32]: in SISO feedback, improving sensitivity at one frequency worsens it elsewhere. The j^2 result has the same conserved-tradeoff flavor, but in the amplification domain rather than the frequency domain. (ii) *Risk-sensitive filtering* [33]: the $\mathcal{H}_\infty$ estimator Riccati introduces a parameter that couples estimation precision to the cost functional. This is the closest algebraic precedent, except that the coupling parameter there is chosen by the designer, whereas j^2 here is imposed by the feedback equilibrium. (iii) *Dual control* [34, 35]: the control action affects both the state trajectory and the information available for estimation. In the present setting the gradient hedge is fixed rather than optimized, so the analogous coupling operates passively rather than through control design.

In a standard LQG problem the separation principle yields two Riccati equations, one for filtering and one for control. Here only the filter Riccati appears because the hedge is not a free control variable. The benchmark therefore isolates an estimation-loss coupling under a fixed hedge, not a full optimal-control problem in traded-coordinate space.

Remark A.4 (General oracle score). When $\mathsf{AI}_s^{\mathrm{true}}$ varies appreciably over the horizon $[t, t+H)$, the Bayes-optimal ranking score is any strictly increasing function of the cumulative drift

$$D_t \;:=\; \sum_{s=t}^{t+H-1} \phi(\mathsf{AI}_s^{\mathrm{true}})\,\Delta t,$$

rather than $\mathsf{AI}_t^{\mathrm{true}}$ alone. The horizon-constancy approximation in theorem A.2 replaces D_t by $H\Delta t\,\phi(\mathsf{AI}_t^{\mathrm{true}})$, which is justified when the horizon H is short relative to the time scale on which AI varies.

Proposition A.5 (AUC stability under score perturbation)**.** *Let*

$$S_t := \mathsf{AI}_t^{\mathrm{true}}, \qquad \widehat{S}_t := S_t + \delta_t.$$

Assume that for two independent copies (S, δ) and (S', δ'),

$$\mathbb{P}\big(|S - S'| \le 2\varepsilon\big) \le r(\varepsilon), \qquad \mathbb{P}\big(|\delta| > \varepsilon\big) \le p(\varepsilon)$$

for some functions $r, p : [0,\infty) \to [0,1]$. Then

$$|\mathrm{AUC}(\widehat{S}) - \mathrm{AUC}(S)| \le r(\varepsilon) + 2p(\varepsilon).$$

In particular, if $r(\varepsilon) \to 0$ and $p(\varepsilon) \to 0$ along a sequence of models, then the AUC of $\widehat{S}$ converges to that of S. Under the high-separation regime of theorem A.2, if the perturbation is asymptotically smaller than the pairwise score separation, then $\widehat{S}$ also attains AUC tending to 1.

Proof. A perturbation can change the pairwise ordering between two scores only if either the original margin is small, $|S - S'| \le 2\varepsilon$, or at least one perturbation exceeds ε in magnitude: $|\delta| > \varepsilon$ or $|\delta'| > \varepsilon$. By a union bound, the probability of a pairwise ordering change is at most

$$\mathbb{P}(|S - S'| \le 2\varepsilon) + \mathbb{P}(|\delta| > \varepsilon) + \mathbb{P}(|\delta'| > \varepsilon) = r(\varepsilon) + 2p(\varepsilon).$$

Since AUC is a pairwise concordance probability, the same bound controls $|\mathrm{AUC}(\widehat{S}) - \mathrm{AUC}(S)|$. □

Remark A.6 (Dependence and inference for reported AUCs). The oracle ranking result is population-level and does not require i.i.d. samples, but empirical AUC estimates and especially *incremental* AUC differences require care in time series: overlapping forward horizons L_t and rolling-window predictors create strong dependence. In practice one should (i) evaluate on non-overlapping event windows (subsample t by steps of H), or (ii) use a block bootstrap for confidence intervals, or (iii) apply a robust ROC/AUC variance estimator designed for dependent data. Small incremental AUCs (e.g. +0.004) should be reported with uncertainty bands and sensitivity to the baseline-selection protocol (avoid choosing the "best baseline" on the same test set).

A.3 Signature isolation: sufficient conditions

We give an explicit algebraic sufficient condition ensuring each channel's signature equations are uncontaminated by the other channel.

Lemma A.7 (Zero-contamination via support restrictions)**.** *Work under the increment model* (13) *with rank-one channels* (29) *and define the signature index set* $S_1 := \{i_1, j_1, k_1\}$ *for channel* 1 *and* $S_2 := \{i_2, j_2, k_2\}$ *for channel* 2*.*

$$K_1 = e_{i_1} e_{j_1}^\top, \qquad K_2 = e_{i_2} e_{j_2}^\top. \tag{29}$$

Fix a time t *and suppose* Σ_t^Z *is any symmetric matrix. If the following three conditions hold:*

(*C1*) (Disjointness) $i_2 \notin S_1$.

(*C2*) (Zero j_2-coupling) $\Sigma^Z_{a,j_2}(t) = 0$ *for all* $a \in S_1$.

(*C3*) (Zero i_2-coupling) $\Sigma^Z_{a,i_2}(t) = 0$ *for all* $a \in S_1$.

Then for every pair $(p, q) \in \{(i_1, j_1), (i_1, k_1)\}$,

$$\Sigma^Y_{pq}(t) = \Sigma^Z_{pq} + b_{1,t}\, \Sigma^Z_{j_1 q},$$

and in particular $\Sigma^Y_{pq}(t)$ *is independent of* $b_{2,t}$*. The analogous statement holds with indices swapped (channel* 2 *independent of* $b_{1,t}$*) under symmetric conditions.*

Proof. Write $J = I + b_1\, e_{i_1} e_{j_1}^\top + b_2\, e_{i_2} e_{j_2}^\top$. Expanding $J\,\Sigma^Z J^\top$ gives nine terms:

$$\begin{aligned} J\,\Sigma^Z J^\top = \Sigma^Z &+ b_1\, e_{i_1} e_{j_1}^\top \Sigma^Z + b_2\, e_{i_2} e_{j_2}^\top \Sigma^Z + b_1\, \Sigma^Z e_{j_1} e_{i_1}^\top + b_2\, \Sigma^Z e_{j_2} e_{i_2}^\top \\ &+ b_1^2\, e_{i_1} e_{j_1}^\top \Sigma^Z e_{j_1} e_{i_1}^\top + b_2^2\, e_{i_2} e_{j_2}^\top \Sigma^Z e_{j_2} e_{i_2}^\top \\ &+ b_1 b_2\, e_{i_1} e_{j_1}^\top \Sigma^Z e_{j_2} e_{i_2}^\top + b_1 b_2\, e_{i_2} e_{j_2}^\top \Sigma^Z e_{j_1} e_{i_1}^\top. \end{aligned}$$

Take $(p, q) = (i_1, j_1)$. Using $e_a^\top e_b = \mathbf{1}_{a=b}$:

Base term: $\Sigma^Z_{i_1 j_1}$.

Linear b_1 *terms:* $+b_1\, e_{i_1}^\top e_{i_1} e_{j_1}^\top \Sigma^Z e_{j_1} = +b_1\, \Sigma^Z_{j_1 j_1}$. The other linear b_1 term is

$$+b_1\, e_{i_1}^\top \Sigma^Z e_{j_1} e_{i_1}^\top e_{j_1} = +b_1\, \Sigma^Z_{i_1 j_1}\, \mathbf{1}_{i_1 = j_1} = 0$$

(since $i_1 \neq j_1$ in any meaningful feedback channel).

Linear b_2 *terms:* $+b_2\, e_{i_1}^\top e_{i_2} e_{j_2}^\top \Sigma^Z e_{j_1} = +b_2\, \mathbf{1}_{i_1 = i_2}\, \Sigma^Z_{j_2 j_1} = 0$ because $i_2 \notin S_1$ implies $i_2 \neq i_1$ by (*C1*). The other linear b_2 term: $+b_2\, e_{i_1}^\top \Sigma^Z e_{j_2} e_{i_2}^\top e_{j_1} = +b_2\, \Sigma^Z_{i_1 j_2}\, \mathbf{1}_{i_2 = j_1}$. If $i_2 = j_1$ this becomes $+b_2\, \Sigma^Z_{i_1 j_2}$, which vanishes under $\Sigma^Z_{i_1 j_2} = 0$. Condition (*C2*) enforces $\Sigma^Z_{a,j_2} = 0$ for all $a \in S_1$, in particular $a = i_1$, so this term is 0.

Quadratic b_2^2 *term:* $b_2^2\, \mathbf{1}_{i_1 = i_2}\, \mathbf{1}_{i_2 = j_1}\, \Sigma^Z_{j_2 j_2} = 0$ since $i_2 \neq i_1$.

Cross terms $b_1 b_2$*:* $b_1 b_2 \, \Sigma^Z_{j_1 j_2} \, \mathbf{1}_{i_2 = j_1} = 0$ under $\Sigma^Z_{j_1 j_2} = 0$ (condition (*C2*), take $a = j_1 \in S_1$). The other cross term: $b_1 b_2 \, \mathbf{1}_{i_1 = i_2} \, \Sigma^Z_{j_2 j_1} \, \mathbf{1}_{i_1 = j_1} = 0$.

Thus all b_2 contributions vanish and $\Sigma^Y_{i_1 j_1} = \Sigma^Z_{i_1 j_1} + b_{1,t} \, \Sigma^Z_{j_1 j_1}$. The same computation with $q = k_1$ gives $\Sigma^Y_{i_1 k_1} = \Sigma^Z_{i_1 k_1} + b_{1,t} \, \Sigma^Z_{j_1 k_1}$. □

Assumption A.8 (Block-variation with invariant signature set)**.** Partition indices into $S \cup R = \{1, \ldots, d\}$ and assume:

(i) $\Sigma^Z_t|_{S \times S}$ and $\Sigma^Z_t|_{S \times R}$ are constant in t,

(ii) all time variation is confined to $\Sigma^Z_t|_{R \times R}$.

Then conditions (*C2*)–(*C3*) can be enforced by choosing $j_2, i_2 \in R$ and $S_1 \subset S$ (or by setting the appropriate $S \times R$ entries to 0). This yields signature isolation for all t.

Remark A.9 (Testable versus assumed). The following are *testable* from observed data (given a covariance estimator): the over-identification residuals $\mathcal{T}_1, \mathcal{T}_2$ in (17) and the stability of signature moments. The following are *assumed*: that there exists a signature set unaffected by Σ^Z_t variation (a modelling choice). The following is a *hypothesis to stress-test*: monotone dependence of loss drift on instability (theorem A.1); in real data, this becomes an empirical question: does estimated AI forecast forward P&L tail losses?

Remark A.10 (Design recipe for signature selection). A mechanically checkable procedure: (i) Choose a signature set S where certain covariances are stable (or can be pinned down by long-window estimates). (ii) Choose $K_1 = e_{i_1} e_{j_1}^\top$ and pick $k_1 \in S$ with $\Sigma^Z_{j_1 k_1} \neq 0$. (iii) Choose channel-2 indices i_2, j_2 outside the channel-1 signature set, and enforce $\Sigma^Z_{S_1, j_2} = 0$ and $\Sigma^Z_{S_1, i_2} = 0$ (or, in block terms, place $i_2, j_2 \in R$ and $S_1 \subset S$). (iv) Repeat symmetrically to isolate channel 2. Then theorem 4.7 applies with no cross-channel contamination.

B Practitioner Procedure, Algorithm, and Detailed Results

This appendix translates the theoretical framework into a concrete implementation. We describe the estimation procedure, the monitoring workflow, and a pseudo-algorithm for daily computation.

B.1 Estimation from observables via filtering

In practice the latent objects driving amplification, most notably the loop-gain matrix B_t and the amplification matrix $J_t = (I - B_t)^{-1}$, are not directly observed. A desk typically observes $\widetilde{Y}$ (market-traded coordinates, which already embed feedback), constructs realized covariation estimates $\widehat{\Sigma}^Y_t$ from high-frequency or intraday data, and has access to covariates/features (order-flow proxies, liquidity indicators, barrier proximity, and contract state ξ_t). A parsimonious and stable approach is to parameterize $B_t = B(\theta_t, \mathrm{feat}_t, \xi_t)$ by a low-dimensional latent parameter θ_t evolving slowly (e.g. random walk or OU) and to fit θ_t sequentially in a state-space model.

A robust choice for online estimation is a Student-t (or Huber) nonlinear filter (UKF/EKF) for θ_t combined with RTS smoothing for offline analysis. For covariance inputs one may filter a low-rank factorization of Σ^Z_t (the latent fundamental covariation) with $\widehat{\Sigma}^Y_t$ (realized covariance from observed returns) as noisy measurement. The output is a time series $\widehat{B}_t$, $\widehat{J}_t$, and $\widehat{\mathsf{AI}}_t = \|\widehat{J}_t\|$ with uncertainty bands.

B.2 Synthetic desk procedure (econometrics-grade, externally validated)

When proprietary market data are unavailable, the framework remains practically testable via a *synthetic desk*, provided the experimental design maintains a strict separation between the data-generating process and the estimation targets so that the simulation does not simply reproduce

its own inputs. We therefore impose a strict separation between (i) the *true data-generating process* (DGP) and its latent quantities, (ii) the *desk information set* (what is observable), and (iii) the *desk model* (potentially misspecified) used for pricing, hedging, and estimation.

Pipeline overview (firewalled by design).

1. *DGP (oracle world).* Simulate fundamental paths $\widetilde{Z}$ and generate feedback-distorted traded prices $\widetilde{Y}$ via the clearing/impact map. Record oracle objects $B_t^*, J_t^*, c_t^{Z,*}, c_t^{Y,*}$.
2. *Information firewall.* Discard oracle objects and retain only the desk-observable time series $\{\widetilde{Y}_{t_k}\}$ (and any publicly observable contract state).
3. *Desk pricing/hedging.* Using only observable $\widetilde{Y}$, compute hedge ratios and hedging P&L; compute realized covariation estimates $\widehat{c}^{Y}$ from returns.
4. *Structural estimation.* From the same desk-observables, estimate $\widehat{B}$ (or $\widehat{J}$), form $\widehat{\mathsf{AI}} = \|\widehat{J}\|$, and, if $\widehat{J}$ is available, compute $\widehat{c}^{Z} = \widehat{J}^{-1}\widehat{c}^{Y}\widehat{J}^{-\top}$.
5. *Validation (oracle only after the fact).*
   (a) *Identity validation* (Layer 1, desk data only): verify the hedging-error identity numerically; this is tautological accounting and involves no oracle.
   (b) *Mechanism validation* (Layer 2, oracle comparison): compare $\widehat{B}, \widehat{J}, \widehat{c}^{Z}$ to oracle $B^*, J^*, c^{Z,*}$; report conjugation R^2 and $\mathrm{corr}(\widehat{\mathsf{AI}}, \mathsf{AI}^*)$.
6. *Horse race / AUC.* Using only desk-estimated predictors (e.g. $\widehat{\mathsf{AI}}$, realized vol, barrier distance), evaluate predictive ranking against *forward* tail-loss events.
7. *Stress grid.* Repeat Steps 1 to 4 across parameter grid points; Steps 5 to 6 are computed per run. Grid points are parallelizable.

Identification guarantee: everything downstream of the firewall uses only desk-observable $\widetilde{Y}$; oracle quantities enter *only* in Step 5(b) (mechanism validation, after all desk estimates are committed).

The procedure begins by simulating a latent fundamental state $X_t = (\widetilde{Z}_t, I_t, \theta_t)$ on $[0, T]$, where $\widetilde{Z}_t \in (0, \infty)^{n+1}$ are fundamental (counterfactual) traded coordinates, $I_t \in \mathcal{I}$ is a regime process (e.g. a Markov chain), and θ_t are slowly-varying impact parameters. A minimal but stress-capable specification is

$$\mathrm{d}\widetilde{Z}_t = \Sigma_Z(t, I_t)^{1/2}\, \mathrm{d}W_t, \qquad \theta_{t+\Delta} = \theta_t + \eta_t, \qquad I_t \sim \mathrm{Markov}(P),$$

Two DGP families are run. (1) *Continuous benchmark*: $\widetilde{Z}$ is a continuous diffusion (possibly with stochastic volatility), so the assumptions of theorems 2.1 and 3.1 hold and the conjugation (theorem 3.11) applies as stated. (2) *Jump robustness*: heavy tails or jumps are introduced by letting W be replaced by a Student-t innovation at discrete sampling times, or by adding a jump component to $\widetilde{Z}$. In this case the Layer 1 identity acquires the jump remainder of theorem 2.10, and diagnostics relying on the continuous-path conjugation are interpreted as approximations whose accuracy the simulation quantifies.

The observed (feedback-distorted) traded coordinates $\widetilde{Y}$ are defined by the clearing relation

$$\widetilde{Y}_t = \widetilde{Z}_t + \Lambda\big(\Phi(t, \widetilde{Y}_t, I_t; \theta_t)\big), \qquad t \in [0, T],$$

consistent with theorem 3.2. Market data are recorded at discrete sampling times (t_k) with microstructure noise:

$$\widetilde{Y}^{\mathrm{obs}}_{t_k} = \widetilde{Y}_{t_k} + \epsilon_{t_k},$$

where (ϵ_{t_k}) is mean-zero measurement noise (e.g. bid–ask bounce), and sampling is discrete (e.g. 1s–1m). Realized covariance estimators $\widehat{\Sigma}^{Y}$ are computed from $(\widetilde{Y}^{\mathrm{obs}}_{t_k})$ using microstructure-robust methods (pre-averaging, subsampling, or refresh-time synchronization). Choose (Φ, Λ) so that the loop gain $B_t = D_Q\Lambda\, D_y\Phi$ approaches the stability boundary $\rho(B_t) \uparrow 1$ in stressed regimes, producing realistically sharp amplification episodes. The canonical product is a down-in up-out (DIUO) put with lower knock-in barrier B_- and upper knock-out barrier B_+ (section 3.4), priced

analytically via Rubinstein–Reiner and Fourier sine spectral methods. The traded coordinate P_t remains a vanilla put, but the desk hedges the DIUO product using both spot (for delta) and vanilla puts (for vega), so that aggregate DIUO hedging demand flows through vanilla put markets and distorts their prices.

To enforce the information firewall, define the desk filtration

$$\mathcal{I}_t^{\text{desk}} := \sigma\big(\{\widetilde{Y}_s : s \le t\}, \{\text{feat}_s : s \le t\}, \xi_t\big),$$

where feat_t are observable covariates (liquidity proxies, barrier distance, order-flow proxies, etc.). All estimation and hedging decisions must be measurable w.r.t. $\mathcal{I}_t^{\text{desk}}$. In particular, the desk *never* observes the fundamental coordinates $\widetilde{Z}_t$, the loop gain B_t, the fundamental covariation c_t^Z, or the impact parameters θ_t.

The desk uses a price-taking covariation operator $\widehat{c}(t, y, \xi)$ calibrated on $\widetilde{Y}$ returns and a parametric loop-gain proxy $B(\hat{\theta}_t, \text{feat}_t, \xi_t)$, with deliberate misspecification such as: (i) wrong functional form for Φ or Λ, (ii) missing/latent regime I_t (only noisy proxies available), (iii) underestimation of correlation/volatility in stress. Compute hedges ϑ_t and curvature H_t from the desk valuation engine (PDE/MC/BSDE/tree), and track the residual term $\int \mathfrak{r}_t \, \mathrm{d}t$ from theorem 2.5 for products represented through a finite-dimensional traded-coordinate state.

From discretely sampled $\widetilde{Y}$, compute realized covariance estimates $\widehat{\Sigma}_t^Y$ on rolling windows. To avoid toy results, include at least one of: (a) subsampling, (b) pre-averaging, (c) refresh-time synchronization (for asynchronous coordinates), and report robustness across sampling frequencies. Treat $\widehat{\Sigma}_t^Y$ as a noisy measurement of the integrated covariance (which includes both fundamental and feedback-induced components).

Using only desk-observable quantities, fit a low-dimensional latent parameter process $\hat{\theta}_t$ (hence $\widehat{B}_t$ and $\widehat{J}_t$) using only $\mathcal{I}_t^{\text{desk}}$: a robust UKF/EKF with Student-t (or Huber) likelihood for outlier resistance, plus RTS smoothing for offline analysis. Compute the desk instability signal $\widehat{\mathsf{AI}}_t = \|(I - \widehat{B}_t)^{-1}\|$ with regularized inversion (eigenvalue clipping/ridge).

For evaluation, split the simulated timeline into a training window and a test window; refit/initialize only on the training window and evaluate strictly out-of-sample. Report two sets of quantities:

$$\text{Oracle (truth): } \mathsf{AI}_t^\star,\ B_t^\star,\ c_t^{Y,\star} \qquad \text{vs} \qquad \text{Desk (estimated): } \widehat{\mathsf{AI}}_t,\ \widehat{B}_t,\ \widehat{\Sigma}_t^Y.$$

Use oracle objects *only* for evaluation, never for estimation.

Define events mechanically, e.g. (A) loss events $L_{t,t+\Delta} \le -q$ for a fixed tail quantile q, (B) instability episodes $\rho(B_t^\star) > 1 - \varepsilon$, and (C) distortion episodes $\|\widehat{\Sigma}_t^Y - \widehat{c}(\cdot)\| > \kappa$. Evaluate: (i) attribution accuracy (correlation / R^2 between predicted leakage $\int \frac{1}{2} H_t : \widehat{\mu}_t \, \mathrm{d}t$ and realized hedging loss), (ii) early-warning AUC/ROC, precision@k, and average lead time, (iii) tail-loss reduction (CVaR_α) when acting on $\widehat{\mathsf{AI}}$. Run a horse race against strong baselines: realized volatility, vol-of-vol proxy, correlation spikes, barrier distance, and PCA factors of $\widehat{\Sigma}^Y$; report incremental AUC and likelihood-ratio improvements when adding $\widehat{\mathsf{AI}}$.

Finally, perform a factorial sweep over maturity/monitoring frequency, moneyness or barrier distance, noise level, regime dwell times, and maximum $\rho(B^\star)$. Repeat across at least three DGP families: Gaussian OU, heavy-tailed SV, and regime-switching correlation. Include at least one monitored or memory product (Asian with state augmentation, or autocall with memory) to quantify $\int \mathfrak{r}_t \, \mathrm{d}t$ and to show the feedback-induced term persists beyond the simplest Markovian vanilla setting once a sufficient finite-dimensional state is introduced.

A single execution script (i) simulates the DGP, (ii) computes $\widehat{\Sigma}^Y$, (iii) runs the filter, (iv) computes signals and hedging losses, and (v) regenerates every table and figure from fixed seeds and configuration parameters. All code is available in the supplementary material; this produces falsifiable, externally validated evidence in the same sense as an empirical econometrics study, but under controlled ground truth. Results from a complete execution of this procedure are reported in figs. 16 to 19.

B.3 Pseudo-algorithm: daily feedback monitoring

The following algorithm is executed once per trading day (or intraday at the desk's discretion). All matrix operations are in $\mathbb{R}^{(n+1)\times(n+1)}$.

Algorithm: Feedback Risk Monitor

---

1. *Input:* Intraday tick data for $\widetilde{Y}_t$ on window $[t-W, t]$ of length W (e.g., 20 days).
2. *Estimate contemporaneous observed covariation:*

$$\widehat{\Sigma}^Y \;=\; \frac{1}{W}\sum_{k} \Delta\widetilde{Y}_k\, \Delta\widetilde{Y}_k^\top \qquad \text{(realized covariation estimator)}.$$

3. *Estimate lead–lag covariance* for lag $s > 0$:

$$\widehat{C}(s) \;=\; \frac{1}{N}\sum_{k=1}^{N} \Delta\widetilde{Y}_{k+\lfloor s/\delta \rfloor}\, \Delta\widetilde{Y}_k^\top.$$

4. *Recover generator:* If $\widehat{C}(s)\,(\widehat{\Sigma}^Y)^{-1}$ has all positive real eigenvalues, compute

$$\widehat{\mathcal{A}} \;=\; \frac{1}{s}\log\!\big(\widehat{C}(s)\,(\widehat{\Sigma}^Y)^{-1}\big).$$

5. *Extract diagnostics:*
   (a) Stability margin: $\widehat{m} = -\max_i \Re(\lambda_i(\widehat{\mathcal{A}}))$.
   (b) Eigenvector structure: identify directional channels (e.g., spot $\to$ surface vs. surface $\to$ spot).
   (c) Loop gain: $\widehat{B} = I + \tau\widehat{\mathcal{A}}$ (requires calibrating τ; alternatively use $\widehat{m}$ directly).
   (d) Amplification index: $\widehat{\mathsf{AI}}_t = \|(I-\widehat{B})^{-1}\|_2 = 1/\sigma_{\min}(I-\widehat{B})$.
6. *Imply fundamental covariation:* $\widehat{c}^Z = (I-\widehat{B})\,\widehat{\Sigma}^Y\,(I-\widehat{B})^\top$.
7. *Compute amplified Hessian:* Given the desk's model Hessian H_t, $\widetilde{H}_t = \widehat{J}_t^\top H_t \widehat{J}_t$.
8. *Pathwise loss bound:* $|\widehat{L}_{\mathrm{fb}}| \le \frac{1}{2}\int_{t-W}^{t} (\widehat{\mathsf{AI}}_u^{\,2}+1)\|H_u\|_2 \mathrm{Tr}(\widehat{c}_u^Z)\,\mathrm{d}u$, where $\widehat{\mathsf{AI}}_u$ and $\widehat{c}_u^Z$ denote the estimates from the rolling window ending at u; when a single window-level estimate is used, replace the integrand by $(\widehat{\mathsf{AI}}^{\,2}+1)\,\mathrm{Tr}(\widehat{c}^Z)\int_{t-W}^{t}\|H_u\|_2\,\mathrm{d}u$.
9. *Rolling Lyapunov exponent:* From the past N windows,

$$\widehat{\lambda}_1 \;=\; \frac{1}{N\Delta t}\log\Big\|\textstyle\prod_{k=1}^{N} e^{\widehat{\mathcal{A}}_{t_k}\Delta t}\Big\|.$$

10. *Riccati filter update:* If scalar, update P_t via $\dot{P}_t = -2\kappa P_t + \nu^2 - j_t^4 P_t^2/R$; compute Kalman gain $K_t = j_t^2 P_t/R$.
11. *Generate alerts:*
    (a) *Yellow:* $\widehat{m} < m_{\mathrm{warn}}$ or $\widehat{\mathsf{AI}} > 2$.
    (b) *Red:* $\widehat{\lambda}_1 > 0$ or $\widehat{\mathsf{AI}} > 4$.

B.4 Supplementary figures

The following figures complement the main-text analysis: a three-dimensional view of the Lyapunov stability boundary (fig. 13), a pedagogical early-warning dashboard (fig. 14), and a heatmap of estimation-loss variance across the amplification-noise plane (fig. 15).

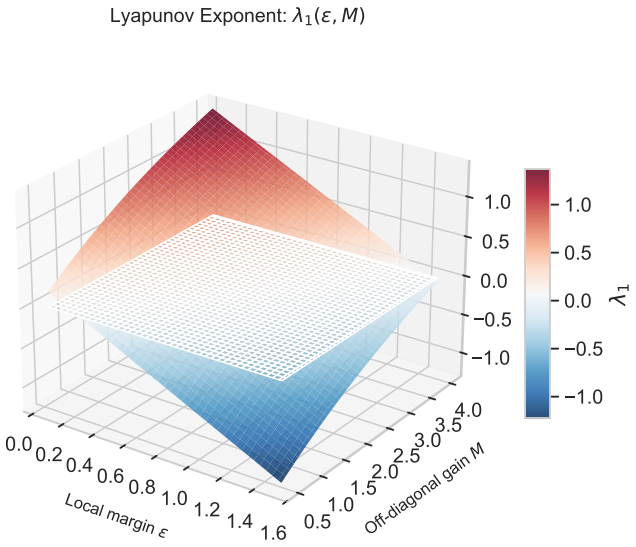

Figure 13: Lyapunov exponent $\lambda_1(\epsilon, M)$ for the two-generator model (theorem 5.8).

In fig. 13, blue: $\lambda_1 < 0$ (stable); red: $\lambda_1 > 0$ (unstable); gray: $\lambda_1 = 0$ (section 5.2). Figure 14:

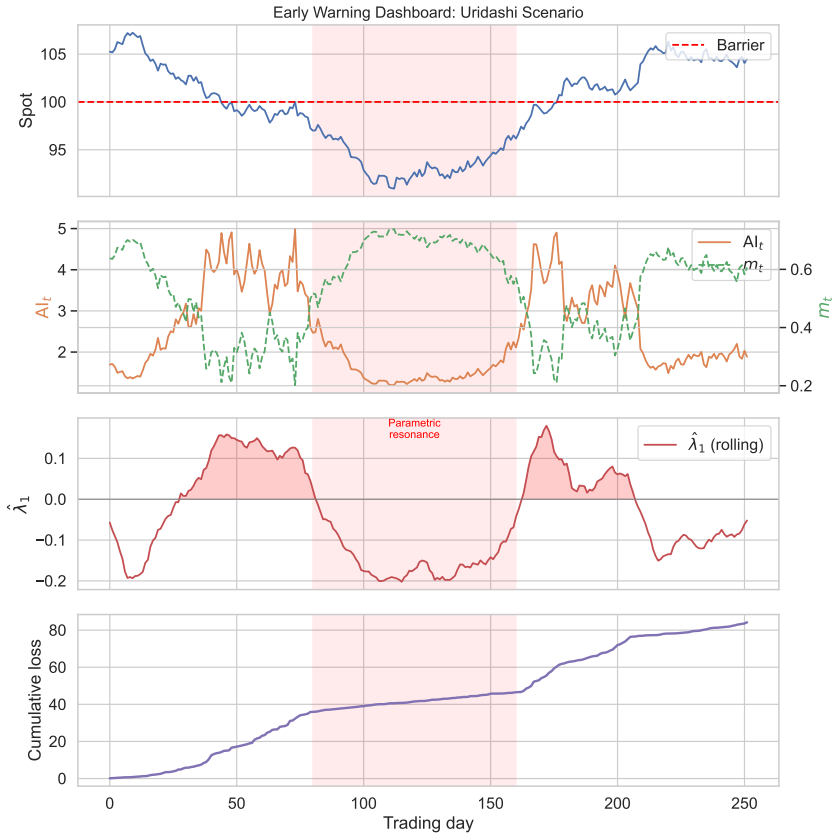

Figure 14: Early-warning dashboard for a 252-day Uridashi-like scenario.

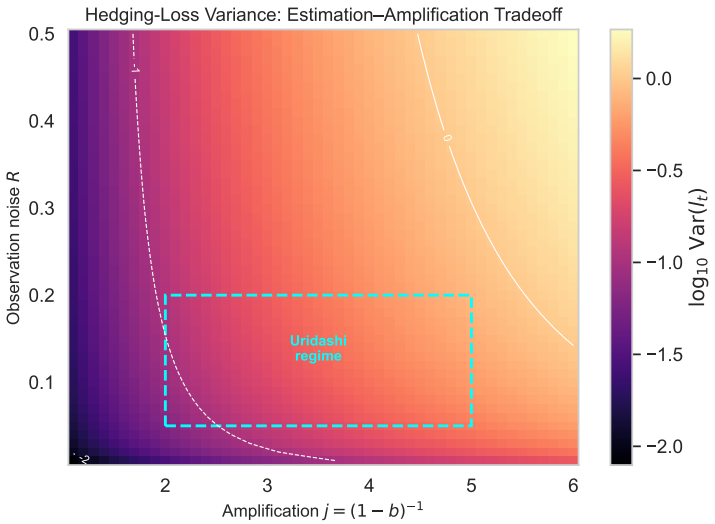

Figure 15: Heatmap of $\log_{10} \mathrm{Var}(l_t)$ over the (j, R) plane.

panels from top: spot vs barrier, amplification index and stability margin, rolling Lyapunov exponent, cumulative feedback loss (section 5.3). Figure 15: cyan box marks the simulated Uridashi regime ($j \approx 2$ to 5, $R \approx 0.05$ to 0.2).

B.5 Externally validated procedure results

Figures 16 to 19 report the full output of the externally validated simulation procedure described in section B.2. Figure 16 summarizes (oracle vs desk-estimated) amplification, estimation accuracy, and loss-bound validation. Figure 17 reports the predictor horse race via ROC/AUC. Figure 18 shows cumulative-risk diagnostics (Lyapunov, over-ID, and Riccati effects). Figure 19 reports robustness across feedback strength and observation noise.

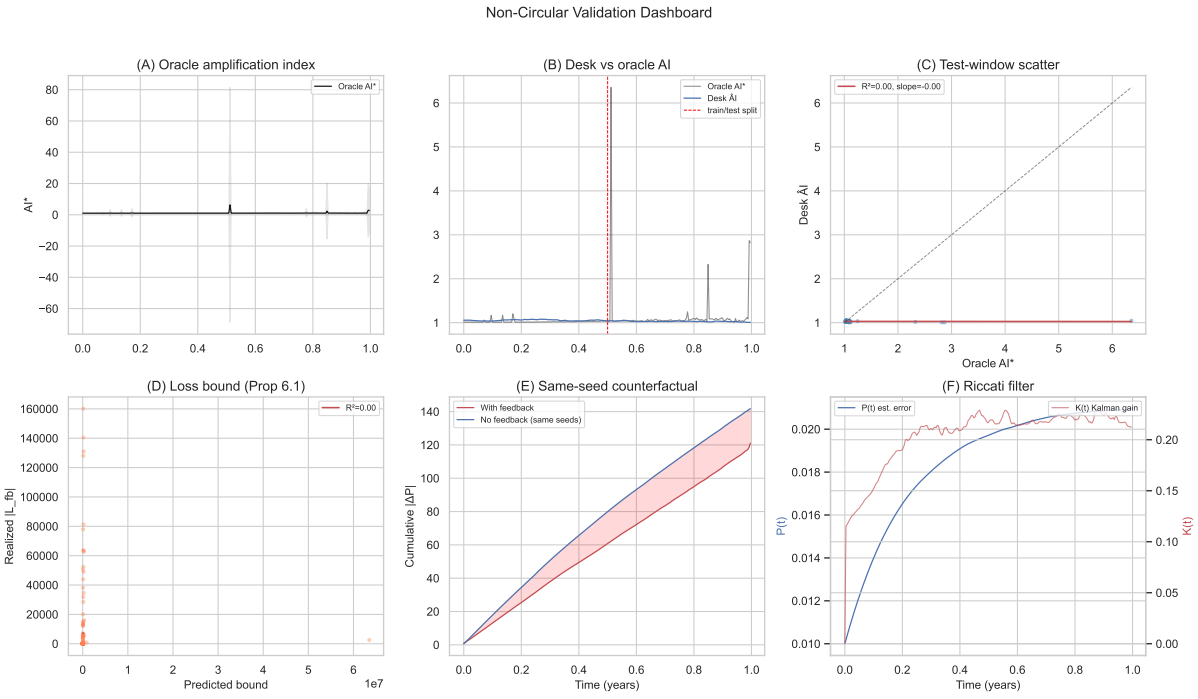

Figure 16: Externally validated simulation dashboard.

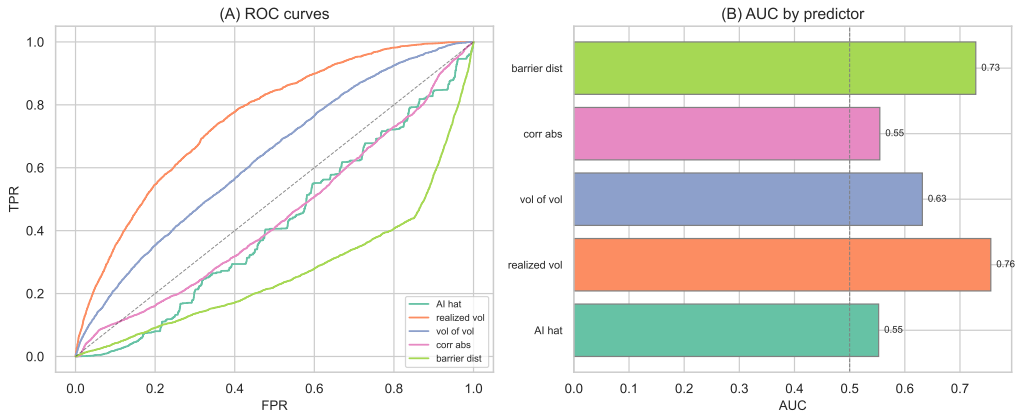

Figure 17: Horse race: ROC curves and AUCs for five tail-loss predictors.

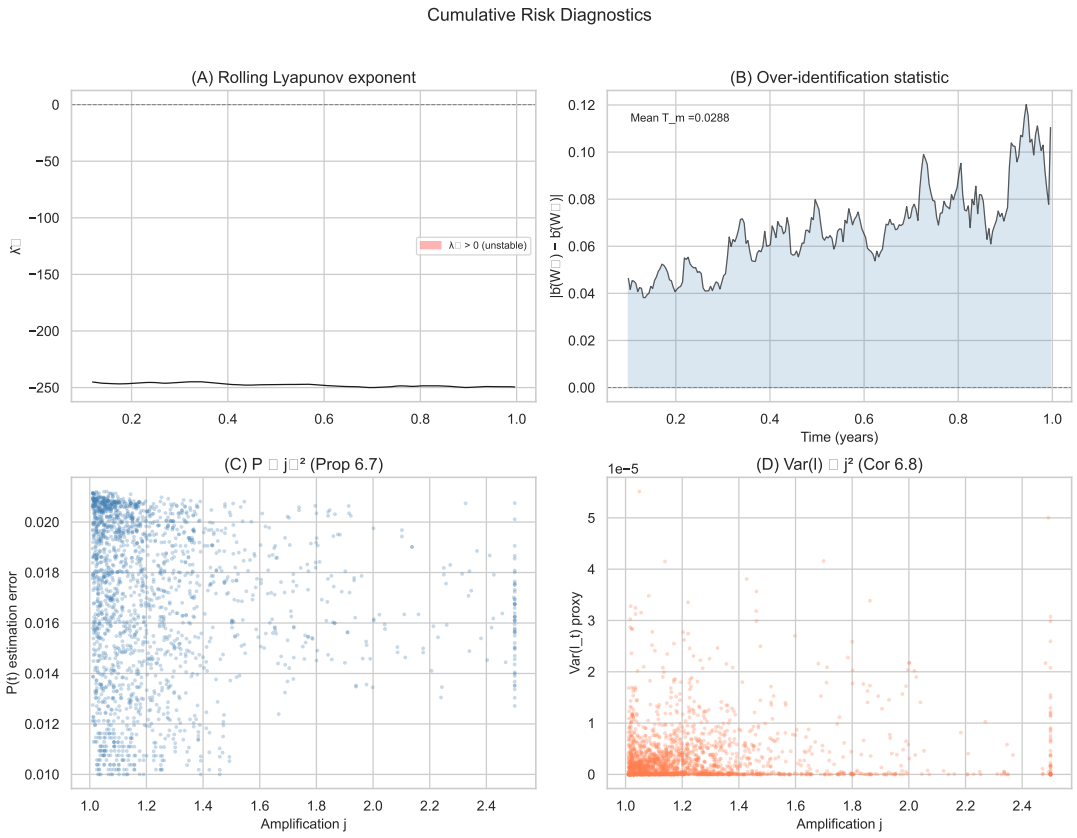

Figure 18: Cumulative risk diagnostics from desk-observable data.

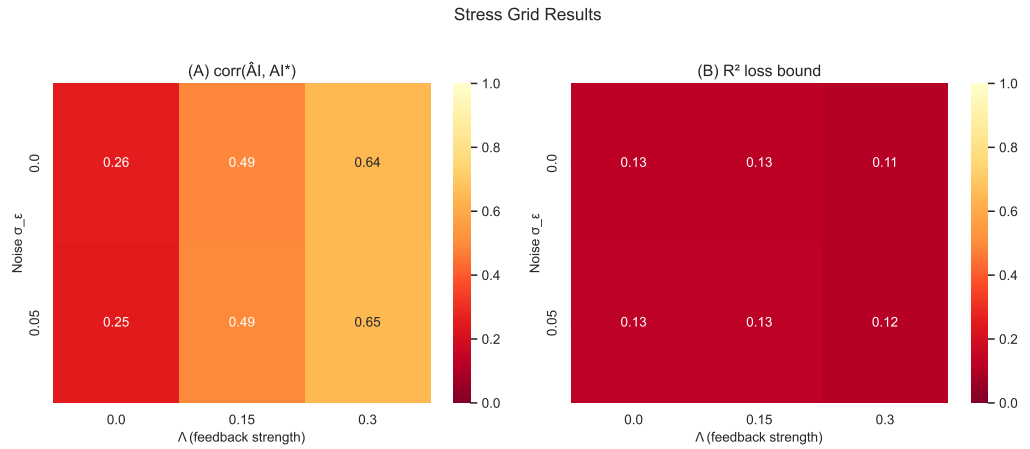

Figure 19: Stress grid: robustness over feedback strength and observation noise.

B.6 Illustrative applications

Before issuing a new autocallable tranche, compute the Hessian H of the prospective product at representative spot/vol levels and the desk's current aggregate flow sensitivity $D_y\Phi$. The amplified Hessian $\widetilde{H} = J^\top H J$ quantifies how much additional curvature the new tranche adds to the book *after* feedback. If $\|\widetilde{H}\|/\|H\| > K$ for a threshold K (e.g., $K = 4$, meaning a fourfold curvature amplification), the structurer can either reduce notional, shift barriers, or add a static hedge overlay.

For intraday monitoring, implement the pseudo-algorithm of section B.3 on a rolling window (e.g., $W = 20$ days, refreshed every 5 minutes). The key early-warning signals are:

- $\widehat{m} \downarrow 0$: feedback loop approaching instability.
- $\widehat{\mathsf{AI}} > 3$: curvature amplification exceeds $9\times$.
- $\widehat{\lambda}_1 > 0$: cumulative system is exponentially unstable despite instantaneous stability.

When red alerts trigger, natural responses include widening bid–ask spreads on the exotic book (reducing Φ), increasing the rebalancing frequency of the fundamental-variance estimator (increasing Kalman gain), and alerting risk management.

After a loss episode, decompose cumulative P&L using the exact identity of theorem 2.7:

$$\begin{aligned}\text{Total hedging loss} &= -\tfrac{1}{2}\int_0^T H_t : \mu_t \,\mathrm{d}t \\ &= \underbrace{-\tfrac{1}{2}\int H_t : (J_t c_t^Z J_t^\top - c_t^Z)\,\mathrm{d}t}_{\text{feedback-induced}} \;+\; \underbrace{-\tfrac{1}{2}\int H_t : (c_t^Z - \widehat{c}_t)\,\mathrm{d}t}_{\text{model mis-specification}}.\end{aligned}$$

This separates feedback-driven losses from pure model error, enabling targeted remediation.

The Lyapunov exponent can also be stressed by shocking M (off-diagonal gain) while holding ϵ (local margin) fixed. Figure 13 shows the $(\epsilon, M) \mapsto \lambda_1$ surface; a regulator or risk committee can identify the critical gain $M^*(\epsilon)$ at which the system transitions from globally stable to globally unstable.

B.7 Degeneration to the Salon vanna-carry correction

We show that the heuristic "ExtraVanna P&L" add-on of Salon [3] is recovered as a special case of the exact hedging-error identity (theorem 2.7) under four simplifying assumptions. We then identify the term that Salon's formula omits and explain why his correction is nonetheless sign-correct under the no-impact hypothesis.

Proposition B.1 (Degeneration to the Salon correction)**.** *Consider the two-channel traded system (S, U) of theorem 3.18, where S is the spot and $U = \sigma_{\text{atm}}$ is the at-the-money implied volatility. Let Π be a cancellable down-and-in put with the product decomposition*

$$\Pi(t, S_t) \;\approx\; \text{Put}_{\text{D\&I}}(t, S_t, T) \;\times\; \text{Surv}(t, S_t, T), \tag{30}$$

where $\text{Put}_{\text{D\&I}}$ is a vanilla-like down-and-in put and Surv is the no-recall survival probability. Impose the following assumptions:

(S1) Vanna dominance. *Only the off-diagonal (S, U) block of $H_t := D^2_{yy}\Pi$ contributes to the carry; diagonal terms (gamma, volga) are hedged or negligible.*

(S2) No hedge–price impact. *The feedback gain vanishes: $B_t = 0$, so $J_t = I$ and $c^Y_t = c^Z_t$. Observed covariation equals fundamental covariation.*

(S3) Product-form Hessian. *Via (30), the exotic vanna satisfies*

$$\frac{\partial^2 \Pi}{\partial S\, \partial \alpha_k} \;\approx\; \frac{\partial\, \text{Surv}}{\partial S}\, \frac{\partial\, \text{Put}_{\text{D\&I}}}{\partial \alpha_k},$$

after discarding two sub-dominant terms (the vanilla vanna × Surv, absorbed into the vanilla book's own P&L, and the mixed second derivative of Surv, numerically negligible).

(S4) Local-volatility benchmark. *The desk prices with a local-volatility model whose implied spot–vol covariance is $\widehat{c}_{SU,t} = \sigma_{\text{atm}}\,(\mathcal{S}_T + T^{-1}\int_0^T \bar{\sigma}^2(u)\mathcal{S}_u/(\hat{\sigma}_u\hat{\sigma}_T)\,\mathrm{d}u)$, following Bergomi [22], Ch. 1.*

Under (S1)–(S4), the exact hedging-error identity of theorem 2.7 reduces to

$$\widetilde{X}_T - \Pi_T \;=\; \widetilde{X}_0 - \Pi_0 \;-\; \int_0^T \mathfrak{r}_t\, \mathrm{d}t \;-\; \int_0^T \frac{\partial\, \text{Surv}}{\partial S}\, \frac{\partial\, \text{Put}_{\text{D\&I}}}{\partial \alpha_k}\, \big[(c^Z_t)_{S\alpha_k} - \widehat{c}_{S\alpha_k,t}\big]\, \mathrm{d}t, \tag{31}$$

where the last integral is precisely the Salon "ExtraVanna P&L" formula (his eq. 9).

Proof. Start from the exact identity (theorem 2.7):

$$\widetilde{X}_T - \Pi_T = \widetilde{X}_0 - \Pi_0 - \int_0^T \mathfrak{r}_t\, \mathrm{d}t - \tfrac{1}{2}\int_0^T H_t : \mu_t\, \mathrm{d}t.$$

The covariation-distortion density is $\mu_t = c^Y_t - \widehat{c}_t$. Under (S2), $c^Y_t = c^Z_t$, so $\mu_t = c^Z_t - \widehat{c}_t$. Under (S1), the Frobenius product $H_t : \mu_t$ reduces to the off-diagonal block:

$$H_t : \mu_t \;\approx\; 2\,\frac{\partial^2 \Pi}{\partial S\, \partial \alpha_k}\, \big[(c^Z_t)_{S\alpha_k} - \widehat{c}_{S\alpha_k,t}\big],$$

where the factor of 2 accounts for both off-diagonal entries of the symmetric matrices, cancelling the $\frac{1}{2}$ prefactor. Under (S3), substitute the product-form vanna approximation. Under (S4), $\widehat{c}_{S\alpha_k,t}$ is the local-vol model covariance (which is nonzero only for $\alpha_k = \sigma_{\text{atm}}$, where it is given by the Bergomi formula). Collecting terms yields (31). □

Remark B.2 (What Salon's formula omits and why it is nonetheless sign-correct).

(i) *Feedback amplification.* When $B_t \neq 0$, assumption (S2) fails. The covariation distortion becomes $\mu_t = J_t c^Z_t J_t^\top - \widehat{c}_t$, and by theorem 3.18 the off-diagonal entry gains the feedback mixing term $\beta_t \sigma^2_{S,t}$. The Salon formula thus under-estimates the true vanna carry by the amount

$$\Delta_{\text{feedback}} = \int_0^T \frac{\partial\, \text{Surv}}{\partial S}\, \frac{\partial\, \text{Put}_{\text{D\&I}}}{\partial \alpha_k}\, \beta_t\, \sigma^2_{S,t}\, \mathrm{d}t,$$

which is strictly positive when $\partial\text{Surv}/\partial S > 0$ and $\partial\text{Put}_{\text{D\&I}}/\partial\alpha_k > 0$ (the standard configuration near barriers). In the full framework, this missing term is exactly the feedback-induced component of the covariation conjugation.

(ii) *Sign correctness under no impact.* Under (S2), observed covariation equals fundamental covariation, so the only source of hedging loss is the gap between $(c_t^Z)_{S\alpha_k}$ and the local-vol model covariance $\widehat{c}_{S\alpha_k,t}$. Since empirically the realized spot–vol covariance is more negative than the local-vol model predicts, and since $\partial^2\Pi/(\partial S\,\partial\alpha_k) < 0$ for the cancellable put (negative vanna), the integrand in (31) is negative, yielding a negative carry. Salon's formula therefore produces the correct sign because the dominant covariance gap (realized vs. local-vol) is a model-specification error that exists independently of feedback. Feedback *amplifies* this gap via the conjugation but does not create it.

(iii) *Recovery of the correct term under no impact.* If one accepts (S2), then the full Layer 1 identity and Salon's formula agree *exactly* (up to the product-form approximation (S3) and the vanna-dominance restriction (S1)). In other words, Salon's add-on is the unique first-order correction to the local-vol model that accounts for the realized–implied covariance gap in the cross term, which is precisely what theorem 2.7 delivers when restricted to a single off-diagonal block with $J = I$. No additional structural insight is needed under the no-impact hypothesis: the Bergomi P&L accounting formula, applied to the product-form proxy of the cancellable put, suffices.

References

[1] Matt Cameron. Uridashi derivatives losses climb to $500 million. *Risk Magazine*, April 2013.

[2] Tom Newton and Max Chambers. All mixed up. *Risk Magazine*, June 2013.

[3] Gregory Salon. Equity autocalls and vanna negative carries: Pricing and hedging with a simple add-on. *SSRN Electronic Journal*, 2019.

[4] Chukwuma Dim, Bjorn Eraker, and Grigory Vilkov. 0DTEs: Trading, gamma risk and volatility propagation. *SSRN Electronic Journal*, 2024.

[5] Christine Bangsgaard and Thomas Kokholm. The stock market impact of volatility hedging: Evidence from end-of-day trading by VIX ETPs. *Journal of Banking & Finance*, 180, 2025.

[6] Rüdiger Frey and Alexander Stremme. Market volatility and feedback effects from dynamic hedging. *Math. Financ.*, 7(4):351–374, October 1997.

[7] Philipp J. Schönbucher and Paul Wilmott. The feedback effect of hedging in illiquid markets. *SIAM Journal on Applied Mathematics*, 61(1):232–272, 2000.

[8] Grégoire Loeper. Option pricing with linear market impact and nonlinear Black–Scholes equations. *Annals of Applied Probability*, 28(5):2845–2890, 2018.

[9] Bruno Bouchard, Grégoire Loeper, and Yiyi Zou. Almost-sure hedging with permanent price impact. *Finance and Stochastics*, 20(3):741–771, 2016.

[10] Eduardo Abi Jaber and Eyal Neuman. Optimal liquidation with signals: The general propagator case. *Mathematical Finance*, 35(4):841–866, 2025.

[11] Eduardo Abi Jaber, Eyal Neuman, and Sturmius Tuschmann. Optimal portfolio choice with cross-impact propagators. Manuscript, revised February 2026, 2026.

[12] Aurélien Alfonsi, Fabian Klöck, and Alexander Schied. Multivariate transient price impact and matrix-valued positive definite functions. *Mathematics of Operations Research*, 41(3):914–934, 2016.

[13] Nicolae Gârleanu and Lasse Heje Pedersen. Dynamic trading with predictable returns and transaction costs. *The Journal of Finance*, 68(6):2309–2340, 2013.

[14] Nicolae Gârleanu and Lasse Heje Pedersen. Dynamic portfolio choice with frictions. *Journal of Economic Theory*, 165:487–516, 2016.

[15] Hans Föllmer and Martin Schweizer. Hedging of contingent claims under incomplete information. *Applied Stochastic Analysis*, pages 389–414, 1991.

[16] Alexander Wehrli and Didier Sornette. The excess volatility puzzle explained by financial noise amplification from endogenous feedbacks. *Scientific Reports*, 12:18895, 2022.

[17] Jun Kyung Auh, Wonho Cho, and Thierry Foucault. Liquidation cascade and anticipatory trading: Evidence from the structured equity product market. *Review of Asset Pricing Studies*, 13(1):53–98, 2023.

[18] Benjamin Anderegg, Florian Ulmann, and Didier Sornette. The impact of option hedging on the spot market volatility. *Journal of International Money and Finance*, 124, 2022.

[19] Sebastian Egebjerg and Thomas Kokholm. A model for the hedging impact of option market makers. *SSRN Electronic Journal*, 2024.

[20] Valery I. Oseledets. A multiplicative ergodic theorem. Lyapunov characteristic numbers for dynamical systems. *Trans. Moscow Math. Soc.*, 19:197–231, 1968.

[21] Rafail Khasminskii. *Stochastic Stability of Differential Equations*. Stochastic Modelling and Applied Probability. Springer, 2 edition, 2012.

[22] Lorenzo Bergomi. *Stochastic Volatility Modeling*. Chapman and Hall/CRC, 2016.

[23] Rama Cont and David-Antoine Fournié. Functional Itô calculus and stochastic integral representation of martingales. *The Annals of Probability*, 41(1):109–133, 2013.

[24] Ole E. Barndorff-Nielsen and Neil Shephard. Power and bipower variation with stochastic volatility and jumps. *Journal of Financial Econometrics*, 2(1):1–37, 2004.

[25] Jean Jacod and Albert N Shiryaev. *Limit theorems for stochastic processes*. Grundlehren der mathematischen Wissenschaften. Springer, Berlin, Germany, 2 edition, December 2010.

[26] Philip Protter. *Stochastic integration and differential equations*. Stochastic Modelling and Applied Probability. Springer, Berlin, Germany, 2 edition, December 2010.

[27] Albert S. Kyle. Continuous auctions and insider trading. *Econometrica*, 53(6):1315–1335, 1985.

[28] Robert Almgren and Neil Chriss. Optimal execution of portfolio transactions. *J. Risk*, 3(2):5–39, January 2001.

[29] Mark Rubinstein and Eric Reiner. Breaking down the barriers. *Risk*, 4(8):28–35, 1991.

[30] Harry Furstenberg and Harry Kesten. Products of random matrices. *Annals of Mathematical Statistics*, 31(2):457–469, 1960.

[31] Brian D. O. Anderson and John B. Moore. *Optimal Filtering*. Prentice-Hall, 1979.

[32] Hendrik W. Bode. *Network Analysis and Feedback Amplifier Design*. Van Nostrand, 1945.

[33] Peter Whittle. *Risk-Sensitive Optimal Control*. Wiley, 1990.

[34] Alexander A. Feldbaum. Dual control theory, i–iv. *Automation and Remote Control*, 21:874–880, 1033–1039, 1280–1285, 1453–1464, 1960.

[35] Yaakov Bar-Shalom and Eden Tse. Dual effect, certainty equivalence, and separation in stochastic control. *IEEE Transactions on Automatic Control*, 19(5):494–500, 1974.