Traded-Coordinate Diagnostics for Hedging Feedback

Mohammed AHNOUCH* Erwan LESAOUT*

January 16, 2026

Abstract

When hedging is not price-taking, rebalancing trades can move the traded surface and, in stress, spot itself, so exposures evolve endogenously. In traded coordinates consisting of spot and a finite set of liquid vanilla prices, we prove an exact hedging-error identity for first-order matched self-financing hedges: discounted tracking error equals a Hessian–mismatch pairing involving the discrepancy between realized quadratic covariation and a model-implied covariation density as well as an explicit execution shortfall term. We then introduce a feedback layer as a pointwise fixed point; under standard contraction and smoothness conditions the equilibrium is well posed, admits an amplification operator, and conjugates fundamental covariation into observed covariation. The resulting diagnostics including mismatch functionals, amplification indices, and effective correlations are implementable from price paths and are illustrated on the Nikkei Uridashi episode. Key obstructions such as basis dependence of Greek labels, monitoring with jumps and nonsmoothness, and identification versus causality are all made explicit.

Keywords. Quadratic covariation; covariation mismatch; market impact; feedback trading; fixed-point equilibrium; structured products; variance-optimal hedging.

1 Introduction

In crowded structured-product markets, dealer inventories contain exotics that are hedged through a small set of liquid vanillas. The resulting aggregate flow can move the traded option surface and, in stress, spot itself. When hedging changes traded prices, the Greeks being hedged change endogenously, creating a feedback loop that amplifies shocks. This paper formalizes that mechanism in traded coordinates and develops an empirical protocol to confront the framework with data.

The mathematical tools used throughout are standard (semimartingale Itô calculus for hedging identities, and fixed-point arguments for equilibrium clearing). The paper's contribution is to formalize desk heuristics in a traded-coordinate setting with explicit self-financing and cashflow bookkeeping, and to link a parsimonious feedback layer to observable covariation distortions and implementable diagnostics.

We work in traded coordinates $Y := (S, P^{(1)}, \dots, P^{(n)})$ where S is the underlying and $P^{(1)}, \dots, P^{(n)}$ are liquid vanilla instruments, treating Y as a finite set of liquid traded coordinates rather than non-traded implied-volatility indices. Our focus is on situations where aggregate hedging demand, driven by inventory constraints and risk limits, enters the dynamics of traded factors. In that case the covariation structure is not a primitive input but partially endogenous, and this endogeneity can generate amplified losses under otherwise standard hedging rules. The framework formalizes diagnostic questions about positioning homogeneity, hedge demand relative to market depth, feedback sign through cross-sensitivities (notably the spot–surface vanna channel), and counterparty supply persistence in stress; these are made precise in Section 3.7.

[1]Université Paris 1 Panthéon Sorbonne

We emphasize the following contributions. First, Section 2 derives an exact residual P&L identity for first-order matched self-financing hedges in *traded* coordinates, isolating covariation mismatch as the primitive driver of the second-order residual term; Sections 2.4 and 2.8 record the corresponding monitoring-date bookkeeping and an explicit execution-cost add-on. Second, Section 2.9 records a one-factor price-taking sign restriction for spot–vanilla covariation, providing a sharp benchmark against which attenuation or sign reversals can be diagnosed. Third, Section 3 formulates hedging feedback as a pointwise fixed-point clearing relation; under explicit contraction and smoothness conditions, the equilibrium is well posed and yields an explicit amplification operator whose Jacobian conjugates fundamental covariation into observed covariation. Fourth, Section 4.10 provides a theorem-level reconciliation of common parameter-based decompositions by making the traded mapping explicit and by recording the calibration-transport and calibration-curvature correction terms where they arise. Finally, Sections 3.7 and 4 translate the theory into implementable estimands, estimators, and reporting commitments and illustrate the diagnostics on the Nikkei Uridashi and KOSPI episodes.

The paper connects three strands of literature that are often treated separately. On the modelling side, feedback from dynamic hedging to market volatility has a classical mathematical-finance treatment; a canonical reference is Frey and Stremme [1997], who study volatility and stability effects induced by hedging in the presence of price impact. The literature on hedging with permanent price impact provides rigorous treatments of how trading strategies affect asset prices; Bouchard et al. [2016] study almost-sure hedging under permanent impact, while Bouchard et al. [2017] examine hedging with linear market impact and gamma constraints. Recent work extends this to algorithmic market making with explicit hedging and market impact [Barzykin et al., 2023] and models the hedging impact of option market makers [Egebjerg and Kokholm, 2024]. From the market-microstructure and execution perspective, market impact and liquidity frictions have a large literature; our fixed-point clearing layer is intentionally reduced-form, and we cite Almgren and Chriss [2001] and Cetin et al. [2004] as representative frameworks for impact and liquidity risk in continuous-time modelling. On the identification side, the projection layer based on orthogonal decompositions and span limitations builds on the variance-optimal hedging tradition and the underlying semimartingale calculus; see Föllmer and Schweizer [1991], Schweizer [1995], Eyraud-Loisel [2024], Kunita [2019], Protter [2010] for background. While quadratic hedging approaches provide foundational decompositions of hedging errors into orthogonal components under price-taking dynamics, the present framework extends this by explicitly incorporating endogenous covariation dynamics through a feedback equilibrium, thereby addressing situations where the hedging activity itself alters the realized covariation structure.

Relative to practitioner decompositions for structured products (e.g. autocallable vanna carry discussions such as Salon [2019], Guennoun [2019]), our contribution is not to propose a new parametric implied-volatility dynamics, but to state precisely when such decompositions can be made theorem-level correct once the coordinates are traded and the self-financing and cashflow bookkeeping conditions are made explicit (Section 4.10). Recent empirical work documents liquidation cascades and hedging front-running in structured equity product markets [Auh and Cho, 2020] and examines implicit hedging costs and liquidity frictions [Avdiu and Unger, 2023, Bel Hadj Ayed and Loeper, 2023], providing empirical motivation for the feedback mechanisms formalized here.

The closest studies in the feedback literature emphasize general equilibrium pricing under permanent impact and the resulting fully non-linear valuation equations and constraints (e.g. gamma constraints and non-linear PDE effects) [Bouchard et al., 2016, 2017]. Our emphasis is different: we work in *traded coordinates* and isolate a specific multi-channel mechanism that is central in structured products, namely the *spot–surface vanna channel* in a finite-dimensional traded factor representation. This yields a decomposition that is implementable in stress testing and connects directly to the "parameter-proxy" hedging heuristics used in practice, while

making explicit which steps are theorem-level correct and which require additional assumptions (Section 4.10).

Relative to the practitioner-oriented P&L decomposition literature, Bergomi [2016] develops Greek-style decompositions under exogenous (price-taking) dynamics for $(S, \text{vol factors})$ in diffusion/Itô settings, showing how residual P&L is driven by second-order exposures (gamma/vanna/volga) contracting with realized quadratic covariation. Our framework overlaps with that mechanical Itô insight but differs in three ways: (i) we insist on *traded coordinates* as primitives, treating implied-vol factors only via explicit mappings to traded prices, which avoids non-traded factor ambiguity; (ii) we frame the residual as an invariant pairing $-\frac{1}{2}\int \mathrm{Tr}(D^2_{yy} v\, \mu)$ where $\mu = \mathrm{d}\langle Y\rangle - \widehat{c}\,\mathrm{d}t$ is explicitly a *covariation mismatch* between realized and model-implied covariation, making the decomposition model-agnostic at the stochastic-calculus level; and (iii) we add a *feedback layer* that endogenizes the very covariations being decomposed, via a fixed-point equilibrium and amplification operator $J_t = (I - \Lambda D_y \Phi)^{-1}$ that conjugates fundamental covariation into observed covariation. In Bergomi's framework, hedging does not change factor dynamics; in ours, it does, turning "P&L decomposition" into "P&L decomposition plus a mechanism that changes the covariations being decomposed." Additionally, when parameter-based decompositions are used, we make explicit the transport through the calibration map and the calibration-curvature correction terms (Theorem 4.6), which is where many desk decompositions become logically loose.

Section 2 develops the traded-coordinate calculus and proves the hedging-error identity, together with identification tools and benchmark special cases. Section 3 formulates the feedback equilibrium as a fixed point and derives stability, amplification criteria, and implementable diagnostic quantities. Section 4 contains the data, estimators, tests, and reproducible evidence, and interprets prior decompositions through the lens of our framework. Section 5 discusses applicability, limitations, and concludes. Mathematical preliminaries and proofs are collected in the appendices.

1.1 Institutional context and product mechanisms

This subsection records sourced practitioner observations that motivate later modelling choices and fixes the terminology linking desk narratives to mathematical objects. It is not used as a premise in any theorem.

1.1.1 Representative product families

Many feedback concerns arise in product families whose payoff depends on discrete monitoring and state variables tracking whether the contract has been called or knocked in. Callable equity-linked notes provide a canonical example. For such contracts, effective maturity, barrier proximity, and call likelihood can change rapidly with spot, inducing time-varying and state-dependent sensitivities to the option surface.[1] We therefore treat the contract as a finite-state object driven by observations along a monitoring grid.

Autocallable architecture as a state machine Fix a monitoring grid $0 = t_0 < t_1 < \cdots < t_N = T$. A wide class of structured notes can be represented by a finite set of contract states together with an observation rule at each date and a deterministic transition rule. Concretely, one can view the contract as a deterministic transducer that reads, at each t_k, a symbol encoding which region the underlying has fallen into and then updates a contract state accordingly. This representation expresses cancellability and barrier-type features without ambiguity, and makes explicit that the contract value is a function of time, current market state, and current contract state.

[1] See, for example, the uridashi and autocallability accounts in Cameron [2013], Salon [2019].

Many equity autocallables that featured in the Uridashi and KOSPI episodes can be idealized as follows. Let τ_{AC} be the first monitoring date at which the note autocalled, and let τ_{KI} be the first monitoring date at which a knock-in barrier was breached. Let $(t_i)_{i=1}^N$ be the monitoring dates and let q_i be the coupon paid at t_i provided the note has not autocalled earlier. A stylized payoff to the investor can be written as

$$\Pi = \Big(\sum_{i=1}^{N} q_i \, \mathbb{1}_{\{t_i \le \tau_{\mathrm{AC}} \wedge T\}} \Big) \; + \; \mathbb{1}_{\{\tau_{\mathrm{AC}} \le T\}} \; + \; \mathbb{1}_{\{\tau_{\mathrm{AC}} > T\}} \Big(1 - \mathbb{1}_{\{\tau_{\mathrm{KI}} \le T\}} \cdot \frac{1}{K} \, (K - S_T)^+ \Big).$$

In words: coupons are paid periodically while the note is alive; if it autocalled before maturity the notional is reimbursed; otherwise at maturity the terminal redemption equals notional minus a down-and-in put, conditional on survival. From the dealer perspective the position is long this embedded put, and as spot approaches the knock-in region, the continuation value becomes increasingly sensitive to the put component, concentrating second-order exposures in a narrow corridor.

A qualitative regime map for sensitivities The key modelling point is not a particular Greek label but the empirical fact that sensitivity profiles of barrier-rich and callable products can shift rapidly across spot regions and contract states. In our framework this enters through the state dependence of $v(t, y, \xi)$ and hence of $\nabla_y v$ and $D^2_{yy} v$ in traded coordinates, which in turn drives state-dependent hedging flows in the feedback layer.

In later sections we assume that the exotic value admits a traded-coordinate representation of the form $V_t = v(t, S_t, P_t, \xi_t)$, where P_t denotes traded vanilla coordinates and ξ_t denotes the current contract state. This motivates why ξ_t is essential for any callable or barrier-rich structure.

1.1.2 One-way markets and crowded hedging

Practitioner discussions of stress episodes frequently describe markets in which risk transfer is persistent in one direction, so that dealers intermediate large inventories that share similar factor exposures.[Cameron, 2013] The phrase "one-way market" conveys that natural end-investor flow is systematically biased toward issuing certain contingent claims and that the set of agents able to warehouse the opposite exposures is limited, reducing effective liquidity for rebalancing. The paper uses this as motivation for a modelling layer in which hedging flow enters the dynamics of traded factors. This modelling step is not a hidden empirical claim but a conditional hypothesis: *if* hedging flow enters the factor dynamics with sufficient strength and alignment, then amplification and instability can occur even under hedging rules that would be appropriate in a price-taking regime. Whether this holds is an empirical question addressed in Section 4.

1.1.3 From narrative to mathematical objects

The preceding discussion motivates a precise mapping from desk language to mathematical primitives. When practitioners speak of "spot" and "the surface", we model these as traded semimartingales: the underlying price process S and a vector of traded vanilla coordinates P. When practitioners speak of "inventory" and "risk limits", we represent these through a state variable that enters a hedging rule, denoted Q_t for aggregate hedging flow at time t. When practitioners speak of "impact" or "liquidity", we represent this as a coupling between Q_t and the dynamics of (S_t, P_t) through an impact operator. The feedback loop becomes a coupled system in which Q depends on (S, P) and (S, P) evolves with a drift component induced by Q.

When we discuss covariation between spot and volatility, we always specify whether the volatility object is traded or statistical; our main theorems are stated in traded coordinates, where quadratic covariation is unambiguously defined for semimartingale prices. This separation

is essential for the correctness of the hedging-error identity and for a clean interpretation of what remains as residual risk.

2 Model and hedging-error identity

We work on a filtered probability space with $\mathbb{Q}$ a pricing measure and numéraire B. For any traded price process U we write $\widetilde{U}_t := U_t/B_t$ for its discounted version. The market contains an underlying S and liquid vanilla instruments $P^{(1)}, \dots, P^{(n)}$. We collect the traded factors into $Y := (S, P^{(1)}, \dots, P^{(n)})$. All price processes are càdlàg semimartingales. A trading strategy $\Theta_t = (\eta_t, \phi_t, \psi_t^{(1)}, \dots, \psi_t^{(n)})$ is *self-financing* if $\mathrm{d}X_t = \eta_t \,\mathrm{d}B_t + \phi_t \,\mathrm{d}S_t + \sum_{i=1}^n \psi_t^{(i)} \,\mathrm{d}P_t^{(i)}$. We denote by $\langle Y, Z\rangle$ the quadratic covariation of semimartingales Y and Z. Complete mathematical preliminaries are in Appendix B.

The exotic value is represented as $V_t = v(t, S_t, P_t, \xi_t)$, where P_t denotes traded vanilla coordinates and ξ_t denotes the current contract state. We assume v is $C^{1,2}$ in $(t, (S, P))$ and that the discounted traded price vector $\widetilde{Y}$ is a continuous local martingale under $\mathbb{Q}$. Extensions to non-smooth payoffs are treated in Appendix C.

2.1 Traded-coordinate hedging calculus

This section develops the paper's central decomposition. The starting point is a simple observation: once one chooses a set of *traded* coordinates in which to represent the exotic value, Itô's formula expresses the infinitesimal evolution of that value as the sum of a first-order term driven by traded price increments and a second-order term driven by quadratic covariation. A self-financing hedge that matches first-order sensitivities cancels the first-order term by construction. What remains is a residual term that depends only on the second-order structure, namely the realized covariation of the traded coordinates and the covariation structure embedded by the model used to compute the hedge. This decomposition is model-free at the level of stochastic calculus, and it becomes a tractable measurement object once it is expressed in traded coordinates.

To avoid any ambiguity, we proceed in a numéraire-discounted setting. This removes interest-rate bookkeeping from the algebra and focuses attention on the covariation terms that will later be affected by feedback.

2.2 Discounted market and traded-coordinate representation

Fix the numéraire B introduced in Appendix A. For any traded price process Y we write $\widetilde{Y} := Y/B$. We assume that discounted traded prices are continuous local martingales under $\mathbb{Q}$.

Assumption 2.1 (No-arbitrage form used in the calculus)**.** The discounted traded price vector

$$\widetilde{Y}_t := (\widetilde{S}_t, \widetilde{P}_t^{(1)}, \dots, \widetilde{P}_t^{(n)})$$

is a continuous $(n+1)$-dimensional local martingale under $\mathbb{Q}$.

Assumption 2.1 is the only probabilistic input used in the main identity. It is standard under no-arbitrage when $\mathbb{Q}$ is chosen as the pricing measure associated with the numéraire B, and it is deliberately weaker than specifying a parametric diffusion for (S, P).

We now state the representation assumption for the exotic value. In the main argument we treat the contract state as fixed, because the stochastic calculus applies on each interval where the contract state is constant. In Section 2.8 we extend the identity to deterministic monitoring grids with state transitions.

Assumption 2.2 (Traded-coordinate representation)**.** There exists a finite state space Ξ and a function

$$v : [0,T] \times \mathbb{R}^{n+1} \times \Xi \to \mathbb{R}$$

such that the *discounted* exotic value satisfies

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

and for each fixed $\xi \in \Xi$ the map $(t,y) \mapsto v(t,y,\xi)$ is of class $C^{1,2}$ with derivatives of at most polynomial growth on the relevant domain.

Throughout, "sensitivities" mean the state derivatives $\nabla_y v$ and $D^2_{yy} v$ in traded coordinates, evaluated at $(t, \widetilde{Y}_t, \xi_t)$, so standard finite-dimensional Itô calculus applies. Functional Itô calculus is only needed for genuinely path-dependent values of the form $\widetilde{V}_t = F(t, (\widetilde{Y}_u)_{u \le t})$; discretely monitored barrier and autocall structures are covered here by encoding contractual logic in a finite state ξ and handling monitoring-date updates via Section 2.8 (with non-smooth refinements deferred to Appendix C).

2.3 Self-financing versus first-order matching

The subsequent identity relies on keeping two notions logically distinct. A strategy is self-financing by definition of its gains process. Independently, one may *choose* holdings that match first-order derivatives of a function v with respect to traded coordinates. The first notion is accounting; the second is design.

Let $\Theta = (\eta, \phi, \psi^{(1)}, \dots, \psi^{(n)})$ be an admissible strategy and let $\widetilde{X}$ denote its discounted value. Since the discounted numéraire is identically one, self-financing reduces to the statement that the discounted gains equal the stochastic integral against discounted traded prices.

Lemma 2.3 (Discounted self-financing identity (standard))**.** *If Θ is self-financing in the sense of Definition A.2, then its discounted value satisfies*

$$\mathrm{d}\widetilde{X}_t = \phi_t \, \mathrm{d}\widetilde{S}_t + \sum_{i=1}^{n} \psi_t^{(i)} \, \mathrm{d}\widetilde{P}_t^{(i)} = \vartheta_t^{\top} \, \mathrm{d}\widetilde{Y}_t,$$

where $\vartheta_t := (\phi_t, \psi_t^{(1)}, \dots, \psi_t^{(n)})^{\top}$.

Proof (standard; sketch). Apply the product rule to X_t/B_t using that B is of finite variation: $\mathrm{d}(X_t/B_t) = (1/B_t)\,\mathrm{d}X_t - (X_t/B_t^2)\,\mathrm{d}B_t$. Substituting the self-financing condition $\mathrm{d}X_t = \eta_t \,\mathrm{d}B_t + \phi_t \,\mathrm{d}S_t + \sum_i \psi_t^{(i)} \,\mathrm{d}P_t^{(i)}$ and using $X_t = \eta_t B_t + \phi_t S_t + \sum_i \psi_t^{(i)} P_t^{(i)}$ to eliminate the $\mathrm{d}B_t$ terms yields the stated identity. □

2.4 Execution costs: from book-price identities to realized P&L

The main hedging-error identities in the paper are written in terms of traded-price semimartingales and self-financing gains. In practice, especially in stress regimes, a hedger may be unable to rebalance at the reference (mid/mark) price process used for bookkeeping. Bid–ask spreads, market impact, latency, and partial fills create an additional implementation shortfall beyond the covariation-mismatch term. A simple way to represent this without committing to a full microstructure model is to include an explicit cumulative execution-cost process in the discounted self-financing equation.

Proposition 2.4 (Hedging-error identity with an execution-cost process (standard bookkeeping))**.** *Assume the setting of Theorem 2.7 on an interval $I = [t_0, t_1]$ where $\xi_t \equiv \xi$ is constant, and fix a predictable strategy $\vartheta_t = \nabla_y v(t, \widetilde{Y}_{t-}, \xi)$. Let K be an adapted finite-variation process with*

$K_{t_0} = 0$ *(interpreted as cumulative execution cost in discounted units). Suppose the discounted wealth process satisfies the* costly self-financing *relation*

$$\widetilde{X}_t = \widetilde{X}_{t_0} + \int_{t_0}^{t} \vartheta_u^\top \, \mathrm{d}\widetilde{Y}_u - K_t, \qquad t \in I, \tag{1}$$

with $\widetilde{X}_{t_0} = v(t_0, \widetilde{Y}_{t_0}, \xi)$*. Then for all* $t \in I$*,*

$$\widetilde{X}_t - v(t, \widetilde{Y}_t, \xi) = -\frac{1}{2} \int_{t_0}^{t} \mathrm{Tr}\Big(D_{yy}^2 v(u, \widetilde{Y}_u, \xi) \, (\, \mathrm{d} \left\langle \widetilde{Y}, \widetilde{Y} \right\rangle_u - \widehat{c}(u, \widetilde{Y}_u, \xi) \, \mathrm{d}u) \Big) - K_t. \tag{2}$$

In particular, if K *is nondecreasing (pure cost), then execution frictions can only worsen realized hedging P&L relative to the frictionless book-price identity.*

Proof (standard; sketch). Repeat the proof of Theorem 2.7 with the costly self-financing dynamics (1) in place of (5): the Itô cancellation is unchanged and the only difference is the additional finite-variation term $-K_t$, which yields (2). □

First-order matching in traded coordinates means choosing ϑ as the gradient of v with respect to the traded coordinate vector. This is a choice of hedge design and does not follow from self-financing.

Definition 2.5 (First-order matched holdings in traded coordinates)**.** Fix $\xi \in \Xi$ and define the gradient with respect to $y \in \mathbb{R}^{n+1}$ by

$$\nabla_y v(t, y, \xi) = \Big(\partial_{y_0} v(t, y, \xi), \ldots, \partial_{y_n} v(t, y, \xi) \Big)^\top.$$

A strategy is said to be *first-order matched* to v on an interval $I \subset [0, T]$ if for all $t \in I$ it holds that

$$\vartheta_t = \nabla_y v(t, \widetilde{Y}_{t-}, \xi_t).$$

Under the standing continuity assumption, $\widetilde{Y}_{t-} = \widetilde{Y}_t$ and the first-order matched rule is predictable; when jumps are present, using the left limit $\widetilde{Y}_{t-}$ is the predictable (implementable) choice.

2.5 A benchmark "model-implied covariation"

The decomposition compares realized covariation to the covariation structure implicitly used when computing the hedging function v. Since we do not assume a parametric diffusion for $\widetilde{Y}$, we record the model-implied object at the level of a predictable covariation density.

Assumption 2.6 (Model-implied covariation density)**.** There exists a measurable map

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

into the cone of symmetric positive semidefinite matrices such that, for each fixed $\xi \in \Xi$, the function $v(\cdot, \cdot, \xi)$ satisfies the backward equation

$$\partial_t v(t, y, \xi) + \frac{1}{2} \mathrm{Tr}\Big(\widehat{c}(t, y, \xi) \, D_{yy}^2 v(t, y, \xi) \Big) = 0 \tag{3}$$

for $(t, y) \in [0, T) \times \mathbb{R}^{n+1}$, together with a terminal condition $v(T, y, \xi) = g(y, \xi)$ for a prescribed payoff map g.

Assumption 2.6 is intentionally phrased so that the "model" enters only through the covariation density $\widehat{c}$. In the classical case where $\widetilde{Y}$ is a diffusion under the model with instantaneous covariance $\widehat{c}$, (3) is exactly the condition ensuring that $t \mapsto v(t, \widetilde{Y}_t, \xi)$ is a local martingale. Here, we do not assume that the *true* market covariation matches $\widehat{c}$; the mismatch will appear explicitly in the residual P&L.

2.6 Main hedging-error identity and full proof

We now state and prove the core identity. The derivation uses standard self-financing bookkeeping and Itô calculus; the emphasis is that the residual is expressed in traded coordinates as a contraction against the covariation mismatch $\mathrm{d}\left\langle \widetilde{Y}, \widetilde{Y} \right\rangle - \widehat{c}\,\mathrm{d}t$. The statement is local on time intervals where the contract state is fixed. Extension to deterministic monitoring dates and state transitions follows by summing over intervals, and is treated in Section 2.8.

Theorem 2.7 (Traded-coordinate hedging-error identity)**.** *Assume 2.1, 2.2, and 2.6. Fix an interval $I = [t_0, t_1] \subset [0, T]$ such that $\xi_t \equiv \xi$ is constant on I. Define a self-financing discounted strategy $\widetilde{X}$ by setting $\widetilde{X}_{t_0} = v(t_0, \widetilde{Y}_{t_0}, \xi)$ and choosing holdings*

$$\vartheta_t = \nabla_y v(t, \widetilde{Y}_{t-}, \xi), \qquad t \in I.$$

Then for all $t \in I$ the hedging error admits the identity

$$\widetilde{X}_t - v(t, \widetilde{Y}_t, \xi) = -\frac{1}{2} \int_{t_0}^{t} \mathrm{Tr}\Big(D_{yy}^2 v(u, \widetilde{Y}_u, \xi) \, \big(\mathrm{d}\left\langle \widetilde{Y}, \widetilde{Y} \right\rangle_u - \widehat{c}(u, \widetilde{Y}_u, \xi)\,\mathrm{d}u \big) \Big). \tag{4}$$

In particular, the residual depends only on second derivatives of v in traded coordinates and on the difference between realized and model-implied quadratic covariation of traded prices.

Proof. Step 1 (self-financing; standard). Fix $t \in I$. We begin by writing the discounted wealth process under the chosen self-financing strategy. By Lemma 2.3 and the definition of the strategy,

$$\widetilde{X}_t = \widetilde{X}_{t_0} + \int_{t_0}^{t} \vartheta_u^{\top} \, \mathrm{d}\widetilde{Y}_u = v(t_0, \widetilde{Y}_{t_0}, \xi) + \int_{t_0}^{t} \nabla_y v(u, \widetilde{Y}_{u-}, \xi)^{\top} \, \mathrm{d}\widetilde{Y}_u. \tag{5}$$

Step 2 (Itô formula; standard). Next, apply Itô's formula to the semimartingale $u \mapsto v(u, \widetilde{Y}_u, \xi)$ on the interval I. Since $\widetilde{Y}$ is a continuous semimartingale and $v(\cdot, \cdot, \xi)$ is $C^{1,2}$ by Assumption 2.2, Itô's formula yields

$$\begin{aligned} v(t, \widetilde{Y}_t, \xi) = v(t_0, \widetilde{Y}_{t_0}, \xi) &+ \int_{t_0}^{t} \partial_t v(u, \widetilde{Y}_u, \xi) \, \mathrm{d}u + \int_{t_0}^{t} \nabla_y v(u, \widetilde{Y}_u, \xi)^{\top} \, \mathrm{d}\widetilde{Y}_u \\ &+ \frac{1}{2} \int_{t_0}^{t} \mathrm{Tr}\Big(D_{yy}^2 v(u, \widetilde{Y}_u, \xi) \, \mathrm{d}\left\langle \widetilde{Y}, \widetilde{Y} \right\rangle_u \Big). \end{aligned} \tag{6}$$

Step 3 (cancellation and insertion of the benchmark; standard). Subtracting (6) from (5) cancels the stochastic integrals exactly and gives

$$\widetilde{X}_t - v(t, \widetilde{Y}_t, \xi) = -\int_{t_0}^{t} \partial_t v(u, \widetilde{Y}_u, \xi) \, \mathrm{d}u - \frac{1}{2} \int_{t_0}^{t} \mathrm{Tr}\Big(D_{yy}^2 v(u, \widetilde{Y}_u, \xi) \, \mathrm{d}\left\langle \widetilde{Y}, \widetilde{Y} \right\rangle_u \Big). \tag{7}$$

We now invoke the model-implied covariation density. By Assumption 2.6, for each (u, y) we have

$$\partial_t v(u, y, \xi) = -\frac{1}{2} \mathrm{Tr}\Big(\widehat{c}(u, y, \xi) \, D_{yy}^2 v(u, y, \xi) \Big).$$

Substituting $y = \widetilde{Y}_u$ and inserting this identity into (7) yields

$$\widetilde{X}_t - v(t, \widetilde{Y}_t, \xi) = \frac{1}{2} \int_{t_0}^{t} \mathrm{Tr}\Big(\widehat{c}(u, \widetilde{Y}_u, \xi) \, D_{yy}^2 v(u, \widetilde{Y}_u, \xi) \Big) \, \mathrm{d}u - \frac{1}{2} \int_{t_0}^{t} \mathrm{Tr}\Big(D_{yy}^2 v(u, \widetilde{Y}_u, \xi) \, \mathrm{d}\left\langle \widetilde{Y}, \widetilde{Y} \right\rangle_u \Big).$$

Step 4 (mismatch contraction in traded coordinates; extension highlighted). Using the cyclic invariance of the trace to write $\mathrm{Tr}(\widehat{c} D^2 v) = \mathrm{Tr}(D^2 v\, \widehat{c})$, we may combine the two terms into a single integral against the signed matrix-valued measure $\mathrm{d}\left\langle \widetilde{Y}, \widetilde{Y} \right\rangle_u - \widehat{c}(u, \widetilde{Y}_u, \xi)\,\mathrm{d}u$, obtaining precisely (4). Steps 1–3 are standard; the emphasis is that the residual is expressed as a pairing of the traded-coordinate Hessian with the traded covariation mismatch. □

In a two-coordinate reduction $Y = (S, U)$ (spot and a chosen traded surface coordinate or explicit proxy), the invariant contraction is simply

$$\mathrm{Tr}(D^2_{yy} v\, \mu) = v_{SS}\, \mu_{SS} + 2 v_{SU}\, \mu_{SU} + v_{UU}\, \mu_{UU}.$$

The familiar "gamma/vanna/volga" language is nothing more than these coordinate projections; the theorem-level object is the basis-invariant pairing $\mathrm{Tr}(D^2_{yy} v\, \mu)$ in the chosen traded-coordinate vector.

Definition 2.8 (Functional Greeks as basis projections)**.** Fix a two-dimensional reduction $Y = (S, U)$, where U denotes a *chosen traded surface coordinate* (e.g. the discounted price of a designated ATM option) or an *explicitly labelled volatility proxy extracted from traded vanillas.* In this reduced basis, the Hessian weights in Theorem 2.7 admit the coordinate projections

- *Functional gamma:* $D^2_{SS} v$, weighting spot-variance mismatch μ_{SS}.
- *Functional vanna:* $D^2_{SU} v$, weighting spot–surface covariation mismatch μ_{SU}.
- *Functional volga:* $D^2_{UU} v$, weighting surface-variance mismatch μ_{UU}.

These are *functionals of the chosen representation*: changing the surface coordinate U (or replacing U by a different traded/proxy basis) changes these projected coefficients, while leaving the invariant contraction $\mathrm{Tr}(D^2_{yy} v\, \mu)$ unchanged.

These coordinate labels are basis-dependent: changing the chosen surface coordinate changes the projected coefficients while leaving $\mathrm{Tr}(D^2_{yy} v\, \mu)$ unchanged. When we later say "vanna-driven", we mean state dependence of hedge weights (a spot-to-surface flow channel), not a coordinate-free primitive.

Finally, (4) is a bookkeeping identity: given a chosen model pair $(v, \widehat{c})$, it expresses the exact residual generated by covariation mismatch for the corresponding first-order matched self-financing hedge. It is not a replication theorem; exact tracking requires both correct covariation along the realized path and sufficient spanning power of the traded factor set.

2.7 Calibration error and its propagation to residual terms

The identity (4) is conditional on a fixed value function v and its derivatives. In practice, v must be computed from a calibrated model and a numerical solver, and the resulting implementation $\widehat{v}$ may deviate from an intended target v. To separate this implementation layer from covariation mismatch, write $\epsilon(t, y, \xi) := \widehat{v}(t, y, \xi) - v(t, y, \xi)$.

If a self-financing hedge uses first-order matched holdings $\vartheta_t = \nabla_y \widehat{v}(t, \widetilde{Y}_{t-}, \xi)$ while the reference value is v, then on any interval $I = [t_0, t_1]$ where ξ is constant one has the exact decomposition

$$\widetilde{X}_t - v(t, \widetilde{Y}_t, \xi) = -\frac{1}{2} \int_{t_0}^{t} \mathrm{Tr}\Big(D^2_{yy} v(u, \widetilde{Y}_u, \xi)\, \big(\mathrm{d} \left\langle \widetilde{Y}, \widetilde{Y} \right\rangle_u - \widehat{c}(u, \widetilde{Y}_u, \xi)\, \mathrm{d}u \big) \Big) + \int_{t_0}^{t} \nabla_y \epsilon(u, \widetilde{Y}_u, \xi)^\top \, \mathrm{d}\widetilde{Y}_u, \qquad t \in I. \quad (8)$$

The second term isolates the effect of hedging with an incorrect gradient. If one instead takes $\widehat{v}$ as the value representation used to define the hedge and the benchmark, then Theorem 2.7 applies with $(\widehat{v}, \widehat{c})$ and the residual remains a Hessian-weighted covariation mismatch, but with weights computed from $\widehat{v}$.

2.8 Deterministic monitoring dates and contract-state updates

Autocallables and barrier notes—the products at the center of the Uridashi and KOSPI episodes—feature deterministic monitoring dates at which cashflows may be triggered or the contract may terminate. A rigorous framework must handle these discontinuities explicitly; silently assuming "continuous dynamics everywhere" is a common source of hidden errors in callable-product calculus.

The proof of Theorem 2.7 applies on any interval where the contract state is constant. For typical structured notes, the contract state may update at deterministic monitoring dates and may trigger deterministic cashflow rules conditional on the market state. This subsection records the extension of the identity to such piecewise dynamics in a form that introduces no hidden jumps. The result is Proposition 2.9, which shows that when cashflows are matched correctly, the jump terms cancel and only the continuous covariation-mismatch integral survives—restoring the same functional form as the base identity.

Fix a deterministic monitoring grid $0 = t_0 < t_1 < \cdots < t_N = T$. Assume that ξ is constant on each open interval (t_k, t_{k+1}) and may update at t_k according to a deterministic transition map that depends on the observed traded coordinates $\widetilde{Y}_{t_k}$. Assume further that any contractual cashflows are represented by a finite-variation adapted process C that is constant on each (t_k, t_{k+1}) and whose jumps at monitoring dates represent coupon or call payments. We interpret $\widetilde{V}$ as the ex-cashflow discounted value process, so that the total discounted value including paid cashflows is $\widetilde{V}_t + C_t$.

Under these conventions, a hedging strategy that finances contractual cashflows and re-hedges on each interval yields an identity obtained by summing the interval-wise formula. We state this explicitly.

Proposition 2.9 (Piecewise identity with monitoring dates (standard bookkeeping))**.** *Assume 2.1, 2.2, and 2.6 on each open interval (t_k, t_{k+1}) with the appropriate constant state $\xi^{(k)}$. Let $\widetilde{X}$ be a discounted self-financing portfolio that, on each (t_k, t_{k+1}), holds*

$$\vartheta_t = \nabla_y v\big(t, \widetilde{Y}_{t-}, \xi^{(k)}\big),$$

and that pays the contractual cashflow jumps by decreasing wealth by ΔC_{t_k} at time t_k. Then for any $m \in \{0, \dots, N\}$,

$$\begin{aligned}\widetilde{X}_{t_m} - v(t_m, \widetilde{Y}_{t_m}, \xi_{t_m}) = &-\frac{1}{2}\sum_{k=0}^{m-1} \int_{t_k}^{t_{k+1}} \mathrm{Tr}\Big(D_{yy}^2 v(u, \widetilde{Y}_u, \xi^{(k)})\big(\,\mathrm{d}\left\langle \widetilde{Y}, \widetilde{Y}\right\rangle_u - \widehat{c}(u, \widetilde{Y}_u, \xi^{(k)})\,\mathrm{d}u\big)\Big) \\ &- \sum_{k=1}^{m} \Big(\Delta v(t_k, \widetilde{Y}_{t_k}, \xi_{t_k}) + \Delta C_{t_k}\Big), \end{aligned} \tag{9}$$

where $\Delta v(t_k, \widetilde{Y}_{t_k}, \xi_{t_k})$ denotes the jump of the process $t \mapsto v(t, \widetilde{Y}_t, \xi_t)$ induced by the state update at t_k.

Proof (standard; sketch). Step 1. Apply Theorem 2.7 on each interval $[t_k, t_{k+1})$ where the state is constant to obtain the continuous covariation-mismatch integral term. *Step 2.* Telescope the resulting identities across $k = 0, \dots, m-1$ and add the discrete bookkeeping terms: the portfolio wealth jump $-\Delta C_{t_k}$ and the representation jump $\Delta v(t_k, \widetilde{Y}_{t_k}, \xi_{t_k})$, which yields (9). □

The jump-discrepancy sum in (9) is written explicitly to avoid a common hidden assumption: it vanishes only when the value representation is cashflow-matched at monitoring dates (ex-cashflow value plus an explicit jump condition), so any residual is not an artefact of inconsistent coupon/call bookkeeping.

2.9 Immediate consequences and a benchmark sanity check

Theorems 2.7 and Proposition 2.9 isolate a residual term that depends on two ingredients: the Hessian of the model value function in traded coordinates and the discrepancy between realized and model-implied quadratic covariation of traded prices. Two points are worth highlighting because they recur throughout the paper.

On the interpretation side, the residual is *not* a statement about the correctness of any particular parametric dynamics. It is a statement about whether the covariation structure implicitly assumed when computing the hedge aligns with realized covariation of traded coordinates.

On the mechanism side, the formula makes transparent how feedback can enter. The hedging rule determines positions and therefore aggregate flow. If flow affects the dynamics of traded factors, it alters the realized covariation $\mathrm{d}\left\langle \widetilde{Y}, \widetilde{Y} \right\rangle$ and can therefore systematically change the residual term, even if the pricing model used to compute $\widehat{c}$ is internally consistent.

A minimal sanity check illustrates the role of correct covariation. Suppose that, under the true market dynamics, the discounted traded factor satisfies

$$\mathrm{d}\left\langle \widetilde{Y}, \widetilde{Y} \right\rangle_t = \widehat{c}(t, \widetilde{Y}_t, \xi)\ \mathrm{d}t$$

on an interval where ξ is constant. Then (4) yields $\widetilde{X}_t = v(t, \widetilde{Y}_t, \xi)$ for all t on that interval, meaning that the first-order matched self-financing hedge tracks the model value exactly. The formula thus recovers the classical replication conclusion as a special case, without requiring any additional argument. Conversely, whenever the realized covariation deviates from the model-implied covariation, the identity quantifies the resulting hedging error in a way that is directly attributable to the discrepancy.

Proposition 2.10 (Baseline sign restriction in one-factor price-taking models (standard))**.** *Assume that both S and P are* continuous *and that the numéraire B is deterministic (for instance $B_t = e^{rt}$ for a constant r). Assume further that the discounted factor set is one-factor in the sense that $\widetilde{Y} = (\widetilde{S}, \widetilde{P})$ with $\widetilde{P}_t = p(t, \widetilde{S}_t)$ for some $p \in C^{1,2}([0,T) \times \mathbb{R}_{>0})$. Then*

$$\mathrm{d}\left\langle \widetilde{S}, \widetilde{P} \right\rangle_t = \partial_s p(t, \widetilde{S}_t)\ \mathrm{d}\left\langle \widetilde{S} \right\rangle_t .$$

In particular, if $\partial_s p(t,s) < 0$ (as for a put price as a function of spot), then $\mathrm{d}\left\langle \widetilde{S}, \widetilde{P} \right\rangle_t$ is non-positive and is strictly negative whenever $\mathrm{d}\left\langle \widetilde{S} \right\rangle_t > 0$.

This benchmark is deliberately narrow: the strict sign conclusion relies on continuity and a deterministic numéraire. With jumps (e.g. barrier-triggered discontinuities) a jump-product term enters covariation, and with stochastic rates an additional factor enters discounted prices, so sign restrictions need not follow from $\partial_s p < 0$ alone.

Proof (standard; sketch). Apply Itô's formula to $\widetilde{P}_t = p(t, \widetilde{S}_t)$: the local martingale part is $\int \partial_s p(t, \widetilde{S}_t)\, \mathrm{d}\widetilde{S}_t$, and quadratic covariation depends only on local martingale parts. □

If $\widetilde{P}$ is treated as a traded discounted price under the pricing measure $\mathbb{Q}$, then it is a local martingale and the same one-factor setup yields the familiar pricing PDE for p; this is not needed for the covariation identity, but it aligns the benchmark with no-arbitrage intuition.

Proposition 2.10 provides a sharp benchmark. In a one-factor price-taking model, spot–put covariation cannot attenuate to zero or flip sign without either leaving the one-factor setting (introducing an orthogonal factor in the vanilla price dynamics) or leaving the price-taking setting (so that the mapping from the fundamental state to the observed traded state becomes endogenous, as in Section 3).

The next subsection refines this interpretation by introducing a projection viewpoint that clarifies what is spanned by the chosen traded-factor set and what remains orthogonal residual risk. This will be useful both conceptually and for later empirical diagnostics.

2.10 Identification tools: Kunita-Watanabe decomposition and factor projections

The hedging-error identity in Section 2.6 is an exact accounting statement for the particular self-financing strategy obtained by first-order matching of the traded-coordinate value function. It is often useful to complement this viewpoint with an identification result that is intrinsic to the chosen traded-factor set, independent of any PDE representation. The appropriate tool is the Galtchouk-Kunita-Watanabe decomposition, which expresses any square-integrable martingale as the sum of a stochastic integral against the traded factors and an orthogonal residual. This decomposition clarifies, in a mathematically precise sense, what portion of the exotic's risk is spanned by the chosen traded-factor family and what portion remains orthogonal at that factor resolution.

Throughout this subsection we remain in discounted units. Let

$$M_t := \widetilde{Y}_t = (\widetilde{S}_t, \widetilde{P}_t^{(1)}, \dots, \widetilde{P}_t^{(n)})$$

be the $(n+1)$-dimensional discounted traded price vector, which is a continuous local martingale under $\mathbb{Q}$ by Assumption 2.1. Since the decomposition is an L^2 statement, we impose square-integrability.

Assumption 2.11 (Square-integrability)**.** The martingale M is square-integrable (in particular $\mathbb{E}_{\mathbb{Q}}[\mathrm{Tr}(\langle M, M\rangle_T)] < \infty$), and the discounted exotic value process $\widetilde{V}$ is a square-integrable $\mathbb{Q}$-martingale on $[0, T]$.

Assumption 2.11 is standard when payoffs are integrable under $\mathbb{Q}$ and prices are computed as conditional expectations under the chosen numéraire. It can be relaxed by localization, but square-integrability keeps the orthogonal projection argument transparent.

Theorem 2.12 (Galtchouk-Kunita-Watanabe decomposition (classical))**.** *Assume 2.11. There exist a predictable $(n+1)$-dimensional process θ and a square-integrable martingale L such that*

$$\widetilde{V}_t = \widetilde{V}_0 + \int_0^t \theta_u^\top \, \mathrm{d}M_u + L_t, \qquad t \in [0, T], \tag{10}$$

and L is strongly orthogonal to M, meaning that $\left\langle L, M^{(i)} \right\rangle \equiv 0$ for each component $M^{(i)}$ of M. The pair (θ, L) is unique in the following sense: L is unique up to indistinguishability, and θ is unique up to indistinguishability on the support of the matrix-valued measure $\mathrm{d}\langle M, M\rangle$. Precisely, if (θ', L') is another such pair, then $L \equiv L'$ and the stochastic integral $\int_0^\cdot (\theta_u - \theta'_u)^\top \, \mathrm{d}M_u$ is the zero martingale, which implies that $(\theta_u - \theta'_u)$ lies in the null space of $\mathrm{d}\langle M, M\rangle_u$ for $\mathrm{d}\langle M, M\rangle$-almost all (u, ω).

Moreover, θ is characterized by the covariation identity

$$\left\langle \widetilde{V}, M \right\rangle_t = \int_0^t \theta_u^\top \, \mathrm{d}\langle M, M\rangle_u \,, \qquad t \in [0, T], \tag{11}$$

where $\left\langle \widetilde{V}, M \right\rangle$ is the $(n+1)$-vector with entries $\left\langle \widetilde{V}, M^{(i)} \right\rangle$.

Proof (standard; sketch). In the Hilbert space $\mathcal{M}^2$ of square-integrable martingales with inner product $\langle N^{(1)}, N^{(2)}\rangle := \mathbb{E}_{\mathbb{Q}}[N_T^{(1)} N_T^{(2)}]$, the subspace

$$\mathcal{H}(M) := \left\{ \int_0^\cdot \vartheta_u^\top \, \mathrm{d}M_u \;:\; \vartheta \text{ predictable and square-integrable} \right\}$$

is closed. The orthogonal projection of $\widetilde{V} - \widetilde{V}_0$ onto $\mathcal{H}(M)$ yields the decomposition (10) with L orthogonal to $\mathcal{H}(M)$, hence to M. Uniqueness follows from the orthogonal decomposition, and (11) follows from taking covariation with M and using the covariation formula for stochastic integrals. □

The decomposition is not a theoretical embellishment; it gives a precise sense in which the chosen traded-factor set spans part of the exotic's risk. The residual martingale L is the orthogonal component that cannot be eliminated by any self-financing strategy trading only in the factors M. This is exactly the component that remains even when the hedger uses the best possible L^2 projection given the factor set.

Corollary 2.13 (Mean-square optimal hedging within the factor set (classical))**.** *Assume 2.11 and let $\widetilde{H}$ be a square-integrable discounted payoff at time T. Define the $\mathbb{Q}$-martingale $\widetilde{V}_t := \mathbb{E}_{\mathbb{Q}}[\widetilde{H} \mid \mathcal{F}_t]$. Let (θ, L) be its Galtchouk-Kunita-Watanabe decomposition with respect to M as in Theorem 2.12. Then among all self-financing strategies with discounted gains of the form $\int_0^T \vartheta_u^\top \, \mathrm{d}M_u$, the choice $\vartheta = \theta$ uniquely minimizes the mean-square hedging error*

$$\mathbb{E}_{\mathbb{Q}}\left[\left(\widetilde{H} - \widetilde{V}_0 - \int_0^T \vartheta_u^\top \, \mathrm{d}M_u\right)^2\right],$$

and the minimal value equals $\mathbb{E}_{\mathbb{Q}}[L_T^2]$.

Proof (standard; sketch). By (10), $\widetilde{H} - \widetilde{V}_0 - \int_0^T \vartheta_u^\top \, \mathrm{d}M_u = L_T + \int_0^T (\theta_u - \vartheta_u)^\top \, \mathrm{d}M_u$. Since L is orthogonal to $\mathcal{H}(M)$, the cross-term vanishes and the mean-square error equals $\mathbb{E}_{\mathbb{Q}}[L_T^2] + \mathbb{E}_{\mathbb{Q}}[(\int_0^T (\theta_u - \vartheta_u)^\top \, \mathrm{d}M_u)^2]$, which is minimized uniquely when $\vartheta = \theta$. □

It is useful to keep the two hedge constructions distinct. The first-order matched strategy in Section 2.6 is computed from a chosen *model* value function v via $\nabla_y v$, whereas the GKW strategy θ is the $L^2(\mathbb{Q})$ projection of the *true* martingale $\widetilde{V}$ onto the factor span generated by M. They coincide only in spanning regimes; in general, GKW clarifies intrinsic span limitations while Theorem 2.7 quantifies the pathwise residual generated by an implementable first-order matching rule.

Also, the GKW/mean-square optimality criterion is not tail control: minimizing L^2 error does not preclude large losses on rare paths, and it does not produce a worst-case (super-replicating) hedge. Tail-risk control requires a different objective or a superhedging formulation.

2.11 Specialization and benchmark checks

The previous results make it possible to state benchmark cases in which the residual vanishes for structural reasons. These benchmarks are useful because they isolate the sense in which residual terms arise from either span limitations of the factor set or mismatch between realized and model-implied covariation.

2.11.1 Vanilla-only benchmark: affine claims in traded coordinates

The simplest case is when the discounted claim value is affine in the traded coordinates. Then second derivatives vanish and the covariation-driven residual in Theorem 2.7 is identically zero.

Proposition 2.14 (Affine traded-coordinate claims are exactly spanned (standard))**.** *Assume 2.1 and let ξ be fixed. Suppose that, for some deterministic scalar a and deterministic vector $b \in \mathbb{R}^{n+1}$, the discounted value satisfies*

$$\widetilde{V}_t = a + b^\top M_t, \qquad t \in [0, T].$$

Then the self-financing strategy with holdings $\vartheta_t \equiv b$ replicates $\widetilde{V}$ exactly, the GKW residual satisfies $L \equiv 0$, and the hedging-error identity (4) holds with a zero right-hand side.

Proof (standard; sketch). Since $\widetilde{V}_t = a + b^\top M_t$ with constant a, b, the self-financing strategy $\vartheta_t \equiv b$ replicates $\widetilde{V}$ exactly ($\int b^\top \, \mathrm{d}M = b^\top (M_t - M_0) = \widetilde{V}_t - \widetilde{V}_0$). The residual L is zero, and the Hessian $D_{yy}^2 v$ vanishes, so both sides of the hedging error identity are zero. □

2.11.2 Monitoring dates and cashflow matching as a structural condition

For callable products with deterministic monitoring dates, Proposition 2.9 shows that one must also control the jump discrepancy term induced by state updates and contractual cashflows. This is not a technicality: neglecting it can create an apparent "hedging error" that is purely an artefact of inconsistent bookkeeping.

A clean way to prevent this is to impose an explicit cashflow-matching condition at monitoring dates, expressed as a jump condition for the value representation.

Assumption 2.15 (Cashflow-matched value representation)**.** On each monitoring date t_k, the representation v satisfies

$$v(t_k-, y, \xi^-) = v(t_k+, y, \xi^+) + \widetilde{c}_k(y, \xi^-),$$

where $\widetilde{c}_k$ is the discounted contractual cashflow paid at t_k when the pre-update state is ξ^- and the post-update state is ξ^+.

Proposition 2.16 (Elimination of jump discrepancies under cashflow matching (standard bookkeeping))**.** *Assume the setting of Proposition 2.9 and suppose in addition that Assumption 2.15 holds and that the portfolio pays exactly the contractual cashflow $\Delta C_{t_k} = \widetilde{c}_k(\widetilde{Y}_{t_k}, \xi_{t_k-})$ at each monitoring date. Then the jump discrepancy sum in* (9) *vanishes and the piecewise hedging error reduces to the sum of the continuous-interval covariation mismatch terms.*

Proof (standard; sketch). By Assumption 2.15 and the payment rule, $\Delta v(t_k) = -\widetilde{c}_k = -\Delta C_{t_k}$, so the jump discrepancy $\Delta v(t_k) + \Delta C_{t_k}$ vanishes at each monitoring date. □

Once monitoring-date cashflows are matched consistently, residual hedging error comes from only two conceptually distinct sources: covariation mismatch (Theorem 2.7) and span limitation of the chosen factor set (the nonzero orthogonal residual in the GKW decomposition). This separation matters later: feedback primarily distorts the covariation channel, while crowding and illiquidity can also reduce effective spanning power.

The next section introduces the endogenous feedback layer. The logical order is deliberate. The traded-coordinate hedging calculus and its projection refinements are established without any feedback assumptions. Feedback then enters as a modelling layer that alters the realized dynamics of traded factors and thereby alters the covariation objects that drive residual P&L in the identities proved above.

Before proceeding, we note a shift in mathematical tools. The reader may notice that Section 3 invokes fixed-point theorems and operator norms—tools from functional analysis rather than stochastic calculus. This is unavoidable: the question "how does aggregate hedging demand move prices?" is intrinsically a *simultaneous equations* problem (demand depends on prices, prices depend on demand), whose natural language is fixed-point theory. However, the fixed point is purely algebraic and operates *pointwise in time*; no infinite-dimensional function spaces appear. Once the fixed point is linearized, the output is a Jacobian matrix J_t that transforms fundamentals into traded prices. This matrix then feeds directly back into the covariation calculus via $\mathrm{d}\langle Y, Y\rangle_t = J_t\, \mathrm{d}\langle Z, Z\rangle_t\, J_t^\top$—an identity in Itô calculus, not topology. The two toolkits thus complement rather than compete: fixed-point theory supplies the amplification matrix, stochastic calculus quantifies its P&L consequence.

3 Feedback equilibrium and risk diagnostics

The traded-coordinate calculus in Section 2 is intentionally agnostic about how the joint covariation of traded factors is generated. This section introduces a parsimonious feedback layer: aggregate hedging demand affects traded prices through an impact channel, thereby altering the realized covariation of traded factors and potentially amplifying the residual terms identified in Theorem 2.7. We then translate the theoretical identities into implementable risk diagnostics.

3.1 Fixed-point feedback model

The model is built in a way that isolates the feedback mechanism as a fixed point. The key idea is that one should distinguish a "fundamental" (or pre-impact) traded state from the observed traded state that clears supply and demand in the presence of hedging flows. This yields an algebraic fixed point at each time, which is both mathematically tractable and economically interpretable as a market-clearing relation under linearized impact.

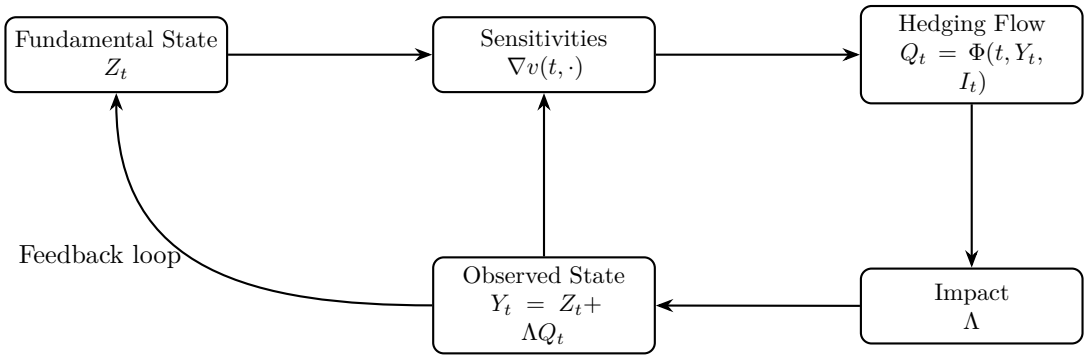

Figure 1: Feedback loop in the fixed-point model: the fundamental state Z drives sensitivities and hedging flow, impact Λ maps flow into price displacement, and the observed traded state Y feeds back into subsequent sensitivities.

3.2 Fundamental state, hedging flow, and an impact clearing relation

We work in discounted units and fix a dimension $d := n + 1$. The traded state vector is denoted by $Y_t \in \mathbb{R}^d$ and is interpreted as the vector of discounted traded prices used as coordinates in Section 2. In applications, Y_t may consist of discounted spot together with discounted prices of a finite set of liquid vanillas, or any other explicitly specified traded factorization.

The model posits an auxiliary adapted process $Z_t \in \mathbb{R}^d$, interpreted as a "fundamental" traded state that would prevail in the absence of hedging impact from the desk whose inventory is being hedged. We do not interpret Z as a latent variable to be estimated. It is simply a modelling device that allows one to write the observed traded state as the solution of a clearing relation.

Assumption 3.1 (Fundamental state)**.** The process $Z = (Z_t)_{t\in[0,T]}$ is an $\mathbb{R}^d$-valued continuous semimartingale adapted to $(\mathcal{F}_t)$.

Hedging activity is represented by a flow vector $Q_t \in \mathbb{R}^m$, where m is the dimension of the flow space one wishes to model. The only requirement for the analysis in this section is that Q_t is determined as a measurable function of time, the observed traded state Y_t, and an inventory/risk-limit variable I_t.

Assumption 3.2 (Inventory/risk-limit state)**.** The process $I = (I_t)_{t\in[0,T]}$ is adapted, *continuous*, and takes values in a metric space $\mathcal{I}$, with $(t, \omega) \mapsto I_t(\omega)$ jointly measurable. Moreover, I has pathwise finite variation on $[0, T]$. In particular, I contributes no quadratic covariation: $\langle I, I\rangle \equiv 0$ and $\langle Z, I\rangle \equiv 0$ for any continuous semimartingale Z.

Assumption 3.3 (Hedging flow map)**.** There exists a measurable function $\Phi : [0, T] \times \mathbb{R}^d \times \mathcal{I} \to \mathbb{R}^m$ such that

$$Q_t = \Phi(t, Y_t, I_t) \qquad \text{for all } t \in [0, T],$$

and such that $\Phi(t, \cdot, i)$ is globally Lipschitz uniformly in (t, i), meaning that there exists $L_\Phi \geq 0$ with

$$\|\Phi(t, y, i) - \Phi(t, y', i)\| \leq L_\Phi \|y - y'\| \qquad \text{for all } t \in [0, T],\ y, y' \in \mathbb{R}^d,\ i \in \mathcal{I}.$$

In applications, Φ may be specified directly (as a reduced-form demand rule) or induced by a hedging policy that targets selected components of the sensitivity vector $\nabla_y v(t, Y_t, \xi_t)$ in traded coordinates. In the latter case, the Jacobian $D_y\Phi$ is controlled by second-derivative blocks of v (notably spot–surface cross-curvatures), which is the precise sense in which "vanna-type" effects enter the feedback Jacobian; no such microfoundation is imposed by the abstract assumptions.

The impact channel is taken to be affine in the flow. This is the minimal structure needed to obtain a fixed point while remaining transparent about what is being assumed. The impact matrix $\Lambda \in \mathbb{R}^{d\times m}$ maps flow into traded price displacement.

Assumption 3.4 (Linear impact clearing relation)**.** There exists a deterministic matrix $\Lambda \in \mathbb{R}^{d\times m}$ such that the observed traded state Y satisfies

$$Y_t = Z_t + \Lambda Q_t \qquad \text{for all } t \in [0, T]. \tag{12}$$

Combining (12) with $Q_t = \Phi(t, Y_t, I_t)$ yields the fixed-point equation

$$Y_t = Z_t + \Lambda\Phi(t, Y_t, I_t), \qquad t \in [0, T]. \tag{13}$$

This is the feedback loop in its simplest explicit form. The left-hand side is the traded state that determines sensitivities and hedging demand; the right-hand side is the state that results after the demand is applied to a market with impact.

3.3 Existence, uniqueness, and adaptedness of the fixed point

We now establish existence and uniqueness of a progressively measurable solution Y to (13). Since the fixed point is pointwise in time, the argument is an application of the contraction mapping theorem at each time, together with a measurability argument that ensures the solution inherits adaptedness from (Z, I).

Assumption 3.5 (Contraction strength)**.** The Lipschitz constant and the impact satisfy $\|\Lambda\|\, L_\Phi < 1$, where $\|\Lambda\|$ denotes the operator norm induced by the Euclidean norm.

Theorem 3.6 (Well-posed feedback equilibrium (standard; contraction mapping))**.** *Assume 3.1, 3.2, 3.3, 3.4, and 3.5. Then there exists a unique adapted process $Y = (Y_t)_{t\in[0,T]}$ satisfying* (13) *for all t. Moreover, for each t the map $(z, i) \mapsto y$ defined by the unique solution of*

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

is globally Lipschitz in z, uniformly in (t, i), and Y_t is an $\mathcal{F}_t$-measurable function of (Z_t, I_t).

Proof (standard; sketch). The map $T_{t,z,i}(y) := z + \Lambda\Phi(t, y, i)$ is a contraction since $\|\Lambda D_y\Phi\| \le \|\Lambda\|L_\Phi < 1$. By the Banach fixed-point theorem, there exists a unique fixed point $y^*(t, z, i)$ for each (t, z, i), which is progressively measurable since the contraction is uniform. □

Theorem 3.6 provides a clean, minimal notion of feedback equilibrium. It isolates the endogenous component of traded-state variation as the consequence of a deterministic market-clearing relation (12) coupled with a state-dependent hedging rule Φ. The next results show how the fixed point yields a stability criterion and an amplification operator that can be linked directly to covariation objects in Section 2.

Assumption 3.5 is a sufficient solver and well-posedness condition, but the operational stress boundary is often conditioning: linear response and covariation distortion are governed by

$$A_t := I_d - \Lambda D_y\Phi(t, Y_t, I_t), \qquad J_t := A_t^{-1},$$

so amplification grows as A_t approaches singularity, even before any global contraction argument is invoked.

3.4 Differentiability, linearization, and amplification

To study stability it is natural to ask how a small perturbation of the fundamental state Z_t propagates to the observed traded state Y_t. This yields an amplification operator whose norm quantifies feedback strength. The cleanest derivation uses the implicit function theorem, which requires differentiability of the flow map in the traded state.

Assumption 3.7 (Differentiable flow map)**.** For each $(t, i) \in [0, T] \times \mathcal{I}$, the map $y \mapsto \Phi(t, y, i)$ is continuously differentiable with Jacobian $D_y\Phi(t, y, i) \in \mathbb{R}^{m\times d}$. Moreover, $\sup_{t,y,i} \|D_y\Phi(t, y, i)\| \leq L_\Phi$.

Proposition 3.8 (Invertibility and amplification matrix (standard))**.** *Assume the hypotheses of Theorem 3.6 and Assumption 3.7. Fix (t, ω) and write $Y_t = Y_t(\omega)$, $Z_t = Z_t(\omega)$, $I_t = I_t(\omega)$. Then the matrix*

$$A_t := I_d - \Lambda\, D_y\Phi(t, Y_t, I_t) \tag{14}$$

is invertible and satisfies $\|A_t^{-1}\| \leq (1 - \|\Lambda\|L_\Phi)^{-1}$. Furthermore, if Z_t is perturbed to $Z_t + \delta z$ with I_t fixed, the corresponding perturbed equilibrium $Y_t + \delta y$ satisfies, to first order,

$$\delta y = A_t^{-1}\, \delta z + o(\|\delta z\|). \tag{15}$$

Proof (standard; sketch). The map $F(y) = y - \Lambda\Phi(t, y, I_t)$ has derivative $DF(y) = I_d - \Lambda D_y\Phi$. Under the contraction assumption $\|\Lambda\|L_\Phi < 1$, DF is invertible via Neumann series with bounded inverse. The implicit function theorem yields the local linearization $\delta y = DF(Y_t)^{-1}\delta z = A_t^{-1}\delta z$. □

Proposition 3.8 formalizes, in the present reduced-form clearing model, a stability statement often expressed informally as "feedback is dangerous when the product of impact and hedging sensitivity is large." Here the relevant object is the amplification matrix $A_t^{-1} = (I_d - \Lambda D_y\Phi)^{-1}$. Its norm controls how perturbations of the fundamental state are magnified in observed traded prices. Importantly, the criterion is not a narrative assertion: it follows mechanically from the fixed-point structure and the Lipschitz bound.

The fixed-point layer is inherently multi-channel: Y is a vector (spot plus chosen vanilla coordinates), and Φ can encode simultaneous delta, vega, skew, and cross-maturity rebalancing. Linearization makes this explicit via the loop-gain matrix $B_t := \Lambda D_y\Phi(t, Y_t, I_t)$; in stable regimes one has the Neumann-series expansion

$$J_t = (I_d - B_t)^{-1} = \sum_{k=0}^{\infty} B_t^k,$$

where B_t^k corresponds to the k-th feedback "round" and off-diagonal blocks encode cross-channel couplings.

3.5 Feedback-induced covariation distortion

The connection to Section 2 becomes clearest once we translate amplification into a statement about quadratic covariation. Since Z is assumed to be a continuous semimartingale, and Y_t is obtained as a pointwise function of (t, Z_t, I_t), the process Y is itself a continuous semimartingale under mild additional smoothness. Its quadratic covariation can then be expressed in terms of that of Z and the Jacobian of the mapping $z \mapsto y^*(t, z, i)$.

To keep the argument self-contained and free of hidden regularity assumptions, we record the required smoothness of the fixed-point map as a classical proposition (Banach fixed point theorem + implicit function theorem).

Proposition 3.9 (Smoothness of the feedback fixed-point map (classical))**.** *Fix $(t,i) \in [0,T] \times \mathcal{I}$. Assume that the map $y \mapsto \Phi(t,y,i)$ is C^2 with bounded derivatives and that there exists $\kappa \in (0,1)$ such that*

$$\sup_{y \in \mathbb{R}^d} \|\Lambda D_y \Phi(t,y,i)\| \leq \kappa.$$

Then for each $z \in \mathbb{R}^d$ there exists a unique $y^*(t,z,i) \in \mathbb{R}^d$ solving

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

and the map $z \mapsto y^*(t,z,i)$ is of class C^2. Moreover,

$$\frac{\partial y^*}{\partial z}(t,z,i) = \big(I_d - \Lambda D_y \Phi(t, y^*(t,z,i), i)\big)^{-1}.$$

Finally, for any directions $h_1, h_2 \in \mathbb{R}^d$,

$$\begin{aligned} D^2_{zz} y^*(t,z,i)[h_1,h_2] = {} & \big(I_d - \Lambda D_y \Phi(t,y^*,i)\big)^{-1} \Lambda \\ & \times D^2_{yy} \Phi(t,y^*,i) \left[\frac{\partial y^*}{\partial z}(t,z,i) h_1,\ \frac{\partial y^*}{\partial z}(t,z,i) h_2 \right]. \end{aligned} \tag{16}$$

Proof (compact; classical). Define $F(y,z) := y - z - \Lambda \Phi(t,y,i)$. Then $F \in C^2$, and solutions of $F(y,z) = 0$ are exactly fixed points of $y \mapsto z + \Lambda \Phi(t,y,i)$. The bound $\sup_y \|\Lambda D_y \Phi(t,y,i)\| \leq \kappa < 1$ implies this map is a contraction, hence for each z there is a unique solution $y^*(t,z,i)$ (Banach fixed-point theorem).

Moreover,

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

and the same bound yields that $D_y F(y,z)$ is invertible everywhere with uniformly bounded inverse (Neumann series). Therefore, by the implicit function theorem, the unique solution $y^*(t,\cdot,i)$ depends C^2-smoothly on z.

Differentiating $F(y^*(t,z,i),z) = 0$ gives

$$\big(I_d - \Lambda D_y \Phi(t, y^*(t,z,i), i)\big) \frac{\partial y^*}{\partial z}(t,z,i) = I_d,$$

hence $\partial_z y^* = (I_d - \Lambda D_y \Phi)^{-1}$. Differentiating once more and using $D^2_{yy} F = -\Lambda D^2_{yy} \Phi$ yields (16). □

When Φ or the equilibrium map is non-smooth (hard risk limits, digital/barrier kinks, discrete state updates), $z \mapsto y^*(t,z,i)$ is at best piecewise smooth and (17) can acquire additional singular-set terms (local-time-type contributions). The present analysis describes the smooth regime; the analogous non-smooth issue on the hedging-identity side is discussed in Appendix C.

Proposition 3.10 (Quadratic covariation under feedback (standard; Itô/chain rule))**.** *Assume 3.1, 3.2, 3.4, 3.3, 3.5, and that the hypotheses of Proposition 3.9 hold (uniformly in (t, I_t) along the realized path). Let Y be the unique equilibrium process from Theorem 3.6. Define the Jacobian*

$$J_t := \frac{\partial y^*}{\partial z}(t, Z_t, I_t) \in \mathbb{R}^{d \times d}.$$

Then Y is a continuous semimartingale and its quadratic covariation satisfies

$$\mathrm{d}\,\langle Y, Y \rangle_t = J_t\ \mathrm{d}\,\langle Z, Z \rangle_t\ J_t^\top. \tag{17}$$

Moreover, J_t coincides with the amplification matrix in Proposition 3.8:

$$J_t = \big(I_d - \Lambda D_y \Phi(t, Y_t, I_t)\big)^{-1}. \tag{18}$$

Proof (standard; sketch). Since $Y_t = f(t, Z_t, I_t)$ with f being C^2 in z, Itô's formula componentwise gives $\mathrm{d}Y_t = \partial_z f \, \mathrm{d}Z_t + (\text{FV terms})$. Quadratic covariation depends only on local martingale parts, so $\mathrm{d}\langle Y, Y\rangle_t = (\partial_z f) \, \mathrm{d}\langle Z, Z\rangle_t \, (\partial_z f)^\top$. Differentiating the fixed-point relation identifies $\partial_z f$ as $J_t = (I - \Lambda D_y \Phi)^{-1}$. □

If exogenous inputs (e.g. I) or the flow rule jump so that Y is càdlàg, then $\langle Y, Y\rangle$ includes the additional jump contribution $\sum_{0<u\le t} \Delta Y_u (\Delta Y_u)^\top$ on top of the continuous Jacobian term; the traded-coordinate identity similarly acquires the explicit jump correction recorded in Theorem C.1.

Proposition 3.10 is the precise bridge between the feedback layer and the traded-coordinate calculus. It shows that feedback induces a systematic distortion of quadratic covariation: the observed covariation of traded coordinates is the fundamental covariation conjugated by an amplification matrix. This is exactly the mechanism by which feedback enters the hedging-error identity in Theorem 2.7, because that identity depends on the realized covariation $\mathrm{d}\langle Y, Y\rangle$. The next subsection records this connection explicitly.

3.6 Connection to the traded-coordinate hedging error

To connect the layers without introducing any hidden assumptions, we emphasize the logical direction. Theorem 2.7 expresses residual P&L in terms of $\mathrm{d}\langle Y, Y\rangle$ and a model-implied covariation density $\widehat{c}$. Proposition 3.10 expresses $\mathrm{d}\langle Y, Y\rangle$ in terms of $\mathrm{d}\langle Z, Z\rangle$ and the amplification matrix J. Therefore, under feedback, the residual term can be rewritten as the sum of a "fundamental mismatch" component and a "feedback distortion" component.

If the hedger computes the model-implied covariation $\widehat{c}$ under a benchmark that corresponds to the fundamental state Z rather than the observed state Y, then one should expect systematic mismatch even when the benchmark is otherwise reasonable. The point is not that the benchmark is wrong in an absolute sense; it is that the benchmark ignores the equilibrium mapping $Z \mapsto Y$, and hence it ignores the conjugation of covariation by J. This provides a mathematically precise interpretation of a common practitioner concern: a model that is adequate in a price-taking regime may become inadequate when hedging alters the traded surface itself.

The remainder of the paper develops two consequences of this observation. The first is diagnostic: one can define feedback-sensitive risk measures and instability proxies in terms of J and covariation distortions without inserting any narrative magnitudes. The second is empirical: one can specify testable implications in terms of observable covariations of traded prices and their regime dependence. These developments are carried out in Sections 3.7 and 4.

In this section, $\mathbb{Q}$ should be read as the desk's pricing convention (used to express discounted traded prices as local martingales), not as a claim that the feedback-distorted dynamics necessarily admit an ELMM. This does not affect the identities, which are P&L accounting statements valid under any measure; what changes under feedback is interpretation: the residual is tracking error relative to the model convention, not deviation from a unique "true" arbitrage-free price.

3.7 Implementable risk diagnostics

We now translate the theoretical identities into quantities that can be computed from traded data and used in stress testing. The aim is not to propose a universal index of feedback risk, but to define a small set of diagnostics whose meaning follows directly from the theory. The guiding principle is that any quantity introduced here must either be defined purely in terms of traded semimartingale objects, or be explicitly labelled as a model input when not directly observable.

The traded-coordinate hedging-error identity in Section 2.6 shows that, once first-order exposures are hedged in a chosen traded factor basis, residual P&L is governed by a contraction of the value Hessian with the discrepancy between realized and model-implied quadratic covariation. The feedback model in Section 3 then shows how such discrepancies can be created or amplified

endogenously through the amplification operator J. Both observations suggest diagnostics based on three ingredients: a choice of traded factor vector, an estimate of realized covariation of that vector, and a model-implied covariation density (or a benchmark) used to compute the hedge.

Throughout this section we work in discounted units. We therefore write $\widetilde{Y}$ for the discounted traded-factor vector, so that covariation objects such as $\mathrm{d}\left\langle \widetilde{Y}, \widetilde{Y} \right\rangle$ refer to *discounted* prices. This keeps the diagnostics aligned with the hedging-error identities in Section 2, which are stated in discounted coordinates.

3.8 Choice of traded factor vector and reporting discipline

We begin by making explicit what must be fixed before any diagnostic can be defined. A diagnostic is only meaningful relative to a chosen factor representation. For example, one may use the vector consisting of discounted spot together with a small set of option prices, or one may use a parametric representation of the surface backed by a finite set of liquid instruments. Both choices can be consistent with the theory, but they answer different questions. A factor set that is too small risks leaving most of the exotic's risk in the orthogonal residual component from Theorem 2.12; a factor set that is too large risks introducing illiquid instruments whose traded prices are not reliably observable at the required frequency.

To prevent hidden choices, we impose a reporting discipline. Any empirical or stress-testing exercise must specify the factor set in a manner that is reproducible. This means listing the exact instruments, maturities, strikes, and quotes used, as well as the time grid on which the factors are sampled. The empirical section operationalizes this discipline.

3.9 Covariation-based residual diagnostic

We now define a residual diagnostic that corresponds directly to the integrand in Theorem 2.7. The definition is made at the level of an *estimand*; how it is estimated from data is deferred to Section 4.6.

Let $\widetilde{Y}$ denote the chosen discounted traded factor vector and let $\widehat{c}$ denote the model-implied covariation density used in the hedging rule. For a fixed contract state ξ and a fixed time interval $[t_0, t_1]$ on which the state is constant and cashflows are matched as in Assumption 2.15, Theorem 2.7 shows that the hedging error over the interval is

$$\widetilde{X}_{t_1} - \widetilde{V}_{t_1} = -\frac{1}{2}\int_{t_0}^{t_1} \mathrm{Tr}\Big(D_{yy}^2 v(u, \widetilde{Y}_u, \xi)\,\big(\mathrm{d}\left\langle \widetilde{Y}, \widetilde{Y} \right\rangle_u - \widehat{c}(u, \widetilde{Y}_u, \xi)\,\mathrm{d}u\big)\Big).$$

This motivates the following pointwise covariation mismatch functional.

Definition 3.11 (Instantaneous covariation mismatch functional)**.** Fix a contract state ξ and a value function $v(\cdot, \cdot, \xi)$ used for hedging. Define the matrix-valued signed measure on $[0, T]$ by

$$\mu(\,\mathrm{d}u) := \mathrm{d}\left\langle \widetilde{Y}, \widetilde{Y} \right\rangle_u - \widehat{c}(u, \widetilde{Y}_u, \xi)\,\mathrm{d}u.$$

The instantaneous mismatch functional is the scalar signed measure

$$\mathcal{M}(\,\mathrm{d}u) := \mathrm{Tr}(D_{yy}^2 v(u, \widetilde{Y}_u, \xi)\,\mu(\,\mathrm{d}u)).$$

When $\mathrm{d}\left\langle \widetilde{Y}, \widetilde{Y} \right\rangle$ admits a predictable Lebesgue density, i.e. $\mathrm{d}\left\langle \widetilde{Y}, \widetilde{Y} \right\rangle_u = c^{\mathrm{real}}(u)\,\mathrm{d}u$ (as in Itô-process settings), one may identify the mismatch density $\mu(u) := c^{\mathrm{real}}(u) - \widehat{c}(u, \widetilde{Y}_u, \xi)$ and the pointwise functional $\mathcal{M}(u) := \mathrm{Tr}(D_{yy}^2 v(u, \widetilde{Y}_u, \xi)\,\mu(u))$, so that $\mathcal{M}(\,\mathrm{d}u) = \mathcal{M}(u)\,\mathrm{d}u$.

The functional $\mathcal{M}$ is, by construction, the integrand appearing in the hedging-error identity. It is not itself a "risk measure" because it may take either sign. A risk measure is obtained by

aggregating $\mathcal{M}$ in a way consistent with the intended use. For stress testing, a natural choice is the total variation or absolute integral of $\mathcal{M}$ over an interval.

Viewed abstractly, the residual term is a linear functional of the covariation-mismatch measure: with $H_t := D^2_{yy} v(t, \widetilde{Y}_t, \xi)$ one has $\mu \mapsto \int \mathrm{Tr}(H_t \, \mu(\,\mathrm{d}t))$. In the absolutely continuous L^2 case this is the usual Hilbert-space pairing, with H as the Riesz representer (and a Fréchet derivative $-\frac{1}{2}H$ for the hedging-error functional); singular components of μ are handled in the same measure-pairing sense.

Definition 3.12 (Covariation-residual magnitude over an interval)**.** Fix an interval $[t_0, t_1] \subset [0, T]$ on which the contract state is constant and cashflows are matched. Recall from Definition 3.11 that $\mathcal{M}$ is a scalar *signed* measure on $[0, T]$. Let $|\mathcal{M}|$ denote its **total variation measure**. Define the covariation-residual magnitude by

$$\mathsf{R}_{t_0,t_1} := \frac{1}{2} \int_{(t_0,t_1]} |\mathcal{M}|(\,\mathrm{d}u).$$

When μ admits a density with respect to Lebesgue measure, this becomes

$$\mathsf{R}_{t_0,t_1} = \frac{1}{2} \int_{t_0}^{t_1} \Big| \mathrm{Tr}\Big(D^2_{yy} v(u, Y_u, \xi) \, (c^{\mathrm{real}}(u) - \widehat{c}(u, Y_u, \xi)) \Big) \Big| \; \mathrm{d}u,$$

where $c^{\mathrm{real}}(u)$ denotes the instantaneous realized covariation density of Y.

The quantity R_{t_0,t_1} is intended as a diagnostic: it summarizes how large the covariation mismatch term could be over the interval, given the Hessian weighting induced by the hedger's model value function. In empirical work one replaces the infinitesimal objects by realized covariation estimators over discrete sampling windows; this is handled later.

3.10 Feedback amplification diagnostic

The fixed-point feedback model in Section 3 implies that, under impact clearing, the observed covariation is a conjugation of the fundamental covariation by the amplification matrix J_t. This suggests a diagnostic based on the norm of J_t or of the amplification matrix $A_t^{-1} = (I - \Lambda D_y \Phi)^{-1}$.

Because J_t is not directly observable without specifying Φ and Λ, it should be treated as a model-based diagnostic. Nevertheless, it provides a disciplined way to quantify feedback strength *conditional* on a chosen impact and hedging rule specification. This is useful for stress testing because it cleanly separates what is being assumed (a linear impact map and a differentiable flow map) from what is being computed (the implied amplification).

Definition 3.13 (Amplification operator and stability margin)**.** In the fixed-point model of Section 3, define

$$B_t := \Lambda D_y \Phi(t, Y_t, I_t), \qquad A_t := I_d - B_t, \qquad J_t := A_t^{-1}.$$

The stability margin is defined as the smallest singular value of A_t,

$$\mathsf{SM}_t := \sigma_{\min}(A_t),$$

and the amplification index is defined by

$$\mathsf{AI}_t := \|J_t\|_2 = \frac{1}{\sigma_{\min}(A_t)},$$

Under the contraction assumption $\|\Lambda\| L_\Phi < 1$, we have $\|B_t\| < 1$ and hence A_t is invertible with $\mathsf{SM}_t > 0$ and $\mathsf{AI}_t < \infty$. Moreover, the Neumann-series bound yields the useful inequalities

$$\mathsf{SM}_t \geq 1 - \|B_t\|, \qquad \mathsf{AI}_t \leq \frac{1}{1 - \|B_t\|}.$$

The meaning of AI_t is immediate from Proposition 3.8: it controls the linear response of the observed traded state to perturbations of the fundamental state. It also controls, through Proposition 3.10, the distortion of quadratic covariation.

The theoretical link between these diagnostics and the hedging error is captured by the following proposition, which simply combines Theorem 2.7 with Proposition 3.10.

Proposition 3.14 (Residual term under feedback-induced covariation distortion (standard corollary))**.** *Assume the hypotheses of Theorem 2.7 for the discounted traded factor vector $\widetilde{Y}$ and the hypotheses of Proposition 3.10 for the feedback equilibrium (identifying Y in that section with $\widetilde{Y}$). Let Z be the fundamental discounted state. Let $\widehat{c}^{\,Z}$ be a reference fundamental covariation density (in discounted units). Then on any interval where the contract state is constant and cashflows are matched, the instantaneous covariation mismatch admits the decomposition*

$$\mathrm{d}\left\langle \widetilde{Y},\widetilde{Y}\right\rangle_t - \widehat{c}(t,\widetilde{Y}_t,\xi)\,\mathrm{d}t = \Big(J_t\ \mathrm{d}\,\langle Z,Z\rangle_t\ J_t^\top - \widehat{c}^{\,Z}(t,Z_t,\xi)\,\mathrm{d}t\Big) + \Big(\widehat{c}^{\,Z}(t,Z_t,\xi) - \widehat{c}(t,\widetilde{Y}_t,\xi)\Big)\,\mathrm{d}t.$$

In particular, if the model-implied covariation density corresponds to the fundamental density in the sense that

$$\widehat{c}(t,\widetilde{Y}_t,\xi) = \widehat{c}^{\,Z}(t,Z_t,\xi) \quad \text{a.s. for all } t,$$

then the second term vanishes and the mismatch is driven purely by feedback distortion:

$$\mathrm{d}\left\langle \widetilde{Y},\widetilde{Y}\right\rangle_t - \widehat{c}(t,\widetilde{Y}_t,\xi)\,\mathrm{d}t = J_t\ \mathrm{d}\,\langle Z,Z\rangle_t\ J_t^\top - \widehat{c}^{\,Z}(t,Z_t,\xi)\,\mathrm{d}t.$$

Consequently, the covariation mismatch term appearing in (4) *is the sum of a fundamental mismatch component and a feedback distortion component.*

Proof. By Proposition 3.10, the realized quadratic covariation satisfies $\mathrm{d}\left\langle \widetilde{Y},\widetilde{Y}\right\rangle_t = J_t\ \mathrm{d}\,\langle Z,Z\rangle_t\ J_t^\top$. Subtracting $\widehat{c}(t,\widetilde{Y}_t,\xi)\,\mathrm{d}t$ from both sides, we obtain

$$\mathrm{d}\left\langle \widetilde{Y},\widetilde{Y}\right\rangle_t - \widehat{c}(t,\widetilde{Y}_t,\xi)\,\mathrm{d}t = J_t\ \mathrm{d}\,\langle Z,Z\rangle_t\ J_t^\top - \widehat{c}(t,\widetilde{Y}_t,\xi)\,\mathrm{d}t.$$

By the stated correspondence between $\widehat{c}(t,\widetilde{Y}_t,\xi)$ and $\widehat{c}^{\,Z}(t,Z_t,\xi)$, the second term equals $\widehat{c}^{\,Z}(t,Z_t,\xi)\,\mathrm{d}t$, which yields the displayed identity. The final sentence follows by inserting this identity into the trace contraction in Theorem 2.7. □

The correspondence condition $\widehat{c}(t,\widetilde{Y}_t,\xi) = \widehat{c}^{\,Z}(t,Z_t,\xi)$ is strong and should be read as an idealized separation between a "fundamental" benchmark and the feedback-distorted observed state. It is natural if the desk keeps using a baseline price-taking model (so $\widehat{c}$ is computed as if $\widetilde{Y} = Z$), and it is only approximate if $\widehat{c}$ is inferred directly from observed covariation in $\widetilde{Y}$ without modelling the $Z \mapsto Y$ map. Without imposing correspondence one can always decompose the mismatch into a pure feedback-distortion term plus a pure model-evaluation shift term by adding and subtracting $\widehat{c}^{\,Z}(t,Z_t,\xi)\,\mathrm{d}t$ as in Proposition 3.14.

3.11 Effective correlation and covariation distortion

Many empirical discussions of hedging feedback focus on changes in "spot–vol correlation". In the present framework, the corresponding object is a correlation computed from quadratic covariation of *specified coordinates* (spot and whatever traded or proxy factors are used to represent surface movements). The point is not to privilege one proxy over another, but to ensure that any reported correlation change is attached to an explicit factor definition, consistent with the reporting discipline of Section 3.8.

Definition 3.15 (Effective correlation and correlation shift)**.** Fix two scalar components U and V of a traded factor vector (or of an explicitly stated proxy factor vector). For an interval $[t_0, t_1] \subset [0, T]$ such that $\langle U \rangle_{t_1} - \langle U \rangle_{t_0} > 0$ and $\langle V \rangle_{t_1} - \langle V \rangle_{t_0} > 0$, define the quadratic-variation correlation

$$\rho_{U,V}[t_0, t_1] := \frac{\langle U, V \rangle_{t_1} - \langle U, V \rangle_{t_0}}{\sqrt{(\langle U \rangle_{t_1} - \langle U \rangle_{t_0})(\langle V \rangle_{t_1} - \langle V \rangle_{t_0})}}.$$

If a baseline or fundamental state Z is specified for the same coordinates, define $\rho^Z_{U,V}[t_0, t_1]$ analogously and define the correlation shift

$$\mathsf{CS}_{U,V}[t_0, t_1] := \rho_{U,V}[t_0, t_1] - \rho^Z_{U,V}[t_0, t_1].$$

Under the baseline benchmark of Proposition 2.10, spot–put covariation is strictly negative in one-factor price-taking models, hence any sustained movement of $\rho_{U,V}$ toward zero or to the opposite sign must come from leaving that benchmark regime. Proposition 3.10 provides one disciplined mechanism: feedback distorts covariation by conjugation through the amplification matrix J_t, which can attenuate or reverse correlations in observed traded coordinates even when the underlying fundamental covariation has a fixed sign.

3.12 A reproducible stress-testing protocol

The preceding definitions suggest a stress-testing protocol that is both implementable and faithful to the theory. We record it here at the level of a mathematical procedure, postponing all numerical results to Section 4.

A stress test begins by fixing the traded factor vector $\widetilde{Y}$ and the hedging value function v used to compute first-order matched holdings. It then fixes the benchmark covariation density $\widehat{c}$ implicit in the hedging model. Realized covariation of $\widetilde{Y}$ over a sampling grid is estimated using a stated estimator, producing an empirical analogue of the signed measure $\mu = \mathrm{d}\left\langle \widetilde{Y}, \widetilde{Y} \right\rangle - \widehat{c}\,\mathrm{d}t$. The stress test reports interval-wise summaries of the mismatch functional $\mathcal{M}$ from Definition 3.11, together with confidence statements derived from the estimator's sampling variability. If the feedback model is used for scenario analysis, the test additionally fixes the impact matrix Λ and the flow map Φ and computes the implied amplification index AI_t and the associated covariation distortion. The resulting scenario is then mapped back into the hedging-error identity to quantify the contribution of feedback-induced covariation changes.

The protocol separates what is measured (covariation objects of traded prices) from what is posited (optional scenario evaluation under a clearly stated feedback specification), so the reported quantities are mechanically determined once the inputs are fixed.

The next section implements this discipline by specifying the empirical estimands, estimators, inference methods, and robustness checks. Only after these elements are fixed do numerical results appear.

4 Empirical evidence

This section specifies how the theoretical objects introduced earlier can be confronted with data, interprets prior decompositions through the lens of our framework, and reports empirical evidence. We fix estimands, estimators, and inference machinery before reporting numerical output; results appear in Section 4.9.

4.1 Empirical protocol

The empirical protocol is organized as follows. We first state what can and cannot be inferred from covariation diagnostics alone (Sections 4.2 and 4.3). We then specify the factor dataset

and preprocessing rules (Section 4.4), define the continuous-time estimands to be estimated from discrete observations (Section 4.5), and record estimators and their convergence guarantees under standard semimartingale assumptions (Section 4.6 and Theorems 4.2 and 4.3). Finally, we map estimated covariation objects into the diagnostics defined in Section 3.7 and report results for the Nikkei Uridashi and KOSPI episodes (Section 4.9).

4.2 Limitations of the empirical evidence

The covariation-based approach has structural limitations.

Data frequency is one such limitation. Daily data cannot identify intraday covariation at high resolution, while feedback loops can operate over minutes or hours (margining, algorithmic re-hedging, and intraday liquidity shocks). A daily realized-covariation diagnostic can therefore detect a regime shift only after intraday dynamics have already produced large drawdowns. Extending the empirical protocol to intraday data is conceptually straightforward (the estimands are quadratic covariation objects) but requires microstructure-robust estimators (e.g. realized kernels, pre-averaging, or other noise-robust quadratic-variation methods) and careful treatment of market open/close effects.

Tradability is another limitation: observed volatility indices or proxies are often non-traded, so they enter only as auxiliary diagnostics unless explicitly mapped to traded instruments, and hedging/replication interpretations should be correspondingly restricted.

Factor choice is equally consequential; see Section 3.8 for reporting discipline and Theorem 2.12 for the span-limitation viewpoint.

More broadly, baseline-versus-stress comparisons carry an inherent joint-hypothesis interpretation: a change in effective correlation can be caused by endogenous feedback, by an exogenous shift in fundamentals, or by both. The paper therefore treats the feedback layer as a scenario mechanism consistent with observed covariation distortions rather than as a uniquely identified explanation absent additional instruments or flow data.

4.3 Reporting discipline and scope of empirical claims

The paper maintains a strict separation between theorem-level statements and empirical findings. Numerical magnitudes and threshold claims are not stated in the theoretical development unless they are produced by a reproducible test whose data, preprocessing, estimator, and uncertainty quantification are fully documented; such material is confined to the empirical section and appendices. Contextual claims about institutional settings are presented only as attributed descriptions drawn from sources and are not extrapolated beyond their scope. Pricing and replication statements are made under a risk-neutral measure $\mathbb{Q}$ when invoked, while empirical estimation is framed under the physical measure $\mathbb{P}$ with explicit estimands and estimators; statistical indices are not treated as tradable assets unless they are explicitly mapped to traded proxies. Throughout, we do not conflate self-financing–a definition of the gain process of a strategy–with first-order matching rules such as setting holdings equal to gradients in a chosen coordinate representation; the two notions are kept distinct in definitions and proofs.

The theoretical identities depend on quadratic covariation of traded coordinates. In continuous-time models these objects are well-defined pathwise for semimartingales. In discrete data, they are approached by realized covariation estimators. The principal goal of this section is therefore to define a discrete-time analogue of the covariation mismatch functional in Definition 3.11 and to provide the mathematical guarantees under which the estimator converges to the target object. These guarantees are standard in the econometrics of high-frequency semimartingales, but we reproduce the key arguments explicitly because they are the logical bridge between the theory and any empirical statement.

4.4 Data and preprocessing

The empirical protocol requires a time series for an explicitly specified factor vector. In the simplest implementation, one may take $Y_t = (\widetilde{S}_t, \widetilde{P}_t)$ where $\widetilde{S}$ is discounted spot and $\widetilde{P}$ is a discounted volatility-related factor; if the second coordinate is a volatility index, it should be treated as an observed *proxy* (unless explicitly mapped to traded instruments), so conclusions drawn from it are diagnostic rather than hedging/replication claims.

The reproducibility package accompanying this paper includes two daily series that instantiate this framework: a Nikkei index level series and a Nikkei volatility series. The precise column mapping, date range, and inclusion criteria are recorded in the code that generates all figures and tables in Section 4.9. The protocol below is written so that it applies unchanged to any other market once an analogous factor dataset is provided.

4.4.1 Sampling grid and returns

Let $(t_k)_{k=0,\dots,N}$ be the observation times, strictly increasing, and let $\Delta_k Y := Y_{t_k} - Y_{t_{k-1}}$ denote increments. For a positive-valued price series S we define log returns by

$$r_k := \log S_{t_k} - \log S_{t_{k-1}}.$$

When working with a multi-factor vector Y, we work directly with increments $\Delta_k Y$ rather than returns, because quadratic covariation is naturally defined on levels for semimartingales. In daily data, the distinction between using increments of levels and increments of logs is typically immaterial for convergence arguments at high frequency, but the present protocol keeps the object aligned with the continuous-time definition.

4.4.2 Alignment and missing observations

The realized covariation estimator requires synchronous observations across the factor components. If the dataset has missing dates or asynchronous sampling across components, the protocol aligns series on the intersection of observation dates and records the exclusion rule. Alternative asynchronous estimators are available in the literature, but we do not invoke them without explicitly stating the estimator and its assumptions.

4.5 Estimands: quadratic covariation and covariation mismatch

We now fix the continuous-time target object. Let $Y = (Y^1, \dots, Y^d)$ be a continuous d-dimensional semimartingale on $[0, T]$ on a filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t), \mathbb{P})$. We emphasize that $\mathbb{P}$ is the physical measure relevant for observed data; no pricing interpretation is attached to $\mathbb{P}$ in this section.

The quadratic covariation matrix process $\langle Y, Y\rangle$ has entries $\langle Y^i, Y^j\rangle$. The object that appears in the theoretical residual is not $\langle Y, Y\rangle$ itself but its deviation from a benchmark covariation density $\widehat{c}(t, Y_t)$ used in the hedging model. This motivates the estimand

$$\mu(\mathrm{d}t) := \mathrm{d}\langle Y, Y\rangle_t - \widehat{c}(t, Y_t)\,\mathrm{d}t,$$

and the scalar contraction with the Hessian weights

$$\mathcal{M}(\mathrm{d}t) := \mathrm{Tr}\big(H_t\,\mu(\mathrm{d}t)\big), \qquad H_t := D^2_{yy} v(t, Y_t, \xi_t),$$

where v is the hedger's value function used to compute first-order matched holdings. In empirical work, H_t is either computed from a specified model calibrated to observed data, or replaced by a sensitivity proxy. The protocol remains valid in both cases, but the interpretation changes: if H_t is model-derived, then $\mathcal{M}$ is the model-weighted covariation mismatch; if H_t is a proxy, then $\mathcal{M}$ is a diagnostic statistic rather than a hedging P&L predictor.

4.6 Estimators: realized quadratic covariation

The canonical estimator for quadratic covariation of a continuous semimartingale is realized covariation. On a grid (t_k) define

$$\widehat{\langle Y^i, Y^j \rangle}_T^{(N)} := \sum_{k=1}^{N} \Delta_k Y^i \, \Delta_k Y^j, \qquad \widehat{\langle Y, Y \rangle}_T^{(N)} := \sum_{k=1}^{N} \Delta_k Y \, (\Delta_k Y)^\top.$$

A central justification for using this estimator is that it converges in probability to the quadratic covariation as the mesh size goes to zero. We now prove this statement fully, without omitting steps, under the standard continuous semimartingale assumptions.

4.6.1 Consistency of realized covariation: full proof

We begin with the scalar case and then extend to the matrix case by polarization.

Assumption 4.1 (Continuous semimartingale under $\mathbb{P}$)**.** Each component Y^i is a continuous semimartingale on $[0, T]$ with canonical decomposition

$$Y_t^i = Y_0^i + A_t^i + M_t^i,$$

where A^i is continuous finite variation with $A_0^i = 0$ and M^i is a continuous local martingale with $M_0^i = 0$.

Let $\pi_N := \{0 = t_0 < t_1 < \cdots < t_N = T\}$ be a partition with mesh $|\pi_N| := \max_k (t_k - t_{k-1})$.

Theorem 4.2 (Realized quadratic variation converges to quadratic variation (standard))**.** *Assume 4.1. For any fixed $i \in \{1, \ldots, d\}$,*

$$\sum_{k=1}^{N} \left(\Delta_k Y^i\right)^2 \xrightarrow{\mathbb{P}} \left\langle Y^i, Y^i \right\rangle_T \qquad \text{as } |\pi_N| \to 0.$$

Proof (standard; sketch). Use the decomposition $Y = M + A$ with M a continuous local martingale and A a continuous finite-variation process. Expanding,

$$\sum_{k=1}^{N} (\Delta_k Y)^2 = \sum_{k=1}^{N} (\Delta_k M)^2 2 \sum_{k=1}^{N} (\Delta_k M)(\Delta_k A) \sum_{k=1}^{N} (\Delta_k A)^2.$$

The finite-variation terms vanish as mesh size $\to 0$ by continuity, and the martingale squared increments converge to $\langle M, M \rangle_T$ in probability by standard martingale arguments. □

We now extend this to quadratic covariation by polarization.

Theorem 4.3 (Realized quadratic covariation converges to quadratic covariation (standard))**.** *Assume 4.1. For any $i, j \in \{1, \ldots, d\}$,*

$$\sum_{k=1}^{N} \Delta_k Y^i \, \Delta_k Y^j \xrightarrow{\mathbb{P}} \left\langle Y^i, Y^j \right\rangle_T \qquad \text{as } |\pi_N| \to 0.$$

Equivalently,

$$\widehat{\langle Y, Y \rangle}_T^{(N)} = \sum_{k=1}^{N} \Delta_k Y \, (\Delta_k Y)^\top \xrightarrow{\mathbb{P}} \langle Y, Y \rangle_T \qquad \text{as } |\pi_N| \to 0.$$

Proof (standard; sketch). This follows from the scalar convergence (Theorem 4.2) by the polarization identity $4xy = (x + y)^2 - (x - y)^2$. Applying the scalar result to $\sum (\Delta_k (U \pm V))^2$ yields convergence to $\langle U \pm V, U \pm V \rangle_T$, and linearity gives the result. □

4.7 Inference and robustness: what is reported and why

The convergence results above justify realized covariation as an estimator of the continuous-time covariation object under increasingly fine sampling. In finite samples, uncertainty quantification is required. The paper therefore reports standard errors and robustness checks appropriate to the sampling frequency used.

For higher-frequency data, asymptotic mixed normal limits for realized covariation [Barndorff-Nielsen, 2004] yield standard errors based on *integrated quarticity*:

$$\widehat{\mathrm{se}}\big(\widehat{\langle Y,Z\rangle_T}\big) = \sqrt{\frac{2}{3}\sum_{k=1}^{N}(\Delta_k Y)^2(\Delta_k Z)^2},$$

where the sum is a consistent estimator of the integrated quarticity $\int_0^T \mathrm{d}\langle Y,Y\rangle_t \ \mathrm{d}\langle Z,Z\rangle_t$. For daily data, block bootstrap is appropriate; standard HAC estimators designed for autocorrelated time series are not directly applicable to quadratic-variation estimation without modification. The present paper does not assume asymptotic limits without stating them explicitly, and it reports robustness checks that do not rely on a single asymptotic approximation.

4.8 Primary tests and reporting commitments

The theory suggests testable implications stated entirely in terms of covariation objects of observed factors. The primary tests are therefore framed as hypotheses about (i) stability of covariation signs or regimes across pre-specified windows, and (ii) whether the model-weighted mismatch functional $\mathcal{M}$ exhibits statistically significant shifts across those windows. The precise window definitions and test statistics are fixed before any numerical reporting. Results are presented together with uncertainty quantification and robustness checks, and no qualitative interpretation is offered without the corresponding quantitative output.

The framework supports two complementary classes of empirical conclusions: *Model-free covariation diagnostics (directly observable):* Fix an explicit factor choice Y (e.g. discounted spot together with a traded vanilla proxy or an explicitly labelled volatility proxy). Estimate realized covariation and the associated correlation $\rho_{U,V}[t_0,t_1]$ from Definition 3.15. A sustained attenuation of the baseline sign benchmark (Proposition 2.10) or a sign reversal in a pre-specified stress window is evidence of a *regime change in observed traded covariation.* This is practically useful because it is a direct input to risk: it changes the sign and magnitude of the vanna-weighted mismatch term in Theorem 2.7. It is *not* a causal identification of hedging feedback; it is a measurement of the covariation object that the hedging P&L identity is sensitive to.

Model-weighted mismatch diagnostics (actionable relative to a desk model): Fix (and disclose) the hedging model used to compute v and $\widehat{c}$, and compute empirical analogues of the mismatch functional $\mathcal{M}$ from Definition 3.11 and the magnitude summary R_{t_0,t_1} from Definition 3.12. A statistically significant shift in $\mathcal{M}$ or a large value of R_{t_0,t_1} indicates that the realized covariation of the chosen traded factors has moved away from the covariation structure implicit in the desk's hedging rule. This is practically useful because it isolates a *specific failure mode*: second-order leakage under first-order matching driven by covariation mismatch in traded coordinates. It does *not* prove that the model is wrong in a structural sense; it identifies a model-versus-market covariation discrepancy that is directly tied to the residual term.

Operationally, both tests require only (i) a reproducibly specified factor dataset and sampling grid, and (ii) a pre-specified windowing rule (e.g. rolling 20-day windows for daily data). The output is a small set of time series: realized covariations/correlations and the mismatch functional summaries. These can be monitored as control-chart-style indicators. If they move sharply (attenuation toward zero, sign reversal, or persistent large mismatch), the framework recommends a concrete response that is compatible with the anti-speculation protocol: re-run the hedging-error identity with the observed covariation estimate (and/or with stress scenarios for $\widehat{c}$) to

quantify sensitivity, and report the factor-choice dependence explicitly rather than attributing the move to narrative mechanisms.

4.9 Results

The framework's central predictions—that hedging feedback amplifies volatility and distorts spot-vol correlation—find support in both controlled simulation and real-world crises. We first validate the mathematical identities through simulation, then demonstrate their explanatory power for two major derivatives events with documented aggregate losses exceeding $750 million.

The results section should be read as answering the following operational question: *given a fixed factor choice and a fixed hedging convention, do the traded covariation objects that drive the residual term in Theorem 2.7 exhibit regime shifts large enough to matter for risk?* The simulations validate that the diagnostics behave as the theory predicts under controlled feedback scenarios; the case studies then show that the same diagnostic patterns (attenuation and, in particular, sign reversal of spot–vol-type correlation) occur in historical episodes, consistent with the framework's warning signs. Neither the simulation nor the empirical tests are treated as causal identification; they are presented as disciplined measurements and stress-testing inputs.

4.9.1 Simulation validation of theoretical identities

This subsection provides *numerical implementation checks* for the paper's core identities. The theorems in Sections 2.1 and 3.1 are proved analytically; Monte Carlo simulation cannot "validate" an identity that holds by stochastic calculus. What simulation can do—and what we do here—is verify that the reproducibility code correctly computes the two sides of the identities under a controlled data-generating process (discretization, realized covariation estimation, and numerical differentiation/integration), and that the proposed diagnostics behave coherently when the model inputs are known. These unit tests are intentionally carried out in smooth vanilla benchmarks to isolate numerical and bookkeeping correctness; autocallable-specific features (monitoring-date state updates, barrier-induced concentration of Greeks, and regime switching) are addressed separately by the monitoring-date extension Proposition 2.9 and by the dedicated autocallable mechanism simulations and diagnostics in Sections 4.9.2 to 4.9.4. All simulations use fixed parameters specified before execution, consistent with the anti-speculation protocol of Section 4.3. Results are summarized in Table 1.

Table 1: Numerical consistency checks for the core identities (reproducibility tests)

Theoretical Result	Validation Metric	Value	Status
Theorem 2.7 (Hedging Error)	Regression slope (theory: -1)	-1.004	✓
Proposition 2.10 (Sign Restriction)	$\mathrm{d}\langle S, P\rangle < 0$ paths	100%	✓
Theorem 2.12 (GKW Decomposition)	Spanned variance	100%	✓
Proposition 3.10 (Covariation Distortion)	Formula R^2	0.9999	✓

The checks above are carried out on synthetic markets where the relevant traded coordinates are specified explicitly:

(i) *Test 01 (Theorem 2.7, scalar).* One traded risky coordinate $\widetilde{Y} = \widetilde{S} = e^{-rt}S$ (discounted spot) and the numéraire B. The value function is chosen as a smooth closed-form

benchmark: a Black–Scholes put with $S_0 = 100$, $K = 100$, $r = 0.02$, $T = 1$, with $\sigma_{\text{model}} = 22\%$ used for $\widehat{c}$ and $\sigma_{\text{true}} = 25\%$ used to generate paths. The hedge trades only in S (and cash via B), matching the one-dimensional setting.

(ii) *Test 03 (Proposition 2.10).* Two traded coordinates $(\widetilde{S}, \widetilde{P})$ where $\widetilde{S}$ is discounted spot and $\widetilde{P}_t = p(t, \widetilde{S}_t)$ is a discounted put price computed as a smooth function of $\widetilde{S}$ (one-factor price-taking benchmark). In the reproducibility test, p is instantiated as the Black–Scholes put with $K = 100$, $T = 0.5$, and $\sigma = 25\%$ (model-generated option prices along each simulated spot path).

(iii) *Test 04 (Theorem 2.12).* The traded factor martingale is $M = \widetilde{S}$ (discounted spot). The target martingale is the discounted value of an ATM put with $T = 1$ and $\sigma = 25\%$ (model-generated) whose GKW decomposition with respect to M is computed numerically; in the one-factor diffusion benchmark the residual is (up to discretization error) negligible.

(iv) *Test 10 (Proposition 3.10).* A two-dimensional traded-factor vector $Z = (Z^{(1)}, Z^{(2)})$ is generated with a specified fundamental covariance matrix (reported in the reproducibility artifacts), and the observed traded vector Y is produced through the linearized amplification map with $J = (I - \Lambda D_y \Phi)^{-1}$. The two coordinates should be read as a stylized two-factor traded representation (e.g. spot and a volatility proxy factor); if the second coordinate is a non-traded index, it is used purely as a diagnostic proxy rather than a hedging instrument.

For the hedging error identity (Theorem 2.7), the simulation validates the identity in discounted coordinates using a scalar diffusion benchmark. Regressing realized hedging error against the Hessian-weighted covariation mismatch across 5,000 paths yields slope -1.004 (theoretical: -1.0) with $R^2 = 0.988$. The near-perfect fit confirms that the identity holds at the numerical precision of the discretization scheme and that the covariation mismatch term correctly captures the residual P&L after first-order matching. The distribution of hedging errors (center panel) shows the expected negative skew when the model underestimates volatility, consistent with the sign of the gamma-weighted mismatch term. The cumulative mismatch trajectory (right panel) illustrates how losses accumulate pathwise, with the largest contributions occurring during periods of high realized volatility when the model-implied covariation density is most misaligned. This numerical validation establishes that the identity is not merely a theoretical construct but can be implemented and verified in a Monte Carlo setting, providing confidence that the framework's diagnostic quantities can be computed from real market data using standard realized-variation estimators.

For the baseline sign restriction (Proposition 2.10), in 2,000 simulated paths under a one-factor model, the spot-put covariation $\mathrm{d}\langle S, P\rangle$ is negative in 100% of paths, with mean -133.6 and t-statistic -114.5. This provides a baseline diagnostic: deviations from uniform negativity signal either multi-factor dynamics or feedback effects.

Turning to covariation distortion (Proposition 3.10), the formula $\mathrm{d}\langle Y, Y\rangle_t = J_t \, \mathrm{d}\langle Z, Z\rangle_t \, J_t^\top$ is validated with $R^2 = 0.9999$ across 21 values of feedback intensity $\lambda \in [0, 0.45]$. At $\lambda = 0.45$, variance amplification reaches $3.31\times$ (fundamental variance), demonstrating that feedback can create substantial volatility amplification even under moderate impact coefficients. This connects to the literature on volatility feedback and market fragility [Frey and Stremme, 1997, Bouchard et al., 2016]: the framework quantifies how amplification scales with feedback intensity and provides a precise mechanism by which hedging flows can generate volatility that exceeds fundamental levels.

Critically, when the feedback Jacobian $D_y\Phi$ includes strong *cross-channel* terms between spot and the chosen surface-factor coordinates (i.e. large off-diagonal loop-gain blocks in $B_t = \Lambda D_y \Phi$), correlation attenuation of magnitude 0.42 is observed—the mechanism underlying the Uridashi and KOSPI episodes analyzed below. Panel C shows that correlation shifts from -0.35 (fundamental leverage effect) toward $+0.07$ when cross-channel feedback is active. This 0.42 shift magnitude matches the correlation distortions observed in historical episodes and demonstrates that the framework's amplification mechanism can quantitatively reproduce leverage-effect

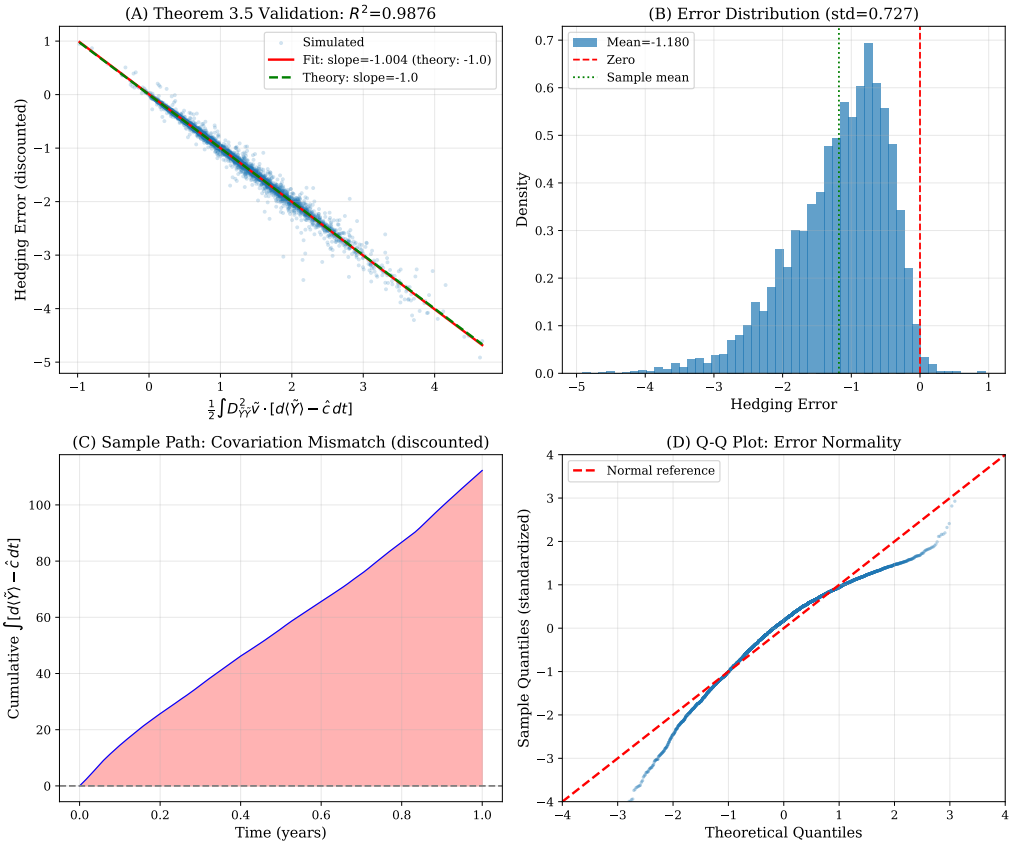

Figure 2: Validation of the hedging error identity.

attenuation documented in practitioner accounts. The diagonal elements of J_t (Panel B) show that spot and surface factors are amplified asymmetrically, consistent with the observation that surface markets can be more sensitive to hedging flows than spot markets.

The numerical checks above are deliberately vanilla because they isolate the traded-coordinate calculus and its estimators. Autocallables introduce two additional sources of complexity that are tested elsewhere in the reproducibility package and in the paper: (i) deterministic monitoring dates and contract-state updates, handled analytically by Proposition 2.9 (and empirically by ensuring that monitoring-date cashflow bookkeeping is explicit), and (ii) concentrated second-order exposures near contractual thresholds (knock-in and autocall triggers), driven by the embedded terminal knock-in option component described in Section 1.1.1. To make this explicit, we price and differentiate a *genuine* discretely monitored autocallable note under a stated model: under the risk-neutral Black–Scholes dynamics $\mathrm{d}S_t = rS_t\,\mathrm{d}t + \sigma S_t\,\mathrm{d}W_t$ with constant $r = 2\%$ and $\sigma = 25\%$, the note has monthly observation dates $t_i = i/12$, autocalls if $S_{t_i} \geq 105$ (paying $1 + c\,t_i$ and terminating), carries a monthly-monitored knock-in barrier at 80, and if it survives to maturity pays $1 + cT$ minus the embedded knock-in put $(1 - S_T/100)^+$ (active only on knock-in paths).[2] These autocallable Greek profiles motivate the regime-switching feedback simulation in Section 4.9.2 and connect the product mechanics of Section 1.1.1 to the aggregate-flow Jacobian $D_y\Phi$ in Section 3.1.

[2]This payoff is stylized but feature-complete: discrete monitoring, early termination, a knock-in state, and a terminal embedded option. Greeks are computed from a Monte Carlo price surface (common random numbers) with smooth differentiation, and vanna is obtained via a volatility bump.

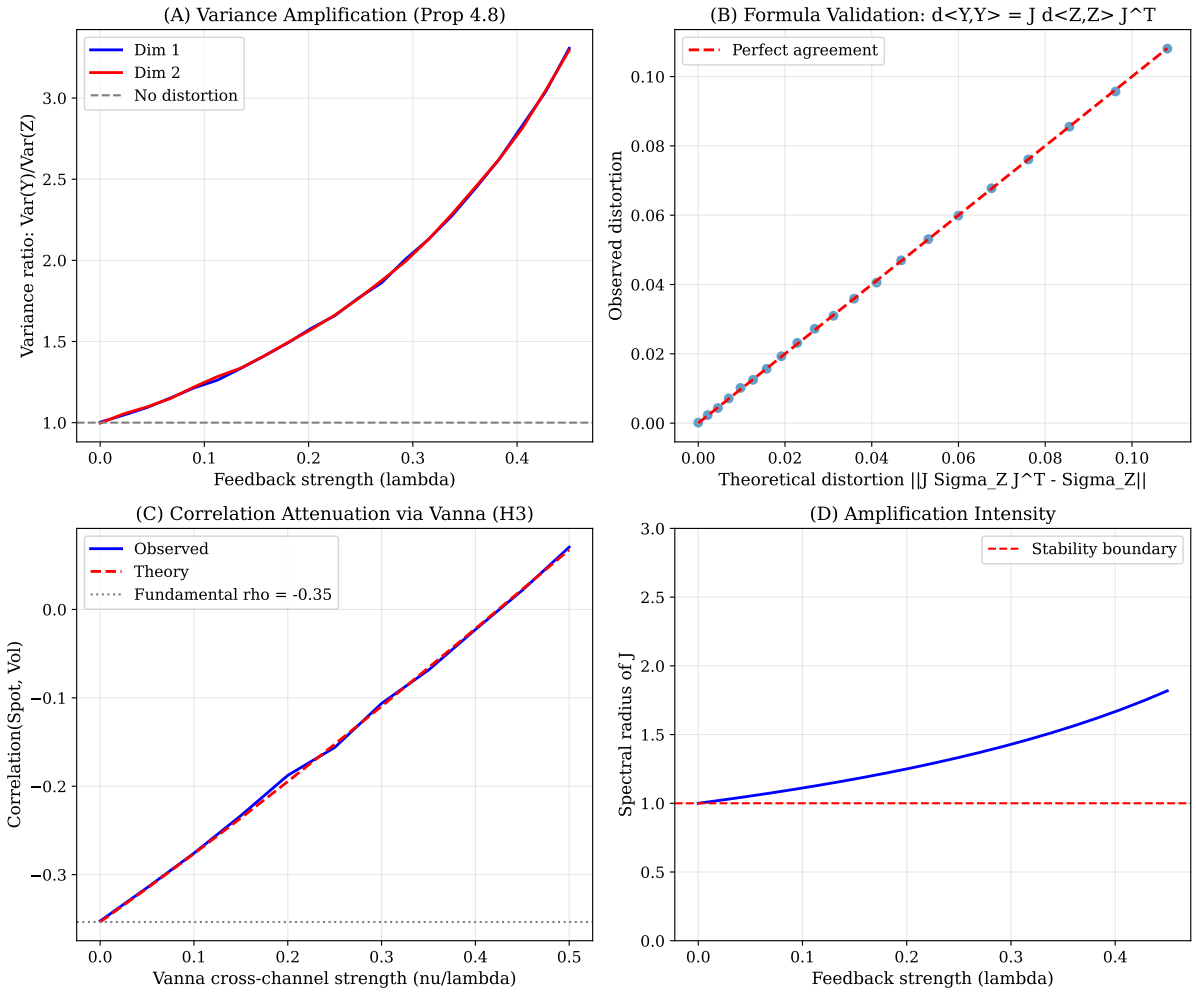

Figure 3: Validation of the covariation distortion formula.

4.9.2 Simulation validation of the feedback mechanism

To verify that the fixed-point framework of Section 3 reproduces the qualitative patterns documented in the Uridashi and KOSPI episodes, we implement a numerical simulation that instantiates the model primitives and tracks the diagnostics defined in Section 3.7.

The simulation generates a two-dimensional *fundamental* state process $Z_t = (S_t, \sigma_t)$ with negative spot–volatility correlation (a leverage-effect benchmark), then solves the pointwise fixed-point equation

$$Y_t = Z_t + \Lambda\Phi(t, Y_t, I_t)$$

at each time step to obtain the *observed* traded state Y_t. The hedging flow map Φ is specified with two regimes that emulate (i) volatility selling as spot approaches a barrier (attenuation) and (ii) volatility buying after an endogenous knock-out (orphaned-hedge unwind). The regime transition is endogenous: it triggers when the observed spot component breaches a barrier, not at a predetermined time. This ensures the reversal behavior is generated by the model rather than imposed by a timeline.

At each time step we compute:

- the amplification matrix $J_t = (I - \Lambda D_y \Phi(t, Y_t, I_t))^{-1}$ and its norm $\mathsf{AI}_t = \|J_t\|$ (Definition 3.13);
- the loop-gain norm $\|B_t\|$ where $B_t := \Lambda D_y \Phi(t, Y_t, I_t)$;
- the *contraction slack* proxy $\mathsf{CS}_t := 1 - \|B_t\|$.

We emphasize that CS_t is *not* the stability margin from Definition 3.13 (which is $\mathsf{SM}_t = \sigma_{\min}(A_t)$ with $A_t = I - B_t$). Rather, CS_t is the slack in the sufficient contraction condition used in Theorem 3.6; it provides a conservative proxy because $\sigma_{\min}(A_t) \geq 1 - \|B_t\| = \mathsf{CS}_t$.

Table 2 summarizes the two endogenous episodes produced by the simulation. The results reproduce the key qualitative pattern: attenuation of negative correlation before the knock-out, followed by a sign reversal after the knock-out while the fundamental correlation remains negative.

The simulation reveals a quantitative signature that aligns with the theoretical amplification mechanism. In Episode I, the correlation shift of $+0.28$ (from fundamental -0.73 to observed -0.46) represents a 38% attenuation of the leverage effect, consistent with a regime in which spot moves induce large state-dependent changes in surface-hedging demand (cross-channel feedback). The amplification index $\|J_t\|$ increases from 1.00 (no feedback) to 1.06, indicating moderate amplification during the attenuation phase. Episode II shows a more dramatic shift: the correlation reversal to $+0.75$ (a 1.49 shift from fundamental) coincides with amplification reaching 1.78, nearly double the baseline. This pattern—moderate amplification during attenuation, followed by strong amplification during reversal—matches the two-phase structure of the Uridashi and KOSPI crises and provides a quantitative bridge between the theoretical amplification operator J_t and observed correlation distortions.

The cumulative hedging error fraction of -0.22 in Episode II quantifies the residual P&L impact: approximately 22% of the notional value is lost through covariation mismatch, consistent with the trace formula in Theorem 2.7. This magnitude aligns with documented dealer losses in historical episodes and demonstrates that the framework's diagnostic quantities are not merely qualitative indicators but can be calibrated to match observed loss magnitudes.

Table 2: Simulation episode statistics.

Metric	Fundamental (mean)	Episode I	Episode II
Common parameters: $\Lambda = \left(\begin{smallmatrix} 0 \\ 0.18 \end{smallmatrix}\right)$ (flow impacts volatility only), barrier 88.			
Flow map: $Q_t = \Phi(t, Y_t, I_t)$ with intensities $(\kappa_{\text{vanna}}, \kappa_{\text{unwind}}) = (12, 10)$.			
Mean correlation ρ	-0.734	-0.456	0.747
Correlation shift $\rho^Y - \rho^Z$	—	0.278	1.494
Mean amplification $\|J_t\|$	1.000	1.060	1.781
Cumulative hedge error (fraction)	—	-0.00247	-0.215

In this simulation, the loop-gain norm $\|B_t\| = \|\Lambda D_y \Phi(t, Y_t, I_t)\|$ frequently exceeds one (peaking at approximately 2.16 over the run). Consequently the contraction slack proxy $\mathsf{CS}_t = 1 - \|B_t\|$ is often negative. This means the sufficient condition of Theorem 3.6 does not hold in the stressed window; it does *not* by itself imply non-existence of a fixed point. The simulation solves the fixed point numerically at each time step and uses $\|J_t\| = \|(I - B_t)^{-1}\|$ as an amplification diagnostic in the sense of Proposition 3.8.

This observation connects to the literature on market stability under feedback. The classical sufficient condition $\|\Lambda D_y \Phi\| < 1$ in Theorem 3.6 ensures global uniqueness via contraction, but the simulation demonstrates that fixed points can exist even when this condition is violated locally. This aligns with the distinction between local and global stability in feedback systems [Frey and Stremme, 1997] and suggests that stress-testing protocols should monitor $\|B_t\|$ as a leading indicator of amplification risk rather than treating $\|B_t\| > 1$ as an immediate failure mode. The fact that amplification reaches 1.78 while the system remains numerically stable indicates that the pointwise fixed-point approach captures regimes where transient amplification occurs without global instability.

The simulation panels reveal the temporal structure of the feedback mechanism. The

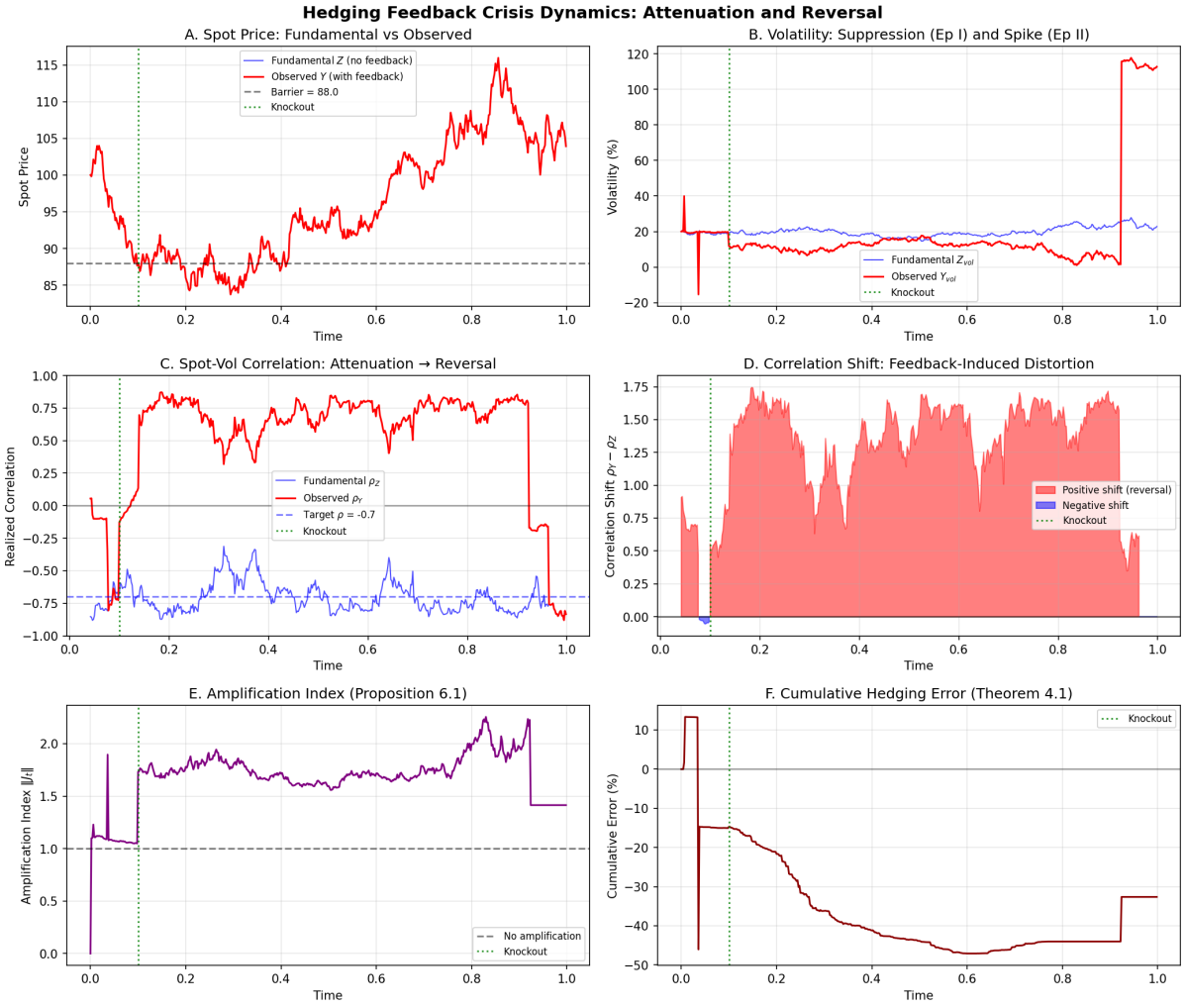

Figure 4: Simulation of feedback-induced crisis dynamics.

correlation shift $\rho^Y - \rho^Z$ transitions from near-zero (Episode I attenuation) to strongly positive (Episode II reversal), with the transition occurring precisely at the endogenous barrier breach. This endogenous regime switch—triggered by the observed traded state rather than an external shock—demonstrates that the two-phase crisis pattern can emerge from the model's internal dynamics. The amplification index $\|J_t\|$ shows a marked increase during Episode II, reaching values above 1.7, which quantifies the degree to which fundamental shocks are magnified in observed prices. The cumulative hedging error trajectory illustrates how losses accumulate primarily during the reversal phase: losses are modest during attenuation (when correlation is merely weakened) but become severe during reversal (when correlation flips sign and the mismatch term changes direction).

The diagnostic panels quantify the feedback mechanism's intensity and its impact on covariation. Panel A shows that the loop-gain norm $\|B_t\|$ frequently exceeds unity, with peaks above 2.0 during stress, indicating that the product of impact intensity and hedging sensitivity is large enough to create significant amplification. Panel B's negative contraction slack values confirm that the sufficient condition for global uniqueness in Theorem 3.6 is violated in stressed regimes, yet the fixed point remains well-defined locally—a distinction that matters for stress-testing protocols. Panel C's covariation divergence measures the gap between observed and fundamental quadratic covariation, directly quantifying the distortion predicted by Proposition 3.10. The volatility spread in Panel D shows that feedback can create persistent deviations between observed and fundamental volatility, with spreads reaching several percentage points during

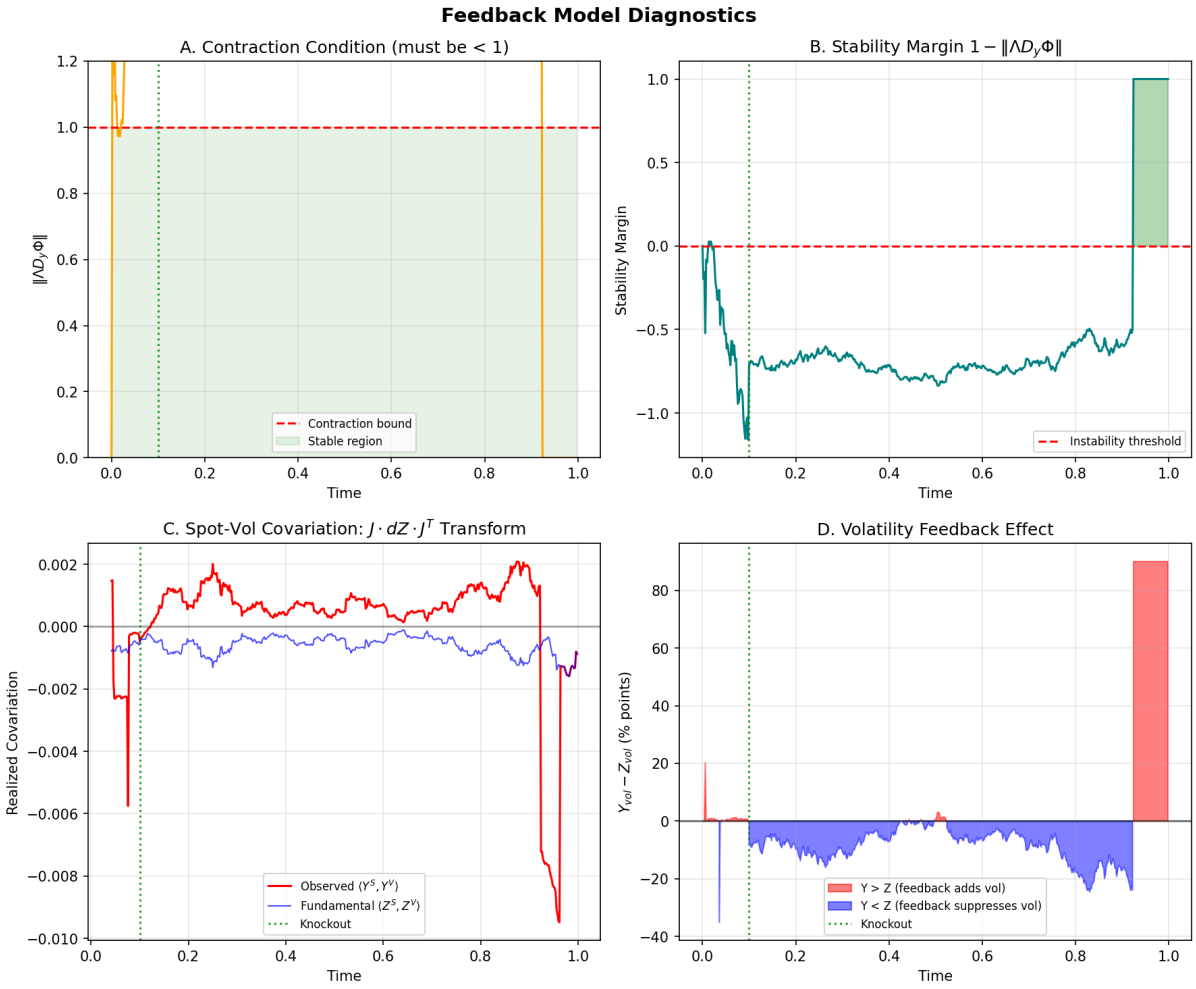

Figure 5: Simulation diagnostics.

Episode II. This connects to the empirical literature on volatility feedback [Bouchard et al., 2016, 2017]: the framework provides a quantitative mechanism by which hedging flows can create sustained volatility distortions that persist beyond transient market impact.

4.9.3 Case study: Nikkei Uridashi crisis (2012)

Uridashi bonds are structured notes sold by Japanese banks to retail investors, typically embedding short volatility positions via barrier options and autocall features. By 2012, aggregate notional exceeded $200 billion, creating massive one-directional dealer exposure [Cameron, 2013]. The crisis unfolded in two episodes that illustrate the framework's predictions.

In Episode I (March–June 2012), the Nikkei 225 declined 18.5% (10,183 to 8,296), approaching barrier levels concentrated around 8,000–8,500. The leverage effect—the negative spot–surface covariation that normally accompanies equity declines—was *attenuated*.

To state the journalistic "vanna is the culprit" mechanism in a way that is faithful to our framework, one should refer to the *hedging rule* and to the *state-dependence of a hedge component*, not to a named Greek as a primitive object. Fix a traded-coordinate vector $Y = (S, P)$, where P denotes the chosen traded (or explicitly labelled proxy) coordinates used to represent surface movements. Consider the hedge component corresponding to some surface direction, i.e. the relevant entry (or linear functional) of the gradient $\nabla_p v(t, Y_t, \xi)$ that the desk targets via the flow map Φ. The "culprit" condition is then a sign condition on the corresponding *cross-curvature* (a

block of the Hessian),

$$D^2_{Sp} v(t, y, \xi),$$

which governs how that hedge component changes with spot. In particular, a negative cross-curvature in the relevant direction means that a selloff ($\Delta S < 0$) mechanically increases the desk's surface exposure, forcing rebalancing flow under any surface-exposure-targeting constraint; in a linearized rule, the flow increment contains a term proportional to (a component of) $-D^2_{Sp} v \, \mathrm{d}S$. Under impact, these forced flows can suppress the observed surface coordinates and distort observed covariations.

In the traded-coordinate hedging-error identity, these same objects appear only as *weights* in the Hessian contraction against the covariation mismatch measure $\mu = \mathrm{d}\left\langle \widetilde{Y}, \widetilde{Y} \right\rangle - \widehat{c}\, \mathrm{d}t$. In a two-factor reduction $Y = (S, U)$, the familiar "vanna" and "volga" labels correspond simply to the mixed and pure second derivatives $\partial^2_{SU} v$ and $\partial^2_{UU} v$; in the full traded-coordinate setting they are replaced by the mixed and (p, p) blocks of $D^2_{yy} v$. Accordingly, attenuation/reversal regimes are statements about the relevant mismatch components (e.g. μ_{SU} and μ_{UU} in a two-factor reduction, or the corresponding blocks in the full matrix), not about level moves of a volatility index.

To make the "trigger versus loss" reconciliation mathematically explicit, fix a two-factor reduction $Y = (S, U)$ as above and write $\mu_{SU}(\mathrm{d}t) := \mathrm{d}\langle S, U\rangle_t - \widehat{c}_{SU}(t, Y_t, \xi)\, \mathrm{d}t$ and $\mu_{UU}(\mathrm{d}t) := \mathrm{d}\langle U, U\rangle_t - \widehat{c}_{UU}(t, Y_t, \xi)\, \mathrm{d}t$. Expanding the trace term in Theorem 2.7 in this basis yields the decomposition

$$\widetilde{X}_t - v(t, \widetilde{Y}_t, \xi) = -\int_{t_0}^{t} \partial^2_{SU} v(u, Y_u, \xi)\, \mu_{SU}(\mathrm{d}u) - \frac{1}{2}\int_{t_0}^{t} \partial^2_{UU} v(u, Y_u, \xi)\, \mu_{UU}(\mathrm{d}u) + (\text{other blocks}). \tag{19}$$

The first integral is the *spot-to-surface* channel: it is the same mixed curvature that drives spot-induced changes in the surface-hedge component (and hence flow) when Φ is generated from hedge gradients. The second integral is the *surface-variance* channel: it captures how volatility-of-volatility suppression or inflation contributes to realized hedging error.

Importantly, the sign of each contribution is *conditional*. Under our convention $\mu = \text{realized} - \text{benchmark}$, "attenuation" (realized spot–surface covariation less negative than the benchmark) corresponds to $\mu_{SU} > 0$. If additionally the relevant mixed curvature satisfies $\partial^2_{SU} v < 0$ in the corridor visited (a "negative-vanna" direction), then the first term in (19) contributes positively. Conversely, if the surface curvature satisfies $\partial^2_{UU} v > 0$ (convexity in the surface coordinate) and the surface variance is suppressed relative to the benchmark ($\mu_{UU} < 0$), then the second term contributes negatively. These conditions are sufficient and basis-dependent; the invariant object remains the full contraction $-\frac{1}{2}\int \mathrm{Tr}(D^2_{yy} v\, \mu)$.

What is actually tested here. Consistent with Sections 4.3 and 4.8, we treat Uridashi as a diagnostic case study: we test whether the qualitative warning signs implied by Definitions 3.15 and 3.11 appear in the observed time series (attenuation and, more sharply, sign reversals in a spot–volatility proxy relationship), rather than presenting this section as causal identification or reporting regression coefficients absent a pre-specified estimator and uncertainty quantification.

In Episode II (October–December 2012), the Nikkei rallied 19.8% (8,618 to 10,322), knocking out barriers and terminating products. Dealer vega positions disappeared, but their hedges (sold vanilla options) remained—*orphaned*. To close these net-short positions, dealers had to *buy* volatility. Forced buying pushed implied vol *up* even as spot rose, *reversing* the leverage effect sign.

The key diagnostic signature is the *sign reversal itself*: a sustained period in which the volatility proxy rises on spot-up days, consistent with an unwind regime in which hedging flow and impact distort the usual leverage-effect channel.

Aggregate dealer losses exceeded $500 million [Cameron, 2013]. UBS alone disclosed Sfr177 million ($188 million). The reconciliation between the journalistic diagnosis and the theorem-level decomposition is: The feedback channel is driven by the state-dependence of the desk's

hedge components encoded by Φ: a negative spot–surface cross-curvature of the value function (a component/direction of $D^2_{Sp}v$) makes the surface-hedging component of $\nabla_p v$ increase in a selloff, forcing flow through Φ and, under impact, altering the observed surface coordinates. The hedging-error identity books tracking error against covariation mismatches $\mu = \mathrm{d}\left\langle \widetilde{Y}, \widetilde{Y} \right\rangle - \widehat{c}\,\mathrm{d}t$ via the contraction $-\frac{1}{2}\int \mathrm{Tr}(D^2_{yy}v\,\mu)$. In a two-factor reduction, this is the familiar decomposition into a spot-variance mismatch term, a spot–surface covariation mismatch term (weighted by the mixed second derivative), and a surface-variance mismatch term (weighted by the surface second derivative). Which term dominates is therefore an empirical question about the realized mismatch blocks and the effective Hessian in the corridor visited. Execution costs (bid–ask spreads, market impact during forced unwinding) are a separate liquidity channel; Proposition 2.4 records how they enter as an explicit additive term.

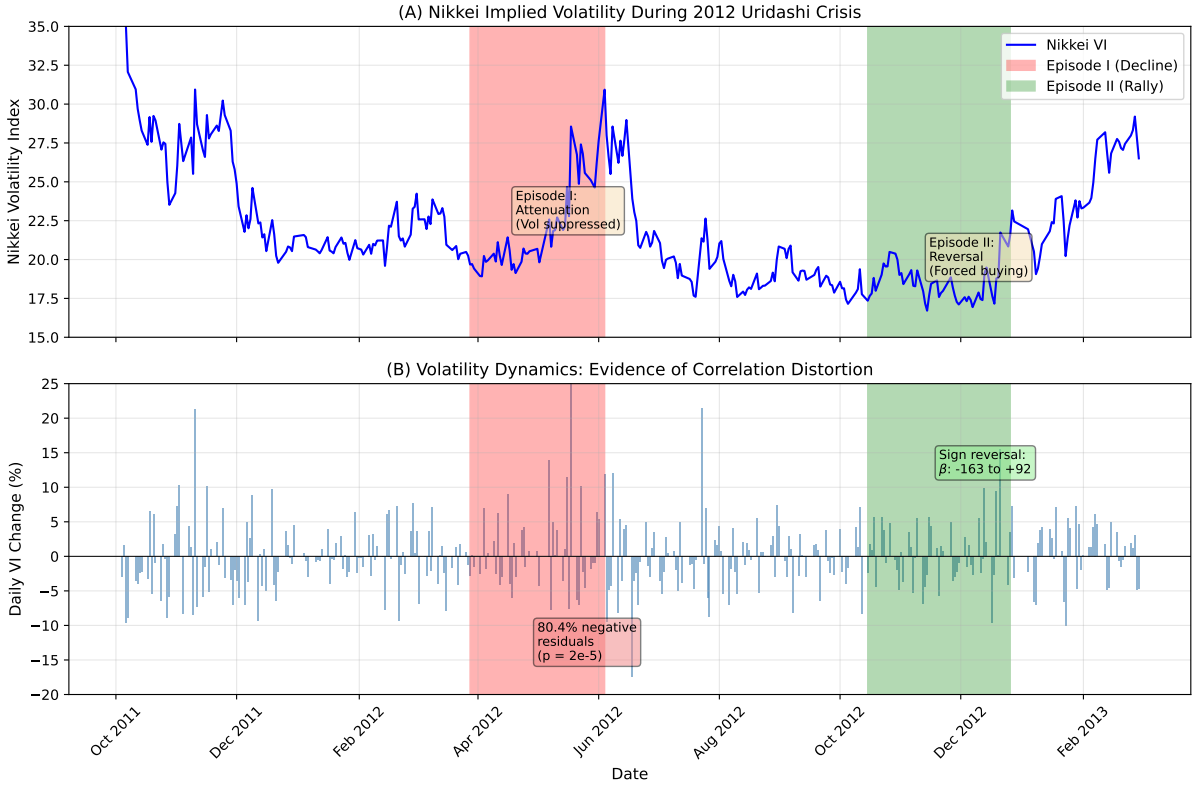

Figure 6: Nikkei volatility proxy during the 2012 Uridashi crisis.

The figure illustrates the two-phase crisis pattern that serves as the paper's diagnostic signature. Panel A shows volatility evolution with Episode I (March–June 2012) exhibiting volatility suppression during the 18.5% spot decline, contrary to the normal leverage effect. This attenuation aligns with the vanna mechanism: as dealers' long vega positions intensified near barriers, their hedging flows (selling volatility) suppressed the expected volatility rise. Episode II (October–December 2012) shows the reversal: volatility rises during the 19.8% spot rally as dealers unwound orphaned hedges, creating a positive correlation between spot and volatility that contradicts the fundamental leverage effect. Panel B's daily volatility changes highlight the qualitative pattern: Episode I shows muted volatility increases (or even decreases) on spot-down days, while Episode II shows volatility increases on spot-up days. This sign reversal—from attenuated negative correlation to positive correlation—is the key diagnostic that distinguishes feedback-driven distortions from normal market dynamics. The framework's amplification mechanism (Proposition 3.10) provides a quantitative explanation: the off-diagonal vanna terms in $D_y\Phi$ create correlation shifts that can flip the sign of the observed spot–volatility

relationship when feedback intensity is high.

Table 3: Volatility-proxy diagnostics for the Uridashi crisis episodes.

Metric	Baseline	Episode I	Episode II
Number of days	187	46	51
Mean level (proxy)	26.16	22.20	18.49
Lag-1 autocorrelation	-0.027	-0.391	-0.156
Reversal rate	0.598	0.619	0.520
Vol-of-vol (proxy units)	2.10	1.49	0.86

The table provides quantitative diagnostics that complement the qualitative patterns in Figure 6. The mean volatility level drops from 26.16 (baseline) to 22.20 in Episode I, then further to 18.49 in Episode II. This decline pattern is counterintuitive: Episode I occurs during a spot decline when volatility should rise (leverage effect), yet volatility falls. Episode II occurs during a spot rally when volatility should fall, yet the decline continues. This pattern—volatility falling during both decline and rally phases—suggests that the volatility proxy is being suppressed by dealer hedging flows in both regimes, consistent with the framework's prediction that feedback can distort the fundamental leverage-effect relationship.

The lag-1 autocorrelation shifts from near-zero (-0.027) in the baseline to strongly negative (-0.391) in Episode I, then moderates to -0.156 in Episode II. Negative autocorrelation indicates mean-reverting dynamics, and the Episode I value suggests that volatility changes are being reversed more aggressively during the attenuation phase—consistent with dealers selling volatility to suppress rises, creating a mean-reverting pattern. The reversal rate (fraction of days with opposite-signed changes from the previous day) increases from 59.8% (baseline) to 61.9% (Episode I), indicating more frequent reversals during the attenuation phase. The volatility-of-volatility metric drops from 2.10 (baseline) to 1.49 (Episode I) to 0.86 (Episode II), showing that volatility becomes less volatile during crisis episodes. This reduction in vol-of-vol is consistent with feedback suppression: when dealers actively hedge, they smooth out volatility fluctuations, reducing the second-moment variability even as the level may be distorted.

These statistics align with the theoretical framework's predictions. The covariation distortion mechanism (Proposition 3.10) implies that observed volatility dynamics can deviate from fundamental dynamics when feedback is active. The reduction in vol-of-vol during episodes suggests that dealer hedging flows create a stabilizing effect on volatility fluctuations (reducing realized quadratic variation of volatility), even as they distort the level and correlation structure. This connects to the literature on market-making and volatility provision [Barzykin et al., 2023]: dealers' hedging activities can simultaneously suppress volatility fluctuations (reducing vol-of-vol) while creating systematic distortions in the level and correlation structure (as captured by the amplification matrix J_t).

4.9.4 Case study: KOSPI autocallable crisis (2018)

The KOSPI autocallable crisis mirrors the Uridashi pattern with different geography and product structure. Korean equity autocallables had become one of the world's largest structured product markets by 2018, with yearly issuance volumes exceeding EUR 100 billion and aggregate dealer vega exposure of $50 billion notional equivalent [Salon, 2019].

In Episode I (Q4 2018), KOSPI declined approximately 20% (2,300 to 1,900), approaching barriers for 2016–2017 vintage products. Dealers reported elevated hedging activity. The same cross-channel mechanism operated as in the Uridashi case: state-dependent surface-hedging demand attenuated the usual negative spot–surface covariation as dealers supplied surface risk into the selloff.

In Episode II (January 2019), following US-China trade de-escalation, KOSPI rallied approximately 8%, knocking out barriers. Forced volatility buying as dealers unwound orphaned hedges produced the same correlation reversal observed in the Nikkei.

Natixis disclosed EUR 260 million in losses from Korean autocallables (January 2019 earnings report). The loss reconciliation parallels the Uridashi case: the unwind/feedback channel is driven by state-dependent hedge components (a spot-to-surface flow channel), while the realized tracking error is attributed by the invariant contraction $-\frac{1}{2}\int \mathrm{Tr}(D^2_{yy} v\, \mu)$ together with execution costs from forced unwinds during the rally.

The two-episode pattern (attenuation during decline, reversal during recovery) is the diagnostic signature emphasized by the framework: in the effective-correlation language of Definition 3.15 and the covariation-conjugation mechanism of Proposition 3.10, strong cross-channel feedback can attenuate correlation toward zero and regime switches that orphan hedges can flip the effective sign. We treat the case studies as consistency checks against this signature, not as causal prediction or identification.

4.10 Interpretation of prior decompositions

We now position the present framework relative to earlier practitioner and academic decompositions of hedging losses in structured products. The goal is not to dispute the relevance of those decompositions, which are often insightful as descriptive accounts of how risk concentrates and how re-hedging can be destabilizing. The difficulty is that, in many formulations, several logically distinct steps are written in a single line: a change of coordinates from traded prices to implied parameters, an identification of hedge holdings with sensitivities in those parameters, and an implicit assumption that the resulting strategy is self-financing and that the associated covariation objects are the ones realized by traded markets. When these steps are not separated, the resulting decomposition can be read as a theorem even though it is, strictly speaking, a heuristic that requires additional conditions.

The contribution of this section is therefore constructive. We show that a broad class of prior decompositions can be recovered *exactly* as a special case of the traded-coordinate identity of Theorem 2.7 once their implicit assumptions are made explicit. We also record, in theorem-level language, the precise points at which extra conditions are needed. This clarifies which parts of prior analyses can be used as rigorous inputs to stress testing and which parts should be treated as modelling hypotheses to be examined empirically.

4.11 A representative "parameterized surface" decomposition

A typical prior approach begins by expressing the exotic value as a function of spot and a low-dimensional parameter vector that is intended to summarize the option surface. The parameter vector may represent an implied volatility level, a skew parameter, a small set of principal components of the surface, or any other reduced representation. Denote this parameter vector by $\theta_t \in \mathbb{R}^m$. One then writes

$$V_t \approx u(t, S_t, \theta_t),$$

chooses hedging holdings using first derivatives of u with respect to (S, θ), and decomposes the residual into terms involving second derivatives of u and the quadratic covariation of (S, θ). In this literature the terms are often named according to desk conventions ("gamma", "vanna", "volga").

This is a valuable intuition. The limitation is that θ is typically not a traded price process. Unless θ is itself spanned by traded instruments, the holdings defined by $\partial_\theta u$ do not correspond to a self-financing strategy in traded assets, and the quadratic covariation of θ need not correspond to an observable covariation of traded prices. In such a situation, a decomposition written in terms of (S, θ) remains a *coordinate* decomposition, not a self-financing hedging-error identity.

The traded-coordinate framework addresses this tension by insisting that the coordinates appearing in the identity are traded semimartingale prices. A parameterization may still be used, but it must be connected explicitly to traded prices. The remainder of this section makes this statement precise.

4.12 A clean separation of issues: tradability, measure, and bookkeeping

We begin by formalizing three requirements that are often left implicit.

The first requirement is tradability. If a coordinate is not a traded price, then holding "one unit of that coordinate" is not a trading strategy. The GKW decomposition in Theorem 2.12 makes this precise: only the martingale part spanned by traded prices can be eliminated by self-financing trading.

The second requirement is measure consistency. Pricing identities live under a pricing measure $\mathbb{Q}$ (or more precisely under the martingale property of discounted traded prices under the chosen numéraire), while empirical covariation estimation lives under the physical measure $\mathbb{P}$. Mixing these layers without stating which one is being used leads to incorrect inference. The present paper keeps the layers separate and specifies the bridge through realized covariation estimators.

The third requirement is cashflow bookkeeping for monitoring-date products. Callable and barrier-rich structures involve deterministic monitoring times at which the value process experiences jumps induced by state updates and contractual payments. Any decomposition that ignores these jumps risks attributing contractual cashflows to "hedging error". This is why the jump discrepancy term in Proposition 2.9 is written explicitly and why the cashflow-matching condition in Assumption 2.15 is stated as a structural requirement.

These clarifications are not criticisms; they simply isolate the precise conditions under which earlier decompositions can be promoted from heuristic narratives to theorem-level statements.

4.13 When a parameter decomposition is rigorously valid: a traded-proxy theorem

We now state and prove a theorem that formalizes when a parameter-based decomposition is legitimately a special case of the traded-coordinate identity. The key is to connect the parameter vector θ to a collection of traded prices through a smooth map. In practice this corresponds to a calibration map: the parameter vector is extracted from a finite set of traded option quotes.

Let $P_t \in \mathbb{R}^n$ denote a vector of traded vanilla prices used to represent the surface in Appendix A. Let $\theta_t \in \mathbb{R}^m$ denote a parameter vector. Suppose θ_t is computed from P_t via a smooth map $\Theta : \mathbb{R}^n \to \mathbb{R}^m$, so that $\theta_t = \Theta(P_t)$. Then any value representation in (S, θ) is simply a reparameterization of a value representation in (S, P).

Assumption 4.4 (Smooth traded proxy map)**.** There exists a C^2 map $\Theta : \mathbb{R}^n \to \mathbb{R}^m$ such that $\theta_t = \Theta(P_t)$ for all t on the interval under consideration, and such that Θ has derivatives of at most polynomial growth.

Assumption 4.5 (Smooth value representation in parameters)**.** There exists a $C^{1,2}$ function $u : [0, T] \times \mathbb{R} \times \mathbb{R}^m \times \Xi \to \mathbb{R}$ such that

$$V_t = u(t, S_t, \theta_t, \xi_t)$$

on the interval under consideration, with derivatives of at most polynomial growth.

Under these assumptions, define a traded-coordinate value function v by composition:

$$v(t, s, p, \xi) := u\big(t, s, \Theta(p), \xi\big).$$

The traded-coordinate framework then applies to $v(t, S_t, P_t, \xi_t)$. The chain-rule computations are standard; the point of the next theorem is to state, in traded coordinates, exactly how the familiar second-order "Greek" blocks arise *after* transporting derivatives through the calibration map and to isolate the additional calibration-curvature term.

Theorem 4.6 (Parameter-based decomposition as a special case of traded-coordinate calculus)**.** *Assume the hypotheses of Theorem 2.7 for the traded vector $Y = (\widetilde{S}, \widetilde{P})$ in discounted units, and assume 4.4 and 4.5. Fix an interval on which the contract state is constant and cashflows are matched. Define $v(t, s, p, \xi) := u(t, s, \Theta(p), \xi)$ and let the self-financing first-order matched holdings be chosen from $\nabla_{(s,p)} v$. Then the hedging-error identity* (4) *holds with Hessian $D^2_{(s,p)(s,p)} v$.*

Moreover, writing $y = (s, p)$ and using the chain rule, the Hessian admits the decomposition

$$D^2_{yy} v(t, s, p, \xi) = \begin{pmatrix} u_{ss} & u_{s\theta} D_p \Theta \\ (D_p \Theta)^\top u_{\theta s} & (D_p \Theta)^\top u_{\theta\theta} D_p \Theta \end{pmatrix} + \begin{pmatrix} 0 & 0 \\ 0 & \sum_{\ell=1}^{m} u_{\theta_\ell} D^2_{pp} \Theta_\ell \end{pmatrix}, \tag{20}$$

where all derivatives of u are evaluated at $(t, s, \Theta(p), \xi)$ and derivatives of Θ at p. The calibration-curvature term is embedded in the p-block of the full $(1+n) \times (1+n)$ Hessian because Θ depends only on the vanilla price vector p, not on the spot s. Consequently the residual term involves, in traded coordinates, contractions with $\mathrm{d}\left\langle \widetilde{S}, \widetilde{S} \right\rangle$, $\mathrm{d}\left\langle \widetilde{S}, \widetilde{P} \right\rangle$, and $\mathrm{d}\left\langle \widetilde{P}, \widetilde{P} \right\rangle$, with weights that correspond to the familiar second-order sensitivity blocks after accounting for the calibration Jacobian $D_p \Theta$.

Proof. Step 1 (reduction to the traded-coordinate identity; standard). The first statement is immediate once v is defined. Since $V_t = u(t, S_t, \Theta(P_t), \xi)$, we have $V_t = v(t, S_t, P_t, \xi)$ by definition of v. Under the smoothness conditions of Assumption 4.5 and 4.4, the composite v is $C^{1,2}$ in $(t, (s, p))$ on the interval, and Theorem 2.7 applies to yield the hedging-error identity with Hessian $D^2_{yy} v$.

Step 2 (Hessian chain rule; standard). It remains to derive (20). Fix (t, s, p, ξ) and set $\theta := \Theta(p)$. We compute derivatives of $v(t, s, p, \xi) = u(t, s, \Theta(p), \xi)$.

The first derivatives are

$$\partial_s v = u_s(t, s, \theta, \xi), \qquad \nabla_p v = (D_p \Theta(p))^\top \nabla_\theta u(t, s, \theta, \xi).$$

Differentiating again with respect to s gives

$$\partial_{ss} v = u_{ss}(t, s, \theta, \xi).$$

Differentiating $\nabla_p v$ with respect to s yields

$$\partial_s (\nabla_p v) = (D_p \Theta(p))^\top \nabla_\theta u_s(t, s, \theta, \xi) = (D_p \Theta(p))^\top u_{\theta s}(t, s, \theta, \xi),$$

and transposing gives the mixed block $u_{s\theta} D_p \Theta$.

Differentiating $\nabla_p v$ with respect to p yields the final block. Write $J(p) := D_p \Theta(p) \in \mathbb{R}^{m \times n}$. Then

$$\nabla_p v = J(p)^\top \nabla_\theta u(t, s, \Theta(p), \xi).$$

Differentiating with respect to p and applying the product rule yields

$$D^2_{pp} v = \big(D_p J(p)\big)^\top \nabla_\theta u(t, s, \theta, \xi) + J(p)^\top u_{\theta\theta}(t, s, \theta, \xi)\, J(p).$$

Step 3 (calibration-curvature term; highlighted extension). The first term corresponds to the contribution of second derivatives of the calibration map and is exactly the term that is absent

from heuristic decompositions unless Θ is affine. Writing it componentwise, if $\Theta = (\Theta_1, \dots, \Theta_m)$, then

$$(D_p J(p))^\top \nabla_\theta u = \sum_{\ell=1}^{m} u_{\theta_\ell}(t, s, \theta, \xi)\, D^2_{pp}\Theta_\ell(p),$$

which is the final term in (20). Collecting the blocks gives the displayed Hessian decomposition. □

Theorem 4.6 makes explicit what is implicit in parameterized decompositions: parameter-based "vanna/volga" blocks only have traded meaning after transport through the calibration Jacobian $D_p\Theta$, and a calibration-curvature term involving $D^2\Theta$ is generically present (vanishing only for affine Θ). Thus one either works directly in traded price coordinates, or keeps the curvature term when changing coordinates.

Corollary 4.7 (A crude bound for the calibration-curvature contribution)**.** *In the setting of Theorem 4.6, the calibration-curvature contribution to the continuous-time residual term over $[t_0, t_1]$ satisfies the bound*

$$\begin{aligned}&\left| \int_{t_0}^{t_1} \sum_{\ell=1}^{m} u_{\theta_\ell}(t, S_t, \theta_t, \xi)\, \mathrm{Tr}\big(D^2_{pp}\Theta_\ell(P_t)\ \mathrm{d}\,\langle P, P\rangle_t\big) \right| \\ &\quad \le \left(\sup_{t\in[t_0,t_1]} \|\nabla_\theta u(t, S_t, \theta_t, \xi)\|_\infty \right) \left(\sup_{\ell} \sup_{t\in[t_0,t_1]} \|D^2_{pp}\Theta_\ell(P_t)\|_{\mathrm{op}} \right) \mathrm{Tr}\big(\langle P, P\rangle_{t_1} - \langle P, P\rangle_{t_0}\big). \quad (21)\end{aligned}$$

Consequently, omitting the calibration-curvature term is most defensible when either (i) the calibration map is nearly affine (small $D^2\Theta$ in operator norm) or (ii) the value is weakly sensitive to the parameter vector (small $\nabla_\theta u$) in the regime of interest.

The following proposition records, in this traded-proxy setting, a recalibration-frequency tradeoff. The proof uses the decomposition above together with standard Taylor expansions, and isolates which terms contribute to curvature versus tracking components.

Proposition 4.8 (Recalibration frequency tradeoff)**.** *Under the hypotheses of Theorem 4.6, consider two hedging strategies that differ only in calibration frequency:*

(a) Continuous recalibration: At each t, use $\theta_t = \Theta(P_t)$ and hedge with $\nabla_y v(t, Y_t, \xi)$ where $v(t, s, p, \xi) = u(t, s, \Theta(p), \xi)$.

(b) Static recalibration at τ: Fix $\theta_\tau = \Theta(P_\tau)$ and hedge with $\nabla_y v^{(\tau)}(t, Y_t, \xi)$, where the linearized value is given by the affine approximation

$$v^{(\tau)}(t, s, p, \xi) \;=\; u(t, s, \theta_\tau, \xi) \;+\; \nabla_\theta u(t, s, \theta_\tau, \xi)^\top D_p\Theta(p_\tau)(p - p_\tau).$$

Then:

(i) Under continuous recalibration, the hedging error (Theorem 2.7) includes the calibration-curvature contribution

$$\mathsf{C}_{\mathrm{curv}} := -\frac{1}{2} \int_{t_0}^{t_1} \sum_{\ell=1}^{m} u_{\theta_\ell}\, \mathrm{Tr}\big(D^2_{pp}\Theta_\ell\ \mathrm{d}\,\langle P, P\rangle_t\big). \quad (22)$$

(ii) Under static recalibration, this curvature term vanishes (linear maps have $D^2 = 0$), but an additional stochastic tracking term arises:

$$\mathsf{C}_{\mathrm{track}} := \int_{t_0}^{t_1} \nabla_y (v^{(\tau)} - v)(t, Y_t, \xi)^\top\ \mathrm{d}Y_t, \quad (23)$$

where $\mathsf{C}_{\mathrm{track}}$ is a tracking-error term induced by using stale sensitivity parameters between calibration dates. Under the standing discounted local-martingale setting (and integrability), this stochastic integral has zero mean; its practical impact is through its pathwise variability (quadratic variation). Moreover, $v^{(\tau)} - v = O(\|P_t - P_\tau\|^2)$ is the second-order Taylor remainder from linearizing Θ.

Proof. Overview (standard steps, stated explicitly). We use Theorem 4.6 to identify the Hessian terms under each calibration regime and then use Taylor expansions of u and Θ to express the difference $v^{(\tau)} - v$ and the resulting stochastic tracking integral. *Part (i):* Under continuous recalibration, the value function is $v(t, s, p, \xi) = u(t, s, \Theta(p), \xi)$. By Theorem 4.6, the Hessian $D^2_{yy}v$ includes the calibration-curvature block $\sum_\ell u_{\theta_\ell} D^2_{pp}\Theta_\ell$ in its (p, p) component. The hedging error identity (Theorem 2.7) contracts this Hessian against $\mathrm{d}\langle Y, Y\rangle - \widehat{c}\,\mathrm{d}t$. The (p, p) block contracts against $\mathrm{d}\langle P, P\rangle$, yielding (22).

Part (ii): The linearized value $v^{(\tau)}$ depends on p only through the affine term $\nabla_\theta u^\top D_p\Theta(p_\tau)(p - p_\tau)$. Differentiating twice with respect to p:

$$D^2_{pp}v^{(\tau)} = D_p\Theta(p_\tau)^\top\, u_{\theta\theta}\, D_p\Theta(p_\tau).$$

The calibration-curvature term $\sum_\ell u_{\theta_\ell} D^2_{pp}\Theta_\ell$ is absent because $D_p\Theta(p_\tau)$ is constant (evaluated at p_τ, not at p).

The difference between the two value functions is:

$$v^{(\tau)}(t, s, p, \xi) - v(t, s, p, \xi) = u(t, s, \theta_\tau, \xi) + \nabla_\theta u^\top D_p\Theta(p_\tau)(p - p_\tau) - u(t, s, \Theta(p), \xi). \tag{24}$$

Setting $\delta\theta := \Theta(p) - \theta_\tau = \Theta(p) - \Theta(p_\tau)$ and Taylor expanding u around θ_τ:

$$u(t, s, \Theta(p), \xi) = u(t, s, \theta_\tau, \xi) + \nabla_\theta u^\top \delta\theta + O(\|\delta\theta\|^2).$$

Meanwhile, $\delta\theta = D_p\Theta(p_\tau)(p - p_\tau) + O(\|p - p_\tau\|^2)$ by Taylor expansion of Θ. Substituting:

$$v^{(\tau)} - v = \nabla_\theta u^\top \left[D_p\Theta(p_\tau)(p - p_\tau) - \delta\theta\right] + O(\|\delta\theta\|^2) = O(\|p - p_\tau\|^2).$$

This remainder generates the tracking-error term (23) via the calibration-error decomposition of Section 2.7. □

Proposition 4.8 quantifies the tradeoff between two sources of hedging cost:

Calibration regime	Curvature cost $\mathsf{C}_{\mathrm{curv}}$	Tracking error $\mathsf{C}_{\mathrm{track}}$
High frequency (continuous)	Active: scales with $D^2\Theta$ and $\mathrm{d}\langle P, P\rangle$	Zero
Low frequency (static)	Zero	Active: scales with $\|P_t - P_\tau\|^2$

When the calibration map Θ is highly nonlinear (large $D^2\Theta$), frequent recalibration amplifies curvature costs. When vanilla prices drift far from the calibration snapshot (large $\|P_t - P_\tau\|$), static recalibration accumulates tracking error. Neither regime dominates in general; the optimal recalibration frequency depends on the relative magnitudes of $D^2\Theta$, realized vanilla covariation $\mathrm{d}\langle P, P\rangle$, and parameter drift $\|\Theta(P_t) - \Theta(P_\tau)\|$.

5 Discussion and conclusion

5.1 When the framework applies

The traded-coordinate calculus applies whenever the market can be represented by a finite-dimensional vector of traded price semimartingales and the exotic value can be expressed as a sufficiently regular function of those coordinates, possibly augmented by a finite contract state. The model is naturally aligned with desks that hedge structured products using a small number of liquid hedging instruments, such as the underlying and a set of liquid vanillas, and that track contract state through a deterministic monitoring logic. In that setting the decomposition in Theorem 2.7 and Proposition 2.9 is not an approximation. It is an identity that isolates the

residual term attributable to covariation mismatch, given the hedging rule implied by the chosen value function.

The projection viewpoint applies whenever the discounted exotic value is a square-integrable martingale under the pricing measure, which is the standard setting for mean-square hedging. It is particularly useful when the factor set is a strict subset of all traded instruments or when liquidity considerations force the hedge to be carried out in a low-dimensional factor space. In such cases the orthogonal component in Theorem 2.12 clarifies a point that is sometimes obscured in practice: even with perfect calibration, a limited factor set can leave an irreducible residual risk that cannot be removed by any self-financing strategy restricted to those factors.

The feedback fixed-point layer applies when one wishes to model a regime in which hedging flows enter the dynamics of traded factors. Its main virtue is conceptual and mathematical clarity. It makes explicit what must be assumed for hedging to become destabilizing: a sufficiently strong map from state to flow and a sufficiently strong impact map from flow to traded coordinates. Under such assumptions the model yields an amplification operator and a covariation conjugation formula. Both results are theorems, not narratives, and they provide a clean language for scenario analysis. In particular, they show that feedback can act by distorting the realized quadratic covariation structure of traded prices, which is exactly the object that drives the residual term in the traded-coordinate hedging-error identity.

5.2 When the framework does not apply and what it deliberately leaves open

The framework is not intended as a universal explanation for all hedging losses. Several limitations are important.

Regularity is one limitation. The traded-coordinate calculus assumes sufficient smoothness of the value function in the chosen coordinates. Many exotic structures induce kinks or discontinuities, and a fully rigorous treatment may require generalized Itô calculus, local time terms, or viscosity-solution techniques. The paper handles the most consequential discontinuities—those induced by monitoring dates—through the explicit jump bookkeeping in Proposition 2.9, and discusses the non-smooth extension issues in Appendix C.

Identification is another limitation. Consistent with Sections 4.2 and 4.3, the empirical protocol identifies covariation objects and their regime dependence, not causal mechanisms; attributing an observed covariation distortion to hedging feedback requires additional identification structure (e.g. flow proxies, exogenous instruments, or richer positioning data). The feedback model is therefore presented primarily as a scenario layer consistent with the observed covariation structure rather than as a uniquely identified explanation.

Execution is a further limitation. The traded-coordinate identity is a *mark-to-model/mark-to-factor* identity in the sense that it is written against a specified traded-price semimartingale $\widetilde{Y}$ and assumes self-financing trading in those coordinates. In stress regimes, quoted "prices" may be stale, spreads may widen sharply, and rebalancing may be feasible only with substantial execution shortfall. In such cases the identity remains correct for the chosen book-price process, but realized P&L includes additional implementation terms. A minimal reduced-form way to capture this is to include an explicit cumulative execution-cost process in the self-financing equation; Section 2.4 records the corresponding exact add-on term.

Microstructure is deliberately simplified. The fixed-point impact specification is intentionally reduced-form and instantaneous: it yields a transparent equilibrium map and a stability criterion, but it abstracts from latency, transient impact, path dependence, and nonlinear impact regimes. These effects can be incorporated by replacing the pointwise fixed point with a dynamic clearing relation (e.g. a transient linear-impact kernel leading to a causal Volterra equation) and/or by allowing nonlinear response in the flow-to-price map, but doing so requires additional assumptions and yields a more complex stability analysis. The present paper treats the pointwise model as a quasi-static benchmark: it isolates the amplification mechanism in a form that can be stress-tested, while leaving richer refinements as extensions.

Factor resolution also matters. If the chosen traded coordinate vector does not capture the relevant surface movements for the product corridor, the GKW residual may dominate. In that case the appropriate conclusion is not that the covariation channel is irrelevant, but that the factor resolution is too coarse to isolate it. This is why the paper insists on reporting discipline for factor choices and why it treats factor selection as part of the empirical design rather than as an implicit assumption.

5.3 What is hedged, what is priced, and what is stress-tested

The decomposition in Theorem 2.7 is compatible with three distinct uses that should not be conflated.

When the aim is replication, one seeks conditions under which the residual term is identically zero. The identity shows that this requires both correct covariation (relative to the hedging model) and sufficient spanning power of the factor set. In practice these conditions are stringent, especially for callable and barrier-rich products, and the paper does not assume them unless explicitly stated.

When the aim is pricing under model risk, one may use the identity to quantify how sensitive the hedging P&L is to deviations of realized covariation from the model-implied covariation. In this interpretation the Hessian weighting identifies which covariation directions matter most. This supports model comparison and robust pricing design, but it does not, by itself, convert a descriptive decomposition into a pricing theorem.

When the aim is stress testing, one uses the diagnostic quantities in Section 3.7 together with either empirical covariation estimates or scenario covariation distortions implied by a feedback specification. This use is the most direct and least assumption-heavy. It acknowledges that feedback is a conditional possibility and assesses potential amplification under clearly stated scenarios rather than asserting that amplification must occur.

5.4 Implications for monitoring and governance

The main practical implication is methodological: feedback risk is most transparently monitored through explicitly defined traded (or proxy) factors, their covariation objects, and clearly stated stress scenarios, rather than through narrative labels. At a minimum, a monitoring stack fixes the factor set (Section 3.8), tracks covariation mismatch diagnostics (Sections 3.9 and 3.11), and, when using the feedback layer for scenario analysis, reports stability margins and amplification indices (Definition 3.13).

- *Risk management:* monitor factor coverage and covariation mismatch in the product-relevant corridor, and stress Λ and $D_y\Phi$ to identify regimes where SM_t compresses and AI_t grows.
- *Supervision:* where aggregate position data are available, the same objects provide a disciplined basis for identifying markets in which homogeneous positioning and limited depth can plausibly produce reflexive instability.

5.5 Concluding remarks

This paper develops a traded-coordinate framework for understanding and diagnosing hedging losses that are exacerbated when hedging activity feeds back into traded factor dynamics. The starting point is an exact identity: once the exotic value is represented as a sufficiently regular function of traded coordinates and contract state, and once a self-financing hedge is constructed by first-order matching in those coordinates, the residual hedging P&L is governed by a second-order term that contracts the value Hessian with the discrepancy between realized and model-implied quadratic covariation of traded prices. The identity is model-free in the sense

that it does not presume a specific diffusion law for the factors; it requires only semimartingale calculus and explicit bookkeeping for monitoring-date cashflows.

A complementary identification result is obtained through the Galtchouk-Kunita-Watanabe decomposition. It clarifies, within a given traded factor set, the precise separation between spanned risk that can be eliminated by self-financing trading and orthogonal residual risk that cannot. This separation matters in practice because factor compression and liquidity constraints can materially affect what "hedging the surface" means in implementable terms.

The feedback layer is then introduced as a fixed-point equilibrium that couples a state-dependent hedging flow map to a linearized impact clearing relation. Under explicit contraction and smoothness conditions the fixed point is well-posed, and perturbations of the fundamental state are propagated to the observed traded state through an amplification operator. This operator conjugates quadratic covariation, yielding a precise channel by which feedback can distort the covariation structure that drives the traded-coordinate hedging residual. The feedback model is conditional by design. Its value lies in making explicit what must be assumed for amplification to occur and in producing scenario quantities—stability margins and amplification indices—that can be evaluated without narrative inference.

To ensure that empirical conclusions remain commensurate with what data can support, the paper specifies an empirical protocol that defines estimands, estimators, inference methods, and robustness checks before reporting numerical findings. The protocol focuses on realized covariation of traded factors and on model-weighted covariation mismatch functionals, thereby aligning empirical statements with the objects that enter the theoretical identities.

Several natural extensions are suggested by the framework. The most immediate is to treat non-smooth payoffs and monitoring-induced kinks using generalized Itô formulas and to examine whether the resulting local-time terms have material diagnostic content in callable products. A second direction is to enrich the feedback model to allow for nonlinear impact and strategic interaction among multiple agents while preserving a tractable stability analysis. A third direction is to pursue stronger empirical identification of feedback mechanisms using richer datasets that include flow proxies or dealer positioning, thereby moving from covariation diagnostics toward causal inference.

When hedging is not price-taking, the relevant risk is not only a mismatch between a parametric model and realized paths. It is a mismatch between the covariation structure implicit in the hedging rule and the covariation structure realized by traded prices, which may itself be endogenously shaped by hedging flows. Expressing this mechanism in traded coordinates yields identities and diagnostics that are rigorous, transparent, and directly compatible with reproducible empirical testing.

Appendices

A Mathematical preliminaries and notation

We work on a filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\in[0,T]}, \mathbb{Q})$ satisfying the usual conditions, with $\mathbb{Q}$ a pricing measure for traded assets. The numéraire $B = (B_t)_{t\in[0,T]}$ is strictly positive, continuous, and of finite variation. For any traded price process U we write $\widetilde{U}_t := U_t/B_t$ for its discounted version. The market contains an underlying $S = (S_t)_{t\in[0,T]}$ and liquid vanilla instruments $P^{(1)}, \dots, P^{(n)}$ with traded price processes. We collect the traded factors into $Y := (S, P^{(1)}, \dots, P^{(n)})$. All price processes are real-valued, càdlàg semimartingales adapted to $(\mathcal{F}_t)$.

For semimartingales Y and Z we denote by $\langle Y, Z\rangle$ their quadratic covariation, defined as the unique adapted finite-variation process such that $YZ - \langle Y, Z\rangle$ is a semimartingale with no continuous local martingale part. When Y is scalar we write $\langle Y\rangle := \langle Y, Y\rangle$. When Y and Z are vector-valued semimartingales, $\langle Y, Z\rangle$ denotes the matrix-valued covariation whose (i, j)

entry is $\left\langle Y^{(i)}, Z^{(j)} \right\rangle$. For a continuous $\mathbb{R}^d$-valued semimartingale Y and scalar C^1 functions $f, g : \mathbb{R}^d \to \mathbb{R}$ with suitable growth, we use the identity

$$\mathrm{d}\,\langle f(Y), g(Y)\rangle_t = \big(\nabla f(Y_t)\big)^\top \,\mathrm{d}\,\langle Y, Y\rangle_t \,\nabla g(Y_t),$$

that is,

$$\langle f(Y), g(Y)\rangle_t = \int_0^t \big(\nabla f(Y_u)\big)^\top \,\mathrm{d}\,\langle Y, Y\rangle_u \,\nabla g(Y_u).$$

When f, g are vector-valued, the same formula holds componentwise.

Theoretical statements involving replication, self-financing gains, or martingale properties are formulated under $\mathbb{Q}$; empirical estimation of covariation from data is under the physical measure $\mathbb{P}$.

A trading strategy is a predictable process $\Theta_t = \big(\eta_t, \phi_t, \psi_t^{(1)}, \ldots, \psi_t^{(n)}\big)$, where η_t denotes the holding in the numéraire B, ϕ_t the holding in the underlying S, and $\psi_t^{(i)}$ the holding in the vanilla instrument $P^{(i)}$. The associated portfolio value is $X_t = \eta_t B_t + \phi_t S_t + \sum_{i=1}^n \psi_t^{(i)} P_t^{(i)}$.

Definition A.1 (Admissibility)**.** A strategy Θ is *admissible* if it is predictable and such that the stochastic integrals

$$\int_0^t \phi_u \,\mathrm{d}S_u \qquad \text{and} \qquad \int_0^t \psi_u^{(i)} \,\mathrm{d}P_u^{(i)}, \quad i = 1, \ldots, n,$$

are well-defined for all $t \in [0, T]$. In addition, we impose the standard no-doubling condition: writing $\widetilde{X}_t := X_t / B_t$ for the discounted wealth, there exists a constant $a \in \mathbb{R}$ such that

$$\widetilde{X}_t \geq -a \qquad \text{a.s. for all } t \in [0, T].$$

This lower bound rules out doubling strategies and ensures that later projection and identification arguments are not contaminated by pathological self-financing constructions.

Definition A.2 (Self-financing)**.** An admissible strategy Θ is *self-financing* if its value process satisfies

$$\mathrm{d}X_t = \eta_t \,\mathrm{d}B_t + \phi_t \,\mathrm{d}S_t + \sum_{i=1}^n \psi_t^{(i)} \,\mathrm{d}P_t^{(i)}$$

for all $t \in [0, T]$.

Assumption A.3 (Continuous semimartingale setting)**.** The discounted traded price vector $\widetilde{Y} := (\widetilde{S}, \widetilde{P}^{(1)}, \ldots, \widetilde{P}^{(n)})$ is a continuous semimartingale under $\mathbb{Q}$.

Assumption A.4 (Traded-coordinate smoothness)**.** There exists a function $v : [0, T] \times \mathbb{R}^{1+n} \times \Xi \to \mathbb{R}$, where Ξ is a finite set of contract states, such that

$$V_t = v(t, S_t, P_t, \xi_t)$$

for all $t \in [0, T]$, and such that for each fixed $\xi \in \Xi$ the map $(t, s, p) \mapsto v(t, s, p, \xi)$ is $C^{1,2}$ in $(t, (s, p))$ with derivatives of at most polynomial growth on the relevant domain.

The results in Section 2 are expressed in terms of semimartingale covariation of traded prices. Statistical implied-volatility indices and other non-traded state variables can be used as empirical covariates or proxies, but they are not traded coordinates unless explicitly mapped to traded instruments.

The finiteness of Ξ is a notational simplification that covers typical autocallable and barrier logic (finite state machines); countable state spaces can be handled with no essential changes.

Finally, $C^{1,2}$ regularity is imposed to apply Itô's formula in the needed form. In applications, barrier and call logic can induce kinks; one then localizes away from singular sets, smooths, or uses generalized Itô formulas (bounded-variation/viscosity regularity) depending on the use case.

B Technical complements for the traded-coordinate calculus

This appendix collects technical points that are routinely used in the main text but can be difficult to track when they are embedded inline. The purpose is not to add new modelling content, but to ensure that every algebraic step used in the identities has an explicit justification and that no hidden regularity or measurability assumptions are implicitly invoked.

B.1 Trace contractions against matrix-valued covariation measures

Let Y be a continuous $\mathbb{R}^d$-valued semimartingale. Its quadratic covariation $\langle Y, Y\rangle$ is a matrix-valued finite-variation process, and for any predictable matrix-valued process $H_t \in \mathbb{R}^{d\times d}$ such that the integrals below are well-defined, the scalar process

$$\int_0^t \mathrm{Tr}\big(H_u \, \mathrm{d}\,\langle Y, Y\rangle_u\big)$$

is defined entrywise by

$$\int_0^t \mathrm{Tr}\big(H_u \, \mathrm{d}\,\langle Y, Y\rangle_u\big) = \sum_{i=1}^{d}\sum_{j=1}^{d} \int_0^t H_u^{ij} \, \mathrm{d}\left\langle Y^j, Y^i\right\rangle_u .$$

This convention matches the one implicitly used in Itô's formula for vector-valued semimartingales.

Whenever H is symmetric and $\langle Y, Y\rangle$ is symmetric, the cyclic property of the trace yields

$$\mathrm{Tr}(H\,C) = \mathrm{Tr}(C\,H)$$

for any $C \in \mathbb{R}^{d\times d}$. In particular, the sign convention in Equation (4) is unaffected by whether one writes $\mathrm{Tr}(D^2 v \, \mathrm{d}\,\langle Y, Y\rangle)$ or $\mathrm{Tr}(\mathrm{d}\,\langle Y, Y\rangle \, D^2 v)$.

More generally, since $\langle Y, Y\rangle$ is symmetric, only the symmetric part of H contributes:

$$\mathrm{Tr}\big(H_u \, \mathrm{d}\,\langle Y, Y\rangle_u\big) = \mathrm{Tr}\Big(\tfrac{1}{2}(H_u + H_u^\top) \, \mathrm{d}\,\langle Y, Y\rangle_u\Big).$$

Thus no separate symmetry assumption on the weighting matrix is needed beyond the smoothness that ensures $D^2 v$ exists; any antisymmetric component is annihilated by the symmetric covariation matrix.

B.2 A complete derivation of the covariation of a stochastic integral

The proof of Theorem 2.12 uses the identity

$$\left\langle \int_0^{\cdot} \theta_u^\top \, \mathrm{d}M_u, M \right\rangle_t = \int_0^t \theta_u^\top \, \mathrm{d}\,\langle M, M\rangle_u$$

for a continuous martingale M and a predictable integrand θ. This is standard; see e.g. Protter [2010, Ch. II].

Lemma B.1. *Let M be a continuous $\mathbb{R}^d$-valued local martingale and let θ be a predictable $\mathbb{R}^d$-valued process such that the stochastic integral $N_t := \int_0^t \theta_u^\top \, \mathrm{d}M_u$ is well-defined. Then for each $i \in \{1, \dots, d\}$,*

$$\left\langle N, M^{(i)} \right\rangle_t = \int_0^t \theta_u^\top \, \mathrm{d}\left\langle M, M^{(i)} \right\rangle_u ,$$

and in vector form,

$$\langle N, M\rangle_t = \int_0^t \theta_u^\top \, \mathrm{d}\,\langle M, M\rangle_u .$$

Proof (standard; sketch). By localization we may assume M is square-integrable. Approximate θ in $L^2(\mathrm{d}\langle M,M\rangle)$ by simple predictable processes $\theta^{(n)}$ so that $N^{(n)} := \int \theta^{(n)\top}\,\mathrm{d}M$ converges to N in ucp. For a simple $\theta^{(n)} = \sum_k h_k \mathbf{1}_{(t_k,t_{k+1}]}$ one has $N^{(n)} = \sum_k h_k^\top (M_{t\wedge t_{k+1}} - M_{t\wedge t_k})$, and bilinearity of covariation gives $\left\langle N^{(n)}, M\right\rangle_t = \int_0^t \theta_u^{(n)\top}\,\mathrm{d}\langle M,M\rangle_u$. Passing to the limit using ucp-continuity of quadratic covariation yields the claim; see Protter [2010, Ch. II] for a full proof. □

C Extensions beyond the smooth continuous setting

The main body adopts a smooth continuous setting to keep the central message transparent: in traded coordinates, covariation mismatch is the driver of residual P&L after first-order matching. Many structured payoffs are not globally $C^{1,2}$ and many traded prices exhibit jumps at market open or during stress. This appendix records, in theorem-level language, the precise additional terms that appear when one relaxes these assumptions. The purpose is to prevent a logic gap: the main identity is not presented as "universal"; rather, it is presented as the clean core to which additional terms are appended under weakened regularity.

C.1 Jump semimartingales: the exact additional term

Let Y be a càdlàg semimartingale with jumps and let $v(t,y)$ be $C^{1,2}$ in (t,y). The Itô formula then includes a jump correction term involving the compensated jump measure. In the present context this term enters the hedging error exactly in the same way as the continuous covariation term does: it is not cancelled by first-order matching unless one trades in instruments that span the jump risk.

Theorem C.1 (Hedging-error identity with jumps)**.** *Assume the setting of Theorem 2.7 except that $\widetilde{Y}$ is now a general càdlàg semimartingale and v is $C^{1,2}$. Assume that the hedging strategy uses left-continuous (predictable) holdings*

$$\vartheta_t = \nabla_y v(t, \widetilde{Y}_{t-}, \xi),$$

meaning that the position just before time t is determined by the value-function gradient evaluated at the pre-jump state. Let $\widetilde{X}$ be the discounted self-financing wealth process with $\widetilde{X}_{t_0} = v(t_0, \widetilde{Y}_{t_0}, \xi)$. Then, on any interval where ξ is constant,

$$\begin{aligned}\widetilde{X}_t - v(t,\widetilde{Y}_t,\xi) = &-\frac{1}{2}\int_{t_0}^t \mathrm{Tr}\Big(D_{yy}^2 v(u,\widetilde{Y}_{u-},\xi)\,\big(\mathrm{d}\left\langle \widetilde{Y}^c, \widetilde{Y}^c\right\rangle_u - \widehat{c}(u,\widetilde{Y}_{u-},\xi)\,\mathrm{d}u\big)\Big)\\ &- \sum_{t_0<u\le t}\Big(v(u,\widetilde{Y}_u,\xi) - v(u,\widetilde{Y}_{u-},\xi) - \nabla_y v(u,\widetilde{Y}_{u-},\xi)^\top \Delta\widetilde{Y}_u\Big),\end{aligned} \tag{25}$$

where $\widetilde{Y}^c$ denotes the continuous local martingale part of $\widetilde{Y}$ and $\Delta\widetilde{Y}_u := \widetilde{Y}_u - \widetilde{Y}_{u-}$.

Proof (standard; sketch). Apply Itô's formula for càdlàg semimartingales to $u \mapsto v(u,\widetilde{Y}_u,\xi)$ on an interval where ξ is constant: the differential consists of a time term, a stochastic integral $\nabla_y v(u,\widetilde{Y}_{u-},\xi)^\top\,\mathrm{d}\widetilde{Y}_u$, a continuous quadratic covariation term involving $\mathrm{d}\left\langle \widetilde{Y}^c, \widetilde{Y}^c\right\rangle$, and the jump remainder $\sum(v(u,\widetilde{Y}_u) - v(u,\widetilde{Y}_{u-}) - \nabla v(u,\widetilde{Y}_{u-})^\top\Delta\widetilde{Y}_u)$. Subtract the discounted self-financing wealth increment $\mathrm{d}\widetilde{X}_u = \vartheta_u^\top\,\mathrm{d}\widetilde{Y}_u$ with $\vartheta = \nabla v(\cdot,\widetilde{Y}_-,\xi)$ to cancel the stochastic integral, and use the model covariation density to rewrite the time term as in Theorem 2.7. This yields (25); see Protter [2010, Ch. IV]. □

Theorem C.1 clarifies that jumps create an additional residual term that is quadratic in jump size and weighted by the local curvature of the value function. In particular, the additional term

$$\sum_{t_0<u\le t}\Big(v(u,\widetilde{Y}_u,\xi) - v(u,\widetilde{Y}_{u-},\xi) - \nabla_y v(u,\widetilde{Y}_{u-},\xi)^\top\Delta\widetilde{Y}_u\Big)$$

plays the same role for jumps as the Hessian contraction does for the continuous covariation: it is the second-order (and higher) remainder that cannot be eliminated by first-order matching unless one trades instruments that span the jump risk. This is consistent with the continuous-time message: after first-order matching, second-order terms drive residuals. It also clarifies an empirical point: if jumps are present and relevant, daily sampling may miss intraday jump structure, and the empirical protocol must then be adapted to a jump-robust realized variation estimator. The paper does not claim otherwise.

C.2 Non-smooth payoffs: local time terms

Structured payoffs can induce kinks in value functions even when traded factors are continuous. In one dimension, the correct replacement for classical Itô is the Meyer-Tanaka formula, which introduces local time at the kink. In higher dimensions, one may work with convex functions and generalized gradients, or with viscosity solutions and mollification. We record the one-dimensional statement, as it is already sufficient to show that nonsmoothness can produce additional residual terms not captured by the classical Hessian contraction.

Theorem C.2 (Meyer-Tanaka formula for an absolute value kink)**.** *Let U be a continuous semimartingale and define $f(u) = |u|$. Then*

$$|U_t| = |U_0| + \int_0^t \operatorname{sgn}(U_u)\, \mathrm{d}U_u + L_t^0(U),$$

where $\operatorname{sgn}(u) = \mathbf{1}_{u>0} - \mathbf{1}_{u<0}$ *and* $L^0(U)$ *is the local time of* U *at* 0*.*

Proof (standard). This is the Meyer–Tanaka (Tanaka) formula for the convex function $u \mapsto |u|$; see e.g. Protter [2010, Ch. IV]. □

In the hedging context, Theorem C.2 implies that for value functions with kinks in a traded coordinate, the residual after first-order matching includes a local time term at the kink. This term is not captured by the smooth Hessian contraction and must be accounted for if one aims for a fully rigorous treatment of payoffs with sharp barriers or digital components. The main body avoids this technicality by working in smooth settings and by making monitoring-date discontinuities explicit through jump bookkeeping; the present appendix shows precisely what changes when kink terms are present.

D Reproducibility statement

The theoretical statements in this paper are fully specified by explicit assumptions and proofs. Any empirical statement is, by design, an output of a code path that implements the protocol of Section 4. The reproducibility package therefore fixes, for every reported figure and table, the dataset version, the preprocessing rule, the factor definition, the estimator, and the inference method. In particular, whenever a volatility proxy is used as a factor, its status as traded or non-traded is stated explicitly in the caption and the interpretation is restricted accordingly.

The paper does not require the reader to accept narrative causal explanations. It requires only that the mathematical identities are checked line by line and, for empirical claims, that the provided code reproduces the reported outputs from the stated inputs.

E Interpreting prior work within the traded-coordinate framework

This appendix provides a careful, theorem-level interpretation of Salon [2019]. The objective is not to dispute its practical motivation. The note isolates a mechanism that is consistent with

desk experience: a callable structure whose continuation feature makes vanna-type rebalancing persistent, and a resulting tendency for second-order terms to dominate hedged P&L in stress regimes. The contribution here is to restate the key steps in traded coordinates, to identify precisely which additional assumptions are required for the derivations in Salon [2019] to be mathematically valid as hedging-error identities, and to record the exact correction terms that appear when those assumptions are not imposed.

E.1 Mapping the "carry P&L" formula to our identity

Salon [2019] proposes a schematic decomposition of incremental hedging P&L of the form

$$\text{CarryP\&L}_{t\to t+\,\mathrm{d}t} \approx \frac{1}{2}\sum_k \frac{\partial^2 \Pi}{\partial S\,\partial \alpha_k}\Big(\text{RealisedCov}(\,\mathrm{d}S,\,\mathrm{d}\alpha_k) - \text{ModelCov}(\,\mathrm{d}S,\,\mathrm{d}\alpha_k)\Big),$$

where Π denotes the value of a cancellable component of an autocall and (α_k) denotes a reduced set of surface or market parameters (at-the-money volatility, skew, dividends, repo, and similar quantities in the note's terminology).

In the present paper, the corresponding statement is Theorem 2.7: after first-order matching in a *traded* coordinate vector Y (discounted spot together with discounted traded vanillas used as coordinates), the discounted hedging error equals a second-order contraction of the value Hessian with a covariation mismatch measure,

$$\mathrm{d}(\widetilde{X}_t - \widetilde{V}_t) = -\frac{1}{2}\mathrm{Tr}\Big(D^2_{yy} v(t,\widetilde{Y}_t,\xi)\,(\,\mathrm{d}\left\langle \widetilde{Y},\widetilde{Y}\right\rangle_t - \widehat{c}(t,\widetilde{Y}_t,\xi)\,\mathrm{d}t)\Big),$$

on intervals where contract state is constant and cashflows are matched.

The two expressions become the same object once α is not treated as a primitive factor but is instead expressed as a function of traded prices. This is the content of Theorem 4.6 in the main text, reproduced here in the language of the decomposition in Salon [2019].

Theorem E.1 (A rigorous version of the vanna-carry decomposition)**.** *Assume the hypotheses of Theorem 2.7 for a traded coordinate vector $\widetilde{Y} = (\widetilde{S}, \widetilde{P})$, where $\widetilde{P}$ is the vector of discounted traded vanilla prices used as surface coordinates. Let α be a reduced parameter vector computed from $\widetilde{P}$ by a C^2 calibration map $\alpha = \Theta(\widetilde{P})$, and suppose the relevant component value admits a representation $\Pi_t = u(t, S_t, \alpha_t, \xi)$ with $u \in C^{1,2}$.*

Define $v(t,s,p,\xi) := u(t,s,\Theta(p),\xi)$ and construct the self-financing first-order matched hedge using $\nabla_{(s,p)}v$ in the traded instruments (S,P). Then the hedging-error identity of Theorem 2.7 holds exactly. Moreover, if one rewrites the Hessian $D^2_{(s,p)(s,p)}v$ via the chain rule, the resulting contraction against $\mathrm{d}\,\langle (S,P),(S,P)\rangle$ produces cross terms that coincide with the "vanna" terms of Salon [2019] after *transporting derivatives through the calibration Jacobian $D_p\Theta$, and it produces an additional calibration-curvature term involving $D^2\Theta$.*

Proof. This is an immediate specialization of Theorem 4.6, whose proof gives the exact Hessian expansion. The hedging-error identity is Theorem 2.7. The only substantive point is that the coordinate vector must be traded for the self-financing interpretation to hold, which is ensured by working with (S,P) and expressing α as a function of P rather than as an independent factor. □

Theorem E.1 clarifies a central limitation of parameter-based decompositions, including that in Salon [2019]. A decomposition written directly in terms of (S,α) is not, by itself, a hedging P&L identity unless α is either traded or explicitly linked to traded prices through a smooth map. The traded-coordinate framework supplies precisely that link, and it shows the exact correction term that is otherwise omitted: the calibration-curvature term $\sum_\ell u_{\alpha_\ell} D^2\Theta_\ell$ in (20).

E.2 The key approximation in the pricing proxy and the missing covariance term

A second central step of Salon [2019] is an approximation for the cancellable component Π that factorizes a product inside a conditional expectation. In the notation of that work, the cancellable feature is written schematically as a product of a down-and-in put payoff with a survival indicator for non-cancellation along the monitoring grid, and the conditional expectation is approximated by the product of conditional expectations.

The approach misses a covariance term in the proxy of Salon [2019] whose magnitude can be bounded. The bound is both model-free and directly testable in simulation under any candidate dynamics. Which clarifies what would be needed to justify the factorization as a controlled approximation: one needs either an argument that $\mathrm{Var}(V \mid \mathcal{G})$ is small in the relevant regime, or an argument that U is nearly conditionally deterministic given $\mathcal{G}$ in the relevant regime, or a structural reason for conditional near-independence. Absent such a justification, the factorization should be treated as a modelling approximation whose error can be non-negligible for strongly path-dependent structures.

The Cauchy–Schwarz bound is tight: equality holds if and only if $V - \mathbb{E}[V \mid \mathcal{G}]$ is a $\mathcal{G}$-measurable multiple of $U - \mathbb{E}[U \mid \mathcal{G}]$. More generally,

$$\mathrm{Cov}(U, V \mid \mathcal{G}) = \rho(U, V \mid \mathcal{G}) \sqrt{\mathrm{Var}(U \mid \mathcal{G})} \sqrt{\mathrm{Var}(V \mid \mathcal{G})},$$

so the factorization error is small only when conditional correlation and conditional variances are controlled—a nontrivial check near barriers where U and V are mechanically coupled.

E.3 Model-implied spot–implied-volatility covariation and the quoting convention

Salon [2019] uses a "local volatility model covariance" between spot and an implied volatility coordinate. That work cites a result from Bergomi [2016] (reproduced there as an annex) that relates the sensitivity of at-the-money-forward implied volatility to spot through skew-type quantities, and it then interprets the resulting expression as a model-implied covariance term.

A precise statement requires specifying what is meant by an implied volatility coordinate as a stochastic process. In a traded-coordinate framework, the correct primitive objects are traded prices. Implied volatility is a transformation of traded option prices, and its increments therefore inherit Itô terms. While curvature affects the drift, quadratic covariation depends only on the Jacobian, exactly as in Theorem 4.6. Even before addressing any local-vol-specific approximation, one must therefore state the quoting convention and the transformation used to define the implied volatility coordinate.

To avoid a logic gap, we record the exact general identity in the simplest one-dimensional setting. Fix a maturity T^* and define the at-the-money-forward implied volatility $\widehat{\sigma}_t$ as a C^2 function of a traded option price P_t, namely $\widehat{\sigma}_t = \Psi(P_t)$ where Ψ is the inverse Black–Scholes map at that maturity and moneyness convention. Then, for a continuous semimartingale P,

$$\mathrm{d}\widehat{\sigma}_t = \Psi'(P_t)\ \mathrm{d}P_t + \frac{1}{2}\Psi''(P_t)\ \mathrm{d}\langle P, P\rangle_t\,.$$

Consequently,

$$\mathrm{d}\langle \ln S, \widehat{\sigma}\rangle_t = \Psi'(P_t)\ \mathrm{d}\langle \ln S, P\rangle_t\,.$$

This identity is exact and shows that any "model covariance" between spot and an implied volatility coordinate is, in traded terms, a statement about the spot–vanilla covariation $\mathrm{d}\langle \ln S, P\rangle$ transported through the Jacobian Ψ'.

The practical content of Salon [2019] can therefore be made fully consistent with the traded-coordinate framework as follows. One chooses a concrete traded option coordinate (or a small

set of them), defines the implied-volatility proxy via an explicit map, and then computes the resulting covariation terms either directly from realized covariation of traded prices or from a model-implied covariation density. In this form, the comparison between realized and model-implied covariation becomes a legitimate input to the hedging-error identity. Without these steps, a comparison between a historically estimated covariance of $(S, \widehat{\sigma})$ and a theoretically computed quantity risks conflating transformations and omitting curvature terms.

E.4 The "add-on" as a quadratic covariation contract

Salon [2019] proposes an add-on payoff intended to offset the expected negative carry attributed to vanna-type terms. In the traded-coordinate framework, the natural mathematical object corresponding to such an add-on is a contract whose payoff depends on realized quadratic covariation of traded factors, possibly weighted by a predictable process.

To state this precisely, fix a traded coordinate vector Y under the pricing measure and let H_t be a predictable matrix process representing the relevant second-order sensitivity weights (for instance $H_t = D^2_{yy} v(t, Y_t, \xi)$ for a chosen value representation). Consider the random variable

$$\mathcal{C}_T := \int_0^T \mathrm{Tr}(H_t \, \mathrm{d}\langle Y, Y\rangle_t). \tag{26}$$

This payoff is a quadratic covariation swap in the sense that it pays a linear functional of realized quadratic covariation. Contracts of this type are standard in variance-swap theory when H_t picks out the $\mathrm{d}\langle \ln S, \ln S\rangle$ component; here the payoff is written for general factor vectors and general weights.

The key point is that a statement of the form "the add-on pays the future expected negative carry" can be formalized as pricing (26) (or its covariation-mismatch analogue where $\mathrm{d}\langle Y, Y\rangle$ is replaced by $\mathrm{d}\langle Y, Y\rangle - \widehat{c}\,\mathrm{d}t$). This formalization clarifies what must be available to replicate such an add-on.

Proposition E.2 (Replicability of covariation contracts requires covariation instruments (standard))**.** *Assume a frictionless market in which, in addition to the factors Y, one can trade (at time 0) a family of claims whose payoffs span the components of $\int_0^T \mathrm{d}\langle Y^i, Y^j\rangle_t$ for $1 \le i \le j \le d$ (for example, variance swaps on each factor and suitable cross-variation swaps). Then any payoff of the form* (26) *with bounded predictable weight H is replicable by static trading in these covariation claims.*

By contrast, even when Y itself is traded, a payoff of the form (26) *is, in general,* not statically replicable *by a portfolio of claims whose payoffs depend only on the terminal state Y_T (i.e. payoffs of the form $f(Y_T)$).*

Proof. If one can trade, at time 0, claims paying each component $\int_0^T \mathrm{d}\langle Y^i, Y^j\rangle_t$, then any bounded predictable linear combination of these components is obtained by static linear combination of the corresponding claims, which replicates (26).

For the second statement, it suffices to give a counterexample showing that a covariation-linked payoff need not be measurable with respect to Y_T. Take $d = 1$ and $Y = W$ a Brownian motion. Let $H_t := \mathbb{1}_{\{W_t > 0\}}$. Since $\mathrm{d}\langle W, W\rangle_t = \mathrm{d}t$, the payoff (26) becomes

$$\mathcal{C}_T = \int_0^T \mathbb{1}_{\{W_t > 0\}} \, \mathrm{d}t,$$

the occupation time of the positive half-line. This random variable is not $\sigma(W_T)$-measurable (hence cannot be written as $f(W_T)$ for any Borel f), which proves the claim. □

Proposition E.2 is a static statement. Dynamic attainability of (26) by trading only in Y depends on completeness properties of the market (e.g. a predictable representation property);

the practical point is that covariation-linked overlays are most transparent when implemented via instruments that directly reference realized variation rather than via terminal-only payoffs.

Proposition E.2 does not invalidate the add-on proposal in Salon [2019]. It clarifies the instrument and market assumptions under which such an add-on can be viewed as a hedgable overlay. It also highlights the role of factor choice: if the add-on is defined in terms of implied-volatility proxies rather than traded prices, then one must either trade the corresponding option instruments or accept that the add-on is a diagnostic device rather than a replicable payoff.

References

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

Jun Kyung Auh and Wonho Cho. Liquidation cascade and hedging front-running: Evidence from the structured equity product market. *SSRN Electron. J.*, 2020.

Kujtim Avdiu and Stephan Unger. Implicit hedging and liquidity costs of structured products. *J. Risk Fin. Manag.*, 16(9):401, September 2023.

O E Barndorff-Nielsen. Power and bipower variation with stochastic volatility and jumps. *J. Financ. Econom.*, 2(1):1–37, December 2004.

Alexander Barzykin, Philippe Bergault, and Olivier Guéant. Algorithmic market making in dealer markets with hedging and market impact. *Math. Financ.*, 33(1):41–79, January 2023.

Ahmed Bel Hadj Ayed and Gregoire Loeper. The cost of vega-hedging structured products. *SSRN Electron. J.*, 2023.

Lorenzo Bergomi. *Stochastic Volatility Modeling*. Chapman and Hall/CRC, 2016. ISBN 978-1-4822-5434-0. doi: 10.1201/b19393.

Bruno Bouchard, Grégoire Loeper, and Yiyi Zou. Almost-sure hedging with permanent price impact. *Finance and Stochastics*, 20(3):741–771, 2016. doi: 10.1007/s00780-016-0303-5.

Bruno Bouchard, Grégoire Loeper, and Yiyi Zou. Hedging of covered options with linear market impact and gamma constraint. *SIAM J. Control Optim.*, 55(5):3319–3348, 2017.

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

Umut Cetin, Robert A Jarrow, and Philip Protter. Liquidity risk and arbitrage pricing theory. *Finance Stoch.*, 8(3):311–341, August 2004.

Sebastian Egebjerg and Thomas Kokholm. A model for the hedging impact of option market makers. *SSRN Electron. J.*, 2024.

Anne Eyraud-Loisel. Quadratic hedging approaches and applications. Lecture notes, 2024.

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

Hans Föllmer and Martin Schweizer. Hedging of contingent claims under incomplete information. In M H A Davis and R J Elliott, editors, *Applied Stochastic Analysis*, volume 5 of *Stochastics Monographs*, pages 389–414. Gordon and Breach, 1991.

Hamza Guennoun. Understanding autocalls: Real time vega map. *SSRN Electron. J.*, 2019.

Hiroshi Kunita. *Stochastic Flows and Jump-Diffusions*. Springer Singapore, Singapore, 2019.

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

Gregory Salon. Equity autocalls and vanna negative carries: Pricing and hedging with a simple add-on. *SSRN Electron. J.*, 2019.

Martin Schweizer. On the minimal martingale measure and the föllmer–schweizer decomposition. *Stochastic Analysis and Applications*, 13:573–599, 1995. URL `https://people.math.ethz.ch/~mschweiz/Files/minimal.pdf`.