Finite Gaussian-mixture martingale kernels: density for projected smile-jet readouts

Zeyu Cao[†] Shaosai Huang[‡]

September 27, 2026

Abstract

The skew-stickiness ratio and its higher-order extensions measure how the implied-volatility smile responds to moves in the spot. They are what a quoted surface reveals about the martingale kernel generating it, a kernel that is never observed. We ask whether finite Gaussian-mixture martingale chains, whose option prices are finite sums of Black–Scholes prices, can reproduce those dynamics. They can. For stochastic-volatility models with a variance floor and correlation bounded away from ± 1, and for local-stochastic-volatility models, under explicit regularity assumptions, we construct exact-martingale finite Gaussian-mixture chains that approximate the marginals and match the finite-step skew-stickiness ratio and its extensions to any prescribed order and accuracy. Part of the contribution is the notion of convergence: the topology generated by the dynamics characteristics themselves, which neither weak nor adapted Wasserstein convergence controls. With M components per node and a d-dimensional latent state the readout error is $O(M^{-1/(d+1)})$ up to logarithms, linear in the quantization error, when returns carry a Gaussian component of fixed variance; without one, adapted approximation forces the innovations preceding a price-dependent kernel to vanish, and our bound degrades with the readout order. The limit is the wings: every finite mixture has zero implied-variance wing slopes. Adopting the class therefore costs nothing in at-the-money smile dynamics of any finite order. In a fixed-fitted-marginal finite-chain experiment on an SPX surface, fixing the first three characteristics narrows the at-the-money forward-start range from 7.0 to 4.0 volatility points, and across fifty-five further monthly surfaces the same experiment removes a median 39% of that range.

Keywords: smile dynamics; skew-stickiness ratio; stochastic-local volatility; Gaussian mixtures; martingale transition kernels; adapted Wasserstein distance; arbitrage-free calibration.

JEL classification: G13; C65; C63; G12.

MSC 2020: 91G20 (primary); 60G42, 60J05, 60B10, 49Q22, 91G60 (secondary).

1 Introduction

An arbitrage-free European surface fixes the risk-neutral marginal law of the underlying at each maturity, but not the martingale kernels joining those laws; kernels with identical marginals produce different continuation smiles, spot–volatility responses and path-dependent prices. For

Working paper. Comments welcome.

[†]Independent Researcher, Long Island City, NY 11101, USA. Email: josephcao891011@gmail.com.

[‡]Kspectra Research Inc., Toronto, ON, M2N 0G3, Canada. Email: arthur.foxie.huang@kspectra.ai.

independent-increment models the continuation smile is even deterministic in advance, however well the marginals fit, which is one form of the surface-versus-dynamics problem [19, Proposition 11.2 and Section 14.3]. The kernel generating those dynamics is never observed. What a surface does reveal is a hierarchy of at-the-money characteristics — the skew-stickiness ratio and its higher-order extensions — so a modelling class is adequate only if no admissible kernel displays dynamics the class cannot express in those coordinates. We ask whether finite Gaussian-mixture martingale chains meet that test: can they reproduce a prescribed finite order of at-the-money smile dynamics to arbitrary accuracy while approximating the marginals and remaining exact asset martingales? They can, at every finite order, for homogeneous stochastic volatility (SV) and for local stochastic volatility (LSV) (Theorems 7.3 and 9.13).

Why martingale kernels and Gaussian mixtures. The kernel is the natural object for calibrating dynamics: it carries exactly the freedom European prices leave, lives on the grid of quoted dates, prices every maturity and path-dependent claim from one law, and combines local and stochastic volatility; as an exact martingale chain it excludes static and dynamic arbitrage by construction. Finite Gaussian mixtures make such kernels tractable: calls and their strike derivatives are sums of Black–Scholes terms; martingality is one exponential-barycentre equation per node, linear in the weights, so price fitting with fixed components is a linear or quadratic program; positive variances give smooth densities and stable smile derivatives; and chains of mixture kernels have mixture marginals. Mixtures are an approximation basis, not an economic mechanism, with a long history in smile fitting [13] and a discrete stochastic-local-volatility implementation in [29]. The costs are thin tails and component counts that grow with accuracy and order.

The observable dynamics. What the kernel settles beyond today's prices is how the smile moves when the spot moves, and calibrating that motion needs a quantity the quotes show; the skew-stickiness ratio [10, 11] is the one desks use, and it is the lowest order of a family. Che and Das [18] grade the response of the smile to spot moves by a transport equation whose at-the-money coefficients extend the ratio to every order; we take that family in finite-step form, computed from the law of a chain on the grid rather than from a pathwise derivative. More specifically, on a grid $t_0 < \cdots < t_J$ with log price X_j and continuation state ζ_j, let $w_{j,\ell}(k,\zeta)$ be the total implied variance at log-moneyness k of the maturity-t_ℓ smile seen from $\zeta_j = \zeta$. The readout consists of at-the-money jets and their projected spot responses over the first step,

$$a_m^\ell = \partial_k^m w_{0,\ell}(0,\zeta_0), \qquad A_m^\ell = \partial_k^m w_{1,\ell}(0,\zeta_1), \qquad d_m^\ell = \frac{\operatorname{Cov}(A_m^\ell, X_1 - X_0)}{\operatorname{Var}(X_1 - X_0)}, \tag{1}$$

where m is the order in log-moneyness and ℓ the maturity, collected over a finite maturity set $\mathcal{L}$ and truncated at an order N in $\mathcal{R}^{\mathrm{raw}}_{N,\mathcal{L}} = \big((a_m^\ell)_{m\le N+1}, (d_m^\ell)_{m\le N}\big)_{\ell\in\mathcal{L}}$. At order zero, a_0^ℓ and a_1^ℓ are the level and the skew, and $v_0^\ell = d_0^\ell / a_1^\ell$ is the finite-step analogue of that ratio; on $a_1^\ell \neq 0$ the rows determine the transport coefficients $v_0^\ell, \dots, v_N^\ell$ of [18] by triangular inversion (Section 2.4). The higher orders are not decoration: the same study rejects $v_1 = v_2 = 0$ at every tenor from one month to twenty-four, on five years of SPX surfaces.

The approximation problem. Quantifying over unknown kernels is not possible, so we quantify instead over the classes practitioners posit. Given a target chain P in such a class, an

order N, a fixed $p \geq 1$ and $\varepsilon > 0$, the problem is to find a finite Gaussian-mixture chain Q with

$$\begin{aligned} \mathbb{E}_Q\big[e^{X_{j+1}} \mid \mathcal{F}_j\big] &= e^{X_j} \quad (0 \leq j < J), \\ \max_{1\leq j\leq J} W_p\big(\mathrm{Law}_Q(X_j), \mathrm{Law}_P(X_j)\big) < \varepsilon, &\qquad \big|\mathcal{R}^{\mathrm{raw}}_{N,\mathcal{L}}(Q) - \mathcal{R}^{\mathrm{raw}}_{N,\mathcal{L}}(P)\big| < \varepsilon. \end{aligned} \tag{2}$$

Density in the topology generated by (1) and the marginals (Section 2) is the statement that this is always possible; if it is, calibrating in finite mixtures costs nothing the readout can see.

The difficulty, and the topology that resolves it. Static density of finite Gaussian mixtures is classical [2], and quantizing a Markov kernel yields weak convergence of finite-dimensional laws; (2) reduces to neither, because its requirements pull against one another. Exact martingality represents every cell of a finite approximation by its exponential barycentre, leaving atoms where the readout asks for derivatives of a conditional density; and a finite chain cannot meet a generic marginal exactly (Proposition 2.5). What has to be settled first is therefore not how to approximate but what approximation should mean, and the ready-made answers do not serve. Weak or Wasserstein closeness of path laws ignores the flow of information, so models close in those topologies can call for very different hedges; the adapted Wasserstein distance repairs that, makes hedging Lipschitz stable [3], and controls the conditional laws the rows d^ℓ_m depend on. It is the natural candidate but still not enough: finite Gaussian-mixture chains can converge to a Black–Scholes target adaptedly while their at-the-money smile curvature diverges. No Wasserstein distance on path laws, adapted or not, makes the readout continuous (Remark 2.4). The notion must therefore come from the readout itself, together with the marginals, which is what (2) asks for and what Section 2 constructs: martingality is kept exact, the marginals are allowed to move, and the readout is required to converge. Placing each demand where it can be met is the first contribution here, and to our knowledge this topology has not been isolated before.

Main results. Theorems 7.3 (SV) and 9.13 (LSV) give, for every target P in the respective class, one sequence Q_n of exact-martingale finite Gaussian-mixture chains whose errors in (2) tend to zero for every N simultaneously, with the velocity coefficients converging on $a^\ell_1 \neq 0$; in the topologies of Section 2,

$$\mathfrak{T}_{\mathrm{SV}} \subseteq \overline{\mathfrak{G}_{\mathrm{SV}}}^{\tau^{\mathrm{SV}}_{\mathrm{mov}}}, \qquad \mathfrak{T}_{\mathrm{LSV}} \subseteq \overline{\mathfrak{G}_{\mathrm{lift}}}^{\tau^{\mathrm{LSV}}_{\mathrm{mov}}}.$$

Both classes assume a compact latent state, Lipschitz kernels with exponential moments, a^ℓ_0 in a compact subset of $(0,\infty)$ and $\mathrm{Var}(X_1 - X_0) > 0$. The SV class adds the Gaussian factor of Assumption 3.1, supplied by a variance floor with correlation bounded away from ± 1 (Proposition 3.3), and contains capped Heston (Proposition 3.4); the LSV class adds one-step return densities with bounded, uniformly continuous derivatives, of which order N uses the first $\max\{N-1, 0\}$ (Remark 9.15). Theorem 8.3 makes the SV statement quantitative: with M components per node and a d-dimensional latent state the readout error is $O(M^{-1/(d+1)})$ up to logarithms, only that dimension entering the exponent, and the quantization exponent is optimal (Proposition 8.4). An LSV target supplies no factor, so the construction adds Gaussian innovations of variance $\beta_n \downarrow 0$. Adapted approximation forces innovations preceding a price-dependent kernel to vanish (Proposition 9.18), and when that dependence is visible the construction's adapted error is of exact order $\sqrt{\beta_n}$ (Proposition 9.19); a final-edge factor may stay (Proposition 9.20). Two results mark the limits of the class: finite mixtures have zero implied-variance wing slopes [33], so density fails for a target with a finite moment index once the wing slope is added to the

readout (Proposition 7.4); and with the marginals held fixed rather than approximated, the finite class caps the dynamics (Proposition 7.5). In a fixed-fitted-marginal finite-chain experiment on an SPX surface, the coupling range leaves the at-the-money forward-start volatility undetermined by 7.0 points, and matching the first three readout coordinates as well narrows this to 4.0 in the fixed-marginal experiment, with each constrained endpoint bracketed by an attained feasible-chain value and a dual outer bound differing by less than 0.01 volatility points; over fifty-five further monthly surfaces the three coordinates remove a median 39% of that range (Section 10.2).

Method. Both proofs quantize each transition by exponential barycentres $\log \mathbb{E}[e^L \mid \text{cell}]$ and couple target and approximant bicausally, which carries conditional laws along the sequence. Three obstacles follow. The martingale-preserving representative is not the optimal one: quantization rates are stated for L^2 centroids, which violate martingality, and martingale-preserving quantization is a problem in its own right [30], so the estimates are rebuilt around exponential barycentres (Lemmas 4.1 and 8.1). The labels move: the cells chosen at one date create the label set of the next, so the continuation jets must converge uniformly over sets that change along the sequence, and errors compound through the recursion (25). And the matched quantity is a jet of implied variance, recovered by an inversion whose denominator, vega, must stay away from zero along the sequence (Assumption 6.3). For an LSV target, density derivatives of order r converge once the quantization error ε_n vanishes faster than $\beta_n^{(r+2)/2}$, and a diagonal in r serves every order at once; this is sufficient rather than necessary, since the readout itself needs only $\beta_n^{-m/2}$ at order m (Corollary 9.17).

Scope and related work. The theorems are existence statements about risk-neutral capacity on a finite grid; fixed component budgets, calibration selection, instantaneous limits and physical-measure use are separate problems (Section 11). The readout meets two requirements that other gradings fail (Section 2.1): (A) order zero recovers the skew-stickiness ratio and higher orders refine it; (B) every coordinate is defined from the grid law, without a Brownian driver or perturbation parameter. Like [18], which confines transport to compact moneyness intervals and defers the wings to [33], the readout here is at-the-money by design. Model-side work computes such quantities for given models rather than asking which classes can match them: a representation formula for the skew-stickiness ratio [24], a second-order volatility-of-volatility expansion [12], and a quintic Ornstein–Uhlenbeck model tested against the skew-stickiness term structure [1]. Arbitrage-consistent frameworks for dynamic surfaces characterize admissible dynamics [34, 35, 16, 17], and [20] derives the motion of the surface inside local volatility.

Organization. Section 2 defines the readout and the topology it generates, and Section 3 the SV class; Sections 4–7 prove Theorem 7.3 and Section 9 proves Theorem 9.13. Section 8 adds a convergence rate and Section 10 the closed-form readout, a cubature bound, two numerical checks of the rates and a constrained band on an SPX surface; Section 11 discusses scope.

AI-use disclosure. The authors used Anthropic Claude Code and OpenAI Codex as interactive research and writing tools. They assisted with exploratory discussion, testing and refinement of ideas, literature and source organization, code development and verification, mathematical error checking, and editorial revision.

2 The jet readout and the topology it generates

Smile dynamics are compared in a finite-step form of the transport coordinates of Che and Das [18], computed from the law of a chain on the grid, and the topology is the one those coordinates generate. Static jets are read off conditional smiles by Black–Scholes inversion, dynamic coordinates are their finite-step projections on the spot increment, and velocity jets follow by triangular inversion.

2.1 The transport construction and what we use

Che and Das [18] write the action of spot moves on total implied variance $w(k,u,T)$, with k the log-moneyness and u the log forward, as a transport

$$\partial_u w(k,u,T) = v(k,u,T;w)\,\partial_k w(k,u,T), \tag{3}$$

in which $v \equiv 0$, $v \equiv 1$ and $v \equiv \beta$ are sticky delta, sticky strike and the skew-stickiness rule. With $\tilde{v}(k;u,T) := v(k,u,T;w(k,u,T))$ and $v_j := \partial_k^j \tilde{v}(0;u,T)$, differentiating (3) at the money gives a triangular identity whose projected form is (7). The family is graded, each v_n being a residual after lower-order transport.

We use (7) only to define coordinates. As an identity, (3) holds with $v = \partial_u w/\partial_k w$ wherever $\partial_k w \neq 0$ and degenerates where $\partial_k w = 0$; one trajectory fixes the composite jets v_j, not the dependence of v on w [18, Remark 5.2]; the jets are coordinates rather than invariants (Remark 2.2); and they see only the spot-spanned response. We therefore replace the pathwise ∂_u by the finite-step projection of Section 2.3 and state the theorems in the raw coordinates (a_m, d_m), which remain defined where the v_j are not. The theorems concern realizability: every projected jet profile generated by a target in the classes of Sections 3 and 9 is realized to arbitrary accuracy by exact-martingale finite chains. Attainability of an arbitrary empirical jet vector is not claimed, nor a characterization of the transport flows compatible with martingale dynamics. Other gradings fail requirement (A) or (B) of the introduction: Bergomi–Guyon functionals [12], Wiener-chaos order and martingale-expansion coefficients [23] need a forward-variance representation or a perturbation parameter, so they are not computable from the grid law of a finite mixture; marginal moments and cumulants are not nested descriptors of the at-the-money spot response; and the number of time points [27] is a complementary axis.

2.2 Conditional smiles and static jets

For any chain considered below, let ζ_j denote a sufficient continuation state: a state variable that determines the conditional law of normalized future returns. For $j < \ell$ and a state value ζ, define

$$C_{j,\ell}(k,\zeta) := \mathbb{E}\left[(e^{X_\ell - X_j} - e^k)^+ \mid \zeta_j = \zeta\right]. \tag{4}$$

Here and below the expectation denotes the Markov continuation started from ζ, fixing one continuation version on the state space. Let $B(k,w)$ denote the normalized Black–Scholes call with log-strike k and total variance w,

$$B(k,w) = \Phi(d_+) - e^k\Phi(d_-), \qquad d_\pm = -\frac{k}{\sqrt{w}} \pm \frac{\sqrt{w}}{2}.$$

The conditional implied total variance $w_{j,\ell}(k,\zeta)$ solves

$$B(k, w_{j,\ell}(k,\zeta)) = C_{j,\ell}(k,\zeta),$$

and its ATM jets are

$$a_m^{j,\ell}(\zeta) := \partial_k^m w_{j,\ell}(0,\zeta). \tag{5}$$

For the homogeneous SV class of Section 3, $\zeta_j = Y_j$, the latent state; for a local-SV target, $\zeta_j = (X_j, Y_j)$; for the lifted approximants of Section 9, $\zeta_j = (\Xi_j^n, \widehat{Y}_j^n)$, the finite proxy label on which their normalized continuation law depends. Quantities of the nth approximant carry a subscript n, as in $a_{m,n}^{j,\ell}$.

Lemma 2.1 (Finite-dimensional static jet map)**.** *Fix $r \geq 0$ and suppose $C \in C^r$ near zero. On any set where the ATM solution $w(0)$ stays in a compact subset of $(0,\infty)$, the vector $(a_0, \dots, a_r)$ is a continuous function of the call derivatives $(C(0), C'(0), \dots, C^{(r)}(0))$. The only implicit-function denominator is the Black–Scholes vega $B_w(0, w(0)) > 0$. If $C \in C^\infty$, then w is smooth near zero.*

Proof. Set $F(k,w) = B(k,w) - C(k)$. For $r = 0$, strict monotonicity of $B(0,\cdot)$ gives continuity of the scalar inverse at the money. For $r \geq 1$, since $F_w = B_w > 0$, the C^r implicit-function theorem gives a C^r local solution $w(k)$ (and a smooth one when C is smooth). The first derivatives are

$$a_1 = -\frac{F_k}{F_w},$$
$$a_2 = -\frac{F_{kk} + 2F_{kw}a_1 + F_{ww}a_1^2}{F_w}.$$

Repeated differentiation gives the same structure at every finite order: the numerator is a polynomial in lower jets and derivatives of F, divided by F_w. On the stated compact set these operations are continuous. □

2.3 Finite-step projection

Fix the observation edge $[t_0, t_1]$ and an option maturity t_ℓ with $\ell \geq 2$. At the known root define

$$a_m^\ell := a_m^{0,\ell}(\zeta_0), \qquad A_m^\ell := a_m^{1,\ell}(\zeta_1).$$

On the domain $\mathrm{Var}(X_1 - X_0) > 0$, the projected spot derivative of the mth smile jet is

$$d_m^\ell := \frac{\mathrm{Cov}(A_m^\ell - a_m^\ell, X_1 - X_0)}{\mathrm{Var}(X_1 - X_0)} = \frac{\mathrm{Cov}(A_m^\ell, X_1 - X_0)}{\mathrm{Var}(X_1 - X_0)}. \tag{6}$$

Since $a_m^{0,\ell}$ and $a_m^{1,\ell}$ refer to the same maturity one step apart, the deterministic part of the calendar decay drops out of the covariance, leaving a spot response. The denominator is positive: $\mathrm{Var}(X_1 - X_0) \geq \alpha_0$ under Assumption 3.1, and for LSV targets and their proxy approximants see (45) and Lemma 9.12.

In regression terms, d_m^ℓ is the population OLS slope of A_m^ℓ on $X_1 - X_0$, with residual vanishing exactly when A_m^ℓ is affine in $X_1 - X_0$. For a Gaussian-mixture edge with component label I, $\mathbb{P}(I = i) = p_i$ and $R := X_1 - X_0 \mid \{I = i\} \sim \mathcal{N}(m_i, s_i^2)$, one has $\mathrm{Var}(R) = \sum_i p_i[s_i^2 + (m_i - \bar{m})^2]$ with $\bar{m} = \sum_i p_i m_i$, and $\mathrm{Cov}(A, R) = \sum_i p_i (A_i - \bar{A})(m_i - \bar{m})$ if $A = A_i$ on $\{I = i\}$; if A varies within components, the law of total covariance adds $\sum_i p_i \,\mathrm{Cov}(A, R \mid I = i)$.

2.4 Triangular inversion and the pivot

The projected velocity jets are defined algebraically by

$$d_n^\ell = \sum_{j=0}^{n} \binom{n}{j} v_j^\ell a_{n-j+1}^\ell. \tag{7}$$

This is the pattern obtained by differentiating (3) at the money, with d_n^ℓ in place of the pathwise $\partial_u a_n$; identifying the two would require the instantaneous limit of Section 11.

If $a_1^\ell \neq 0$, the system is triangular:

$$v_n^\ell = \frac{d_n^\ell - \sum_{j=0}^{n-1} \binom{n}{j} v_j^\ell a_{n-j+1}^\ell}{a_1^\ell}. \tag{8}$$

More generally, if the hierarchy is compatible and $a_1^\ell = \cdots = a_{r-1}^\ell = 0 \neq a_r^\ell$, the rows impose $d_0^\ell = \cdots = d_{r-2}^\ell = 0$ and, for $n \geq 0$,

$$v_n^\ell = \frac{d_{r-1+n}^\ell - \sum_{j=0}^{n-1} \binom{r-1+n}{j} v_j^\ell a_{r+n-j}^\ell}{\binom{r-1+n}{n} a_r^\ell}, \tag{9}$$

so $v_0, \ldots, v_N$ need a_m for $m \leq r+N$ and d_m for $m \leq r-1+N$, and near that stratum (9) is a continuous r-pivoted extension wherever $a_r \neq 0$. Approximation may perturb the vanishing jets, so the theorems use raw coordinates globally and velocity coefficients on the regular chart $a_1 \neq 0$.

For finite $\mathcal{L} \subseteq \{2, \ldots, J\}$ and $N \in \mathbb{N}_0$, the raw readout $\mathcal{R}_{N,\mathcal{L}}^{\mathrm{raw}}(P)$ collects $a_m^\ell(P)$, $m \leq N+1$, and $d_m^\ell(P)$, $m \leq N$, over $\ell \in \mathcal{L}$, as in (1). It recovers $v_0^\ell, \ldots, v_N^\ell$ on the chart $a_1^\ell \neq 0$, where the velocity coefficients are continuous functions of finitely many raw coordinates, so adjoining them leaves the topology unchanged; raw convergence also gives convergence of any fixed r-pivoted extension at a target with $a_r \neq 0$.

Remark 2.2 (Transformation law: the velocity jets are chart coordinates)**.** The v_j are jets of a vector field and transform as such. They are invariant under a change of variance level $w \mapsto \lambda w$, which is why they compare across volatility levels. Under a linear change of moneyness $\tilde{k} = ck$, $v_j \mapsto c^{1-j} v_j$, so v_0 scales like c, v_1 is invariant and v_2 scales like c^{-1}; under $\tilde{k} = \varphi(k)$ with $\varphi(0) = 0$, $v_0 \mapsto \varphi'(0) v_0$ and $v_1 \mapsto v_1 + (\varphi''(0)/\varphi'(0)) v_0$. They are therefore coordinates relative to a parametrization: under the linear subgroup the invariants are the monomials of total weight zero, such as v_1 and $v_0 v_2$, and quoting the smile in $k/(\sigma\sqrt{T})$ divides v_0 by $\sigma\sqrt{T}$. We fix $k = \log(K/F)$; the raw coordinates (a_n, d_n) carry no convention beyond this choice.

2.5 The topology generated by the readout

Let $\mathfrak{M}$ be a class of rooted marked martingale chains. For maps $\mathcal{R}_i : \mathfrak{M} \to Y_i$, $\sigma(\mathcal{R})$ is the initial topology, the coarsest making every $\mathcal{R}_i$ continuous; a sequence converges in it exactly when every coordinate does. The construction is textbook, and the content is the choice of maps. Generated by the readout, the topology adapts to the quantities that describe the dynamics: two chains are close when their at-the-money jets and projected spot responses are close, whatever kernels produce them. Adjoining a coordinate refines the topology and can destroy density, as the wing slope does (Proposition 7.4). $\mathcal{P}_p(F)$ denotes the laws on a metric space F with finite pth moment, with the distance W_p.

Lemma 2.3 (Image density)**.** *Let $Y = \prod_i Y_i$ carry the product topology and let $\mathcal{R} = (\mathcal{R}_i) : \mathfrak{M} \to Y$. For $D \subseteq \mathfrak{M}$,*

$$D \text{ is } \sigma(\mathcal{R})\text{-dense in } \mathfrak{M} \quad \Longleftrightarrow \quad \mathcal{R}(D) \text{ is dense in } \mathcal{R}(\mathfrak{M}).$$

Proof. The sets $\mathcal{R}^{-1}(V)$, with V product-open, form a base of the initial topology. Such a set is nonempty exactly when V meets $\mathcal{R}(\mathfrak{M})$, and it meets D exactly when V meets $\mathcal{R}(D)$. This is precisely image density in the subspace $\mathcal{R}(\mathfrak{M})$. □

For an ambient class $\mathfrak{M}$ on which the raw readout $\mathcal{R}^{\mathrm{raw}}$ is defined, let $\mathsf{m}_j^X(P) := \mathrm{Law}_P(X_j) \in (\mathcal{P}_p(\mathbb{R}), W_p)$ and $\mathsf{m}_j^S(P) := \mathrm{Law}_P(e^{X_j}) \in (\mathcal{P}_1(\mathbb{R}_+), W_1)$, and, for finite N and $\mathcal{L} \subseteq \{2, \dots, J\}$, define

$$\tau_{\mathrm{mov}}^{N,\mathcal{L}}(\mathfrak{M}) := \sigma\big(\mathcal{R}_{N,\mathcal{L}}^{\mathrm{raw}}, \mathsf{m}_j^X, \mathsf{m}_j^S : j \le J\big), \qquad \tau_{\mathrm{mov}}(\mathfrak{M}) := \sigma\big(\mathcal{R}^{\mathrm{raw}}, \mathsf{m}_j^X, \mathsf{m}_j^S : j \le J\big). \tag{10}$$

The full moving-marginal topology τ_{mov} is generated jointly by all $\tau_{\mathrm{mov}}^{N,\mathcal{L}}$, so a basic neighbourhood constrains finitely many coordinates; target and approximant may have different mark spaces, since every generator is a model-level map. The topology $\tau_{\mathrm{mov}}^{N,\mathcal{L}}$ is generated by the pseudometric

$$\begin{aligned} D_{N,\mathcal{L}}(P,Q) := {} & \big\|\mathcal{R}_{N,\mathcal{L}}^{\mathrm{raw}}(P) - \mathcal{R}_{N,\mathcal{L}}^{\mathrm{raw}}(Q)\big\| \\ & + \sum_{j=0}^{J} \Big[W_p\big(\mathsf{m}_j^X(P), \mathsf{m}_j^X(Q)\big) + W_1\big(\mathsf{m}_j^S(P), \mathsf{m}_j^S(Q)\big) \Big], \end{aligned}$$

so P lies in the $\tau_{\mathrm{mov}}^{N,\mathcal{L}}$-closure of a class $\mathfrak{A} \subseteq \mathfrak{M}$ exactly when its finite-order calibration error $\inf_{Q \in \mathfrak{A}} D_{N,\mathcal{L}}(P,Q)$ vanishes; exact martingality constrains the candidates rather than entering the error. For an approximating class $\mathfrak{A} \subseteq \mathfrak{M}$ and a target $P \in \mathfrak{M}$, write

$$k_*(P; \mathfrak{A}) := \sup\big\{ N \in \mathbb{N}_0 : \ P \in \overline{\mathfrak{A}}^{\tau_{\mathrm{mov}}^{N,\mathcal{L}}(\mathfrak{M})} \text{ for every finite } \mathcal{L} \big\}.$$

Thus $k_* = \infty$ means zero calibration error at every finite order and maturity set; both constructions below achieve it with a single sequence.

Density is proved along one constructed sequence, by showing that every generator converges and applying Lemma 2.3; adapted Wasserstein convergence (Section 5) is a step of that proof, not a finer topology.

Remark 2.4 (The readout is not weakly continuous)**.** No Wasserstein distance on path laws, adapted or not, makes the readout continuous on the classes below. Take a two-step Black–Scholes target, $X_{j+1} - X_j \sim \gamma_{v_j}$ independent with $v_0, v_1 > 0$, and let $\alpha_n, h_n \downarrow 0$ with $\alpha_n = o(h_n^4)$. On each edge write $\gamma_{v_j} = \gamma_{v_j - \alpha_n} * \gamma_{\alpha_n}$ and keep the second factor, but quantize the first by exponential barycentres on cells of width of order h_n, placing one cell on each edge so that the two barycentres sum to α_n; a barycentre moves continuously as its cell slides. The result is a finite chain as in Definition 5.1 with edge variances α_n, and coupling the quantized factors optimally and the Gaussian factors identically gives $\mathcal{AW}_p \to 0$. The two placed cells carry masses of order h_n, and the component through both is $\mathcal{N}(0, 2\alpha_n)$, centred at the money, so the density f of $X_2 - X_0$ is at least of order $h_n^2/\sqrt{\alpha_n} \to \infty$ there. With C the call function of $X_2 - X_0$, $C''(0) - C'(0) = f(0)$ while $C(0)$ and $C'(0)$ converge, so the expression for a_2 in the proof of Lemma 2.1 diverges: the root curvature jet at t_2 does not converge. What restores continuity in Section 6 is a Gaussian factor whose variance is held fixed along the sequence. Section 9 lets that variance vanish too, but only with the quantization error driven to zero faster than a power of it; here the variance vanishes faster than the quantization error.

2.6 Moving marginals and exact martingality

For the discounted asset S and $X = \log S$ (so the log forward is $u = X$), martingality is the exponential barycentre identity

$$\int e^{x'} K((x,y), dx', dy') = e^{x}, \tag{11}$$

for every state (x, y) reached by the chain.

Proposition 2.5 (Marginal rigidity)**.** *Let Q start from a known root x_0 and carry a label q_h taking finitely many values at each date, with transitions*

$$K_h\big((x,q), dx', dq'\big) = \sum_r p_r(q)\, \mathcal{N}\big(x + \lambda_r(q) - \tfrac{1}{2}\alpha_h,\ \alpha_h\big)(dx')\, \delta_{q_r^+(q)}(dq'), \qquad \alpha_h > 0,$$

so that the component index and the successor label depend on the current label alone. Then for every $j \geq 1$,

$$\mathrm{Law}_Q(X_j) = \nu_j * \mathcal{N}(0, A_j), \qquad A_j := \sum_{h<j} \alpha_h, \tag{12}$$

for a finitely supported ν_j, which is determined by $\mathrm{Law}_Q(X_j)$ and A_j. In particular a law on $\mathbb{R}$ is the date-j marginal of such a chain only if it is the convolution of a finitely supported law with a centred Gaussian.

Proof. The label chain evolves independently of the Gaussian increments. Conditionally on the finitely many label and component paths up to date j, X_j is x_0 plus the chosen locations, less $A_j/2$, plus j independent centred Gaussians, hence Gaussian with variance A_j; mixing over the finitely many paths gives (12). The Fourier transform of $\mathcal{N}(0, A_j)$ does not vanish, so $\widehat{\nu_j}$, and with it ν_j, is determined. □

Both finite classes below have this form, with α_h the factor variances (Definition 5.1) or $\beta_{h,n}$ (Definition 9.6, Lemma 9.7). Under Assumption 3.1, $\mathrm{Law}_P(X_j) = \mathrm{Law}(x_0 + S_j) * \gamma_{A_j^P}$ with $S_j := \sum_{h<j} L_h$ and A_j^P the sum of the target's factor variances, and matching (12) is impossible if $A_j < A_j^P$, forces S_j to be finitely supported if $A_j = A_j^P$, and forces it to be a finite Gaussian location mixture with a common variance if $A_j > A_j^P$. For generic targets the marginals can therefore be approximated — in W_p by Theorems 7.3 and 9.13, and in total variation, since the densities converge uniformly (Lemmas 6.2 and 9.10) — but not attained. This is why the formulation lets the marginals move; holding them fixed would in addition cap the dynamics (Proposition 7.5). The approximants below satisfy (11) exactly, so their spot laws are in convex order. A nondegenerate time-zero law μ_0 with $\int e^{q|x|}\mu_0(dx) < \infty$ can be included by quantizing $X_0 - \log \int e^x \mu_0(dx)$ with Lemma 4.1, shifting back and adding an independent $\mathcal{N}(-h_n^2/2, h_n^2)$, $h_n \downarrow 0$, which keeps $\mathbb{E}[e^{X_0}]$ exact.

3 Model class and Gaussian factorization

A variance floor with imperfect correlation leaves a Brownian direction unspanned by the volatility, so each transition splits into a residual return carrying the whole exponential barycentre and an independent mean-one Gaussian factor. This restriction on the target is checkable from the model's coefficients (Proposition 3.3); the factor keeps martingality exact under quantization,

supplies the smoothing of Section 6 and yields the rate of Section 8. Section 9 drops it and pays for it.

Fix dates $0 = t_0 < t_1 < \cdots < t_J$ containing the observation step and all maturities, a compact convex $E \subset \mathbb{R}^d$, and states $Z_j = (X_j, Y_j) \in \mathbb{R} \times E$ with known root (x_0, y_0), where Y is the latent volatility state. With $\mathcal{F}_j = \sigma(Z_h : 0 \le h \le j)$, homogeneity means that for kernels K_j^{ret},

$$\mathrm{Law}((X_{j+1} - X_j, Y_{j+1}) \mid \mathcal{F}_j) = K_j^{\mathrm{ret}}(Y_j) \qquad \text{a.s.}$$

For $\alpha > 0$ put $\gamma_\alpha := \mathcal{N}(-\alpha/2, \alpha)$, so that $\mathbb{E}[e^G] = 1$ for $G \sim \gamma_\alpha$.

Assumption 3.1 (Independent mean-one Gaussian factor)**.** For every edge $j < J$, there are $\alpha_j > 0$ and a kernel

$$H_j : E \longrightarrow \mathcal{P}(\mathbb{R} \times E), \qquad y \longmapsto H_j(y; d\ell, dy'),$$

such that, jointly over the whole grid,

$$X_{j+1} = X_j + L_j + G_j, \qquad \mathrm{Law}((L_j, Y_{j+1}) \mid \mathcal{F}_j) = H_j(Y_j), \qquad G_j \sim \gamma_{\alpha_j}, \tag{13}$$

where the G_j are mutually independent and

$$(G_0, \ldots, G_{J-1}) \;\perp\!\!\!\perp\; \big((L_h, Y_{h+1})\big)_{h<J}.$$

Moreover,

$$\int e^{\ell} H_j(y; d\ell, dy') = 1 \qquad \text{for every } y \in E. \tag{14}$$

By (14) and $\mathbb{E}[e^{G_j}] = 1$, (11) holds exactly. Fix $p > 2$ and the metric $d((\ell, y), (\bar\ell, \bar y)) := |\ell - \bar\ell| + |y - \bar y|$ on $\mathbb{R} \times E$.

Assumption 3.2 (Kernel regularity and moments)**.** There are $q > p$ and $C, L < \infty$ such that for every j and $y, \bar y \in E$,

$$W_p(H_j(y), H_j(\bar y)) \le L|y - \bar y|, \tag{15}$$

$$\int e^{q|\ell|} H_j(y; d\ell, dy') \le C. \tag{16}$$

Proposition 3.3 (Stochastic volatility supplies the factor)**.** *Suppose*

$$dX_t = -\tfrac{1}{2}\sigma(t, Y_t)^2 dt + \sigma(t, Y_t)\left(\rho(t, Y_t)^\top dW_t + \sqrt{1 - \|\rho(t, Y_t)\|^2}\, dB_t\right), \tag{17}$$

$$dY_t = b(t, Y_t)dt + \Gamma(t, Y_t)dW_t, \tag{18}$$

where B is independent of W, $S = e^X$ is a true martingale from every initial state (t, x, y) under consideration, Y is a well-posed strong Markov solution adapted to the augmented filtration of W, and

$$\sigma \ge \underline{\sigma} > 0, \qquad \|\rho\| \le \bar\rho < 1.$$

Then, after passing to an extension of the probability space carrying the auxiliary normals constructed in the proof, the chain satisfies Assumption 3.1 with, for any fixed $\theta \in (0, 1)$,

$$\alpha_j = \theta\, \underline{\sigma}^2 (1 - \bar\rho^2)(t_{j+1} - t_j). \tag{19}$$

Proof. Let $\mathcal{G} = \sigma(W_s : 0 \le s \le t_J)$ (augmented by the initial state), and condition first on $\mathcal{G}$. On $[t_j, t_{j+1}]$ the log return is conditionally Gaussian,

$$X_{j+1} - X_j \mid \mathcal{G} \sim \mathcal{N}(M_j, Q_j),$$

where

$$M_j = -\frac{1}{2}\int_{t_j}^{t_{j+1}} \sigma_s^2 ds + \int_{t_j}^{t_{j+1}} \sigma_s \rho_s^\top dW_s,$$
$$Q_j = \int_{t_j}^{t_{j+1}} \sigma_s^2(1 - \|\rho_s\|^2) ds \ge \underline{\sigma}^2(1 - \bar{\rho}^2)(t_{j+1} - t_j) > \alpha_j.$$

On an extension of the probability space, do this simultaneously on every disjoint grid edge using mutually independent pairs of standard normals $(Z_{0,j}, Z_{1,j})$, independent of the W-driven path, and set

$$G_j = -\frac{\alpha_j}{2} + \sqrt{\alpha_j} Z_{0,j}, \qquad L_j = M_j + \frac{\alpha_j}{2} + \sqrt{Q_j - \alpha_j} Z_{1,j}.$$

Conditionally on $\mathcal{G}$, $G_j + L_j$ has law $\mathcal{N}(M_j, Q_j)$, independently across edges, as do the original increments, whose B-parts live on disjoint edges; so $X_{j+1} - X_j := G_j + L_j$ leaves the law of the grid chain unchanged. The vector $(Z_{0,j})_{j<J}$ is independent of $\mathcal{G}$ and $(Z_{1,j})_{j<J}$, which gives the path-level independence in Assumption 3.1, and (L_j, Y_{j+1}) is a function of Y_j, the W-increments on $[t_j, t_{j+1}]$ and $Z_{1,j}$, all but Y_j independent of $\mathcal{F}_j$, so its law given $\mathcal{F}_j$ is a kernel $H_j(Y_j)$. Finally, homogeneity and the asset-martingale property give

$$1 = \mathbb{E}[e^{X_{j+1} - X_j} \mid Y_j] = \mathbb{E}[e^{L_j} \mid Y_j]\mathbb{E}[e^{G_j}],$$

and $\mathbb{E}[e^{G_j}] = 1$, proving (14). □

Proposition 3.4 (A floored and capped square-root model lies in the class)**.** *Fix $0 < \underline{v} < \overline{v} < \infty$, put $E = [\underline{v}, \overline{v}]$, and let $b, \Gamma : \mathbb{R} \to \mathbb{R}$ be bounded and Lipschitz with*

$$\Gamma(\underline{v}) = \Gamma(\overline{v}) = 0, \qquad \Gamma > 0 \text{ on } (\underline{v}, \overline{v}), \qquad b(\underline{v}) > 0 > b(\overline{v}).$$

Let $|\rho| < 1$, let B be independent of W, let $V_0 \in E$, and consider

$$dX_t = -\tfrac{1}{2} V_t \, dt + \sqrt{V_t}\left(\rho \, dW_t + \sqrt{1 - \rho^2} \, dB_t\right), \qquad dV_t = b(V_t) \, dt + \Gamma(V_t) \, dW_t. \tag{20}$$

Then V takes values in E, and the grid skeleton of (20) satisfies Assumptions 3.1, 3.2 and 6.3; hence it lies in $\mathfrak{T}_{\mathrm{SV}}$ and Theorem 7.3 applies to it. With b and Γ equal to $\kappa(m - v)$ and $\xi\sqrt{v}$ away from the endpoints this is floored and capped Heston; if $\underline{v} < m < \overline{v}$, only Γ needs modifying.

Proof. Replacing $\sqrt{v}$ by $\sqrt{v \vee \underline{v}}$, which agrees with it on E, makes the system globally Lipschitz, so it has a unique strong solution. The coefficient Γ and the drift $b - b(\underline{v}) \le b$ both vanish at $\underline{v}$, so the constant $\underline{v}$ solves the equation with that drift, and comparison for one-dimensional equations with a common Lipschitz diffusion coefficient [31, Section 5.2.C] gives $V_t \ge \underline{v}$ from every start in E; the drift $b - b(\overline{v}) \ge b$ gives $V_t \le \overline{v}$ in the same way. Hence E is invariant, endpoints included.

Assumption 3.1. On E one has $\sigma = \sqrt{V} \in [\sqrt{\underline{v}}, \sqrt{\overline{v}}]$, so the volatility is bounded and floored, and $\|\rho\| = |\rho| < 1$. Boundedness of σ gives Novikov's condition, so e^X is a true martingale from every state, and V is a well-posed strong Markov solution adapted to the augmented filtration

of W. Proposition 3.3 therefore applies and yields (13) with $\alpha_j = \theta\,\underline{v}(1-\rho^2)(t_{j+1}-t_j)$ for any $\theta \in (0,1)$.

Assumption 3.2. In the representation of Proposition 3.3, couple the copies started from y and $\bar{y}$ through the same W and the same auxiliary normals. Then $L_j - \bar{L}_j = (M_j - \bar{M}_j) + (\sqrt{Q_j - \alpha_j} - \sqrt{\bar{Q}_j - \alpha_j})Z_{1,j}$, the square root is Lipschitz because $Q_j - \alpha_j \geq (1-\theta)\underline{v}(1-\rho^2)(t_{j+1}-t_j)$, and Gronwall with the Burkholder–Davis–Gundy inequality bounds $V - \bar{V}$, hence $M_j - \bar{M}_j$, $Q_j - \bar{Q}_j$ and the marks, by a multiple of $|y - \bar{y}|$ in L^p; this is (15). For (16), condition on the path of W: the residual return is Gaussian with variance at most $\overline{v}(t_{j+1}-t_j)$ and mean bounded by $\frac{1}{2}\overline{v}(t_{j+1}-t_j)$ plus a stochastic integral of quadratic variation at most $\overline{v}\rho^2(t_{j+1}-t_j)$. Both terms are sub-Gaussian with parameters depending only on $\overline{v}$ and the mesh, so every exponential moment is finite uniformly in the starting mark; in particular (16) holds for any q.

Assumption 6.3. The volatility satisfies $\sqrt{\underline{v}} \leq \sigma \leq \sqrt{\overline{v}}$ pathwise, and a European call is a convex payoff, so by the comparison theorem for misspecified volatility [21] the conditional call price lies between the Black–Scholes prices at those two volatilities. Monotonicity of $B(0,\cdot)$ then places the ATM implied total variance in $[\underline{v}(t_\ell - t_j), \overline{v}(t_\ell - t_j)]$, a compact subset of $(0,\infty)$ independent of the mark. □

4 Log-barycentric quantization

Quantization theory [25] gives rates for the unconstrained problem; the point here is that representing cells by exponential barycentres preserves martingality exactly at every level.

Lemma 4.1 (Exact exponential-barycentre quantization)**.** *Let $H \in \mathcal{P}(\mathbb{R}\times E)$ have unit exponential mean, $\int e^{\ell} H(d\ell, dy) = 1$, and a finite exponential moment $\int e^{q|\ell|} H(d\ell, dy) < \infty$ for some $q > p$. Then there are finitely supported laws $H^n = \sum_{r=1}^{M_n} p_r^n \delta_{(\lambda_r^n, y_r^n)}$ such that*

$$\sum_r p_r^n e^{\lambda_r^n} = 1, \qquad W_p(H^n, H) \longrightarrow 0, \qquad \sup_n \sum_r p_r^n e^{q|\lambda_r^n|} < \infty.$$

Proof. Let $(L, Y) \sim H$. Since $\mathbb{R} \times E$ is standard Borel, choose increasing finite sigma-fields $\mathcal{A}_n$ whose union generates $\sigma(L, Y)$. Set

$$U_n = \mathbb{E}[e^L \mid \mathcal{A}_n], \qquad L_n = \log U_n, \qquad Y_n = \mathbb{E}[Y \mid \mathcal{A}_n].$$

The convexity of E gives $Y_n \in E$, and finiteness of $\mathcal{A}_n$ makes (L_n, Y_n) finitely valued. Moreover,

$$\mathbb{E}[e^{L_n}] = \mathbb{E}[U_n] = \mathbb{E}[e^L] = 1.$$

Martingale convergence yields $U_n \to e^L$ and $Y_n \to Y$, hence $L_n \to L$, almost surely. Conditional Jensen for the convex maps $u \mapsto u^q$ and $u \mapsto u^{-q}$ gives

$$e^{qL_n} \leq \mathbb{E}[e^{qL} \mid \mathcal{A}_n], \qquad e^{-qL_n} = U_n^{-q} \leq \mathbb{E}[e^{-qL} \mid \mathcal{A}_n].$$

Since $e^{q|x|} \leq e^{qx} + e^{-qx}$, integration gives the asserted uniform two-sided exponential bound. Therefore $(|L_n|^p)_n$ is uniformly integrable. Since E is compact, $Y_n \to Y$ in every L^p. Thus $(L_n, Y_n) \to (L, Y)$ in L^p, which implies $W_p(\mathrm{Law}(L_n, Y_n), H) \to 0$. Taking $H^n = \mathrm{Law}(L_n, Y_n)$ proves the result. □

The partitions must generate $\sigma(L, Y)$, not $\sigma(Y)$: if the cells C_r are unions of full y-fibres, per-fibre normalization gives every $\lambda_r = \log \mathbb{E}[e^L \mid (L, Y) \in C_r]$ the value 0, leaving no spot–mark covariance.

5 Finite marked trees and adapted convergence

Quantizing every node gives a finite marked tree of exact martingale kernels, and a recursive bicausal coupling gives adapted convergence.

For path laws P, Q on $(\mathbb{R} \times E)^{J+1}$ define

$$\mathcal{AW}_p(P,Q)^p := \inf_{\pi \in \Pi_{\mathrm{bc}}(P,Q)} \mathbb{E}_\pi \left[\sum_{j=0}^{J} \big(|X_j - \widehat{X}_j| + |Y_j - \widehat{Y}_j| \big)^p \right], \tag{21}$$

where Π_{bc} denotes the bicausal couplings [3, 5]. For Markov chains on the grid, recursively coupling next-step kernels as functions of the two current states gives such a coupling.

Definition 5.1 (Finite marked GM chain)**.** A finite marked GM chain has finite continuation-label sets Q_j^n and edgewise constants $\alpha_j > 0$, all part of the chain's data. The label space may be the target latent space, as in the homogeneous construction, or may include a finite proxy log-price, as in Section 9. From (x, q), $q \in Q_j^n$, its transition is

$$K_j^n((x,q), dx', dq') = \sum_{r=1}^{M_{j,q}^n} p_{j,q,r}^n \mathcal{N}\left(x + \lambda_{j,q,r}^n - \frac{\alpha_j}{2}, \alpha_j \right)(dx') \delta_{q_{j,q,r}^{n,+}}(dq'), \tag{22}$$

where

$$\sum_r p_{j,q,r}^n e^{\lambda_{j,q,r}^n} = 1. \tag{23}$$

Every kernel (22) is an exact asset-martingale kernel:

$$\int e^{x'} K_j^n((x,q), dx', dq') = e^x \sum_r p_{j,q,r}^n e^{\lambda_{j,q,r}^n} \mathbb{E}[e^{G_j}] = e^x.$$

Theorem 5.2 (Finite-tree approximation)**.** *Under Assumptions 3.1 and 3.2, there are finite marked GM chains P^n starting from (x_0, y_0) such that*

1. *every transition is an exact asset-martingale transition;*
2. *$\mathcal{AW}_p(P^n, P) \to 0$;*
3. *for every $j \geq 1$, $\mathrm{Law}(X_j^n)$ is a finite Gaussian mixture, and $\mathrm{Law}(X_j^n, Y_j^n)$ converges to $\mathrm{Law}(X_j, Y_j)$ in W_p.*

Proof. Fix $\varepsilon_n \downarrow 0$ and put $E_0^n = \{y_0\}$. Suppose the finite set E_j^n has been constructed. For every $y \in E_j^n$, apply Lemma 4.1 to $H_j(y)$ and choose $H_j^n(y) = \sum_r p_{j,y,r}^n \delta_{(\lambda_{j,y,r}^n, y_{j,y,r}^{n,+})}$ with $W_p(H_j^n(y), H_j(y)) \leq \varepsilon_n$ and exact exponential barycentre. The conditional-Jensen estimates in that lemma and (16) give the same uniform q-exponential bound over all selected nodes and all n. Let E_{j+1}^n be the union of the finitely many successor marks. This recursively defines (22); exact martingality was checked above.

We construct a bicausal coupling recursively. Given current states (x, y) and $(\widehat{x}, \widehat{y})$, first note that, by (15),

$$W_p(H_j(y), H_j^n(\widehat{y})) \leq L|y - \widehat{y}| + \varepsilon_n. \tag{24}$$

Choose a Borel measurable ε_n-optimal coupling kernel for this pair of laws; such measurable selections exist for Wasserstein costs on Polish spaces [36, Corollary 5.22]. Its conditional L^p cost is bounded by the right side of (24) plus ε_n, hence by $L|y-\widehat{y}|+2\varepsilon_n$. Use the same independent $G_j \sim \gamma_{\alpha_j}$ in the two transitions. If the coupled residual variables are (L, Y') and $(\widehat{L}, \widehat{Y}')$, then

$$X' - \widehat{X}' = X - \widehat{X} + L - \widehat{L}.$$

Writing $e_j := \big(\mathbb{E}[(|X_j - \widehat{X}_j| + |Y_j - \widehat{Y}_j|)^p]\big)^{1/p}$ for the date-j coupling cost in the metric of (21), the price gap carries over and the coupled residual–mark cost adds at most $L\|Y_j - \widehat{Y}_j\|_{L^p} + 2\varepsilon_n$, so Minkowski's inequality gives

$$e_{j+1} \le (1+L)e_j + 2\varepsilon_n, \qquad e_0 = 0. \tag{25}$$

Iterating over the finite grid gives an $O(\varepsilon_n)$ bound at every date. The selected coupling kernel at each step depends only on the pair of current states, and the common G_j is drawn independently of the coupled residuals, so each coordinate receives its own transition kernel; the coupling is therefore bicausal. This proves $\mathcal{AW}_p(P^n, P) \to 0$ and the marginal W_p convergence.

It remains to identify the approximating marginal. Conditional on a finite residual branch path π up to date j,

$$X_j^n \sim \mathcal{N}\left(x_0 + \sum_{h<j} \lambda_{\pi,h} - \frac{1}{2}A_j,\ A_j\right), \qquad A_j = \sum_{h<j} \alpha_h.$$

There are only finitely many branch paths, so $\mathrm{Law}(X_j^n)$ is a finite Gaussian mixture. □

Since $e^{X_j^n} \Rightarrow e^{X_j}$ with common mean e^{x_0}, the finite-lognormal-mixture spot laws converge in W_1, and so do the call prices C_j^n, C_j on $S_{t_j}^n = e^{X_j^n}$ and $S_{t_j} = e^{X_j}$, uniformly in strike:

$$\sup_{K\ge 0} |C_j^n(K) - C_j(K)| \le W_1\big(\mathrm{Law}(S_{t_j}^n), \mathrm{Law}(S_{t_j})\big). \tag{26}$$

6 Continuation stability and Gaussian smoothing

The recursive coupling controls future conditional laws, and the Gaussian factor on the last edge before each maturity upgrades this to uniform convergence of densities and their derivatives.

Lemma 6.1 (Future conditional laws)**.** *Fix $j < \ell$. Let $y_n \in E_j^n$ and suppose $y_n \to y \in E$. Couple the target chain started from y at t_j and the approximating continuation started from y_n, with the same normalized current log price, by the construction in Theorem 5.2. Then their future path laws through t_ℓ converge in $\mathcal{AW}_p$. Moreover, with*

$$Z_{j,\ell} := \sum_{h=j}^{\ell-2}(L_h + G_h) + L_{\ell-1}, \qquad X_\ell - X_j = Z_{j,\ell} + G_{\ell-1},$$

and the analogous $Z_{j,\ell}^n$, one has

$$W_p\big(\mathrm{Law}(Z_{j,\ell}^n), \mathrm{Law}(Z_{j,\ell})\big) \le C_{j,\ell}(|y_n - y| + \varepsilon_n).$$

Proof. The initial state error is $|y_n - y|$. At each later edge, the measurable coupling estimate following (24) applies with the same constants. The finite-horizon induction used in Theorem 5.2 therefore bounds the future path error by $C_{j,\ell}(|y_n - y| + \varepsilon_n)$, which tends to zero. The coupling is again bicausal. On its extended space every G_h is coupled identically and every $L_h - \widehat{L}_h$ is controlled by the corresponding edge estimate. Minkowski's inequality applied directly to the displayed residual sums gives the stated W_p bound. □

Lemma 6.2 (Common-Gaussian smoothing)**.** *Let $\eta_n, \eta \in \mathcal{P}_1(\mathbb{R})$ with $W_1(\eta_n, \eta) \to 0$, and fix $\alpha > 0$. If $f_n = \eta_n * \varphi_\alpha$ and $f = \eta * \varphi_\alpha$, where φ_α is the density of γ_α, then for every $r \geq 0$,*

$$\|f_n^{(r)} - f^{(r)}\|_\infty \leq \|\varphi_\alpha^{(r+1)}\|_\infty W_1(\eta_n, \eta). \tag{27}$$

Proof. Take a coupling (U_n, U) attaining, or arbitrarily approaching, W_1. For every x, the mean value theorem gives

$$|\varphi_\alpha^{(r)}(x - U_n) - \varphi_\alpha^{(r)}(x - U)| \leq \|\varphi_\alpha^{(r+1)}\|_\infty |U_n - U|.$$

Take expectations and then the supremum over x. □

Since $X_\ell - X_j = Z_{j,\ell} + G_{\ell-1}$ with $G_{\ell-1}$ independent and additive at maturity, the conditional return law is the residual law convolved with $\gamma_{\alpha_{\ell-1}}$, and Lemma 6.2 applies.

Assumption 6.3 (Implied-variance nondegeneracy)**.** For every date pair under consideration, the target ATM implied total variances stay in one compact subset of $(0, \infty)$ uniformly over $y \in E$.

Lemma 6.4 (Static jets from converging return densities)**.** *Fix $p > 2$, $p' = p/(p-1)$ and an order $r \geq 0$. For each n let $\mathcal{Z}_n$ be a set of labels and, for $z \in \mathcal{Z}_n$, let $R_n(z)$ and $R(z)$ be coupled log returns with unit exponential mean. Suppose that*

1. *$\sup_{z \in \mathcal{Z}_n} \|R_n(z) - R(z)\|_{L^p} \to 0$;*
2. *$\sup_n \sup_{z \in \mathcal{Z}_n} \big(\mathbb{E}[e^{p' R_n(z)}] + \mathbb{E}[e^{p' R(z)}]\big) < \infty$;*
3. *$R_n(z)$ and $R(z)$ have densities $f_n(\cdot\,; z)$ and $f(\cdot\,; z)$ of class C^r such that, for every $i \leq r$, $\sup_n \sup_{z \in \mathcal{Z}_n} \|\partial^i f(\cdot\,; z)\|_\infty < \infty$ and $\sup_{z \in \mathcal{Z}_n} \|\partial^i f_n(\cdot\,; z) - \partial^i f(\cdot\,; z)\|_\infty \to 0$;*
4. *the ATM implied total variances of the $R(z)$ lie in one compact subset of $(0, \infty)$.*

Then, writing $a_m(R)$ for the mth ATM jet of the implied total variance of $k \mapsto \mathbb{E}[(e^R - e^k)^+]$,

$$\sup_{z \in \mathcal{Z}_n} \big|a_m(R_n(z)) - a_m(R(z))\big| \longrightarrow 0 \qquad (m \leq r + 2).$$

Proof. Write C_n, C for the two call functions. By $|e^u - e^v| \leq |u - v|(e^u + e^v)$ and Hölder's inequality, (1) and (2) give $\sup_z |C_n(0) - C(0)| \leq \sup_z \mathbb{E}|e^{R_n} - e^R| \to 0$. Next, $C'(0) = -\mathbb{P}(R > 0)$, and for every $\delta > 0$, $|\mathbb{P}(R_n > 0) - \mathbb{P}(R > 0)| \leq \delta^{-1}\mathbb{E}|R_n - R| + 2\delta \sup f$; letting $n \to \infty$ and then $\delta \downarrow 0$ gives uniform convergence of $C_n'(0)$. Since $C''(k) - C'(k) = e^k f(k)$, each $C^{(i)}(0)$ with $2 \leq i \leq r+2$ is a fixed linear combination of $C'(0)$ and the values $f^{(h)}(0)$, $h \leq i - 2$, so (3) gives uniform convergence of all of them, and the vectors stay in a bounded set. By (4) and the convergence of $C_n(0)$, the approximating ATM variances lie in a slightly enlarged compact subset of $(0, \infty)$ for large n, on which the map of Lemma 2.1 is uniformly continuous on bounded sets. □

Proposition 6.5 (Convergence of conditional static jets)**.** *Under Assumptions 3.1, 3.2, and 6.3, for every $j < \ell$, every finite m,*

$$\max_{z \in E_j^n} \left| a_{m,n}^{j,\ell}(z) - a_m^{j,\ell}(z) \right| \longrightarrow 0.$$

The target map $y \mapsto a_m^{j,\ell}(y)$ is continuous on E. Consequently, if $y_n \in E_j^n$ and $y_n \to y$, then $a_{m,n}^{j,\ell}(y_n) \to a_m^{j,\ell}(y)$.

Proof. Apply Lemma 6.4 with $\mathcal{Z}_n = E_j^n$, with $R_n(z), R(z)$ the coupled normalized returns $X_\ell - X_j$ started from z, and with every r; all conditional forwards equal one exactly. Hypothesis (1) is Lemma 6.1 with both continuations started from the same label $z \in E_j^n$: the residual sums are then within $C_{j,\ell}\varepsilon_n$ in L^p, with a constant independent of z, and the Gaussian factors are coupled identically. For (2), the bound follows edge by edge from the tower property: conditioning on $\mathcal{F}_h$ and using $\int e^{p'\ell} H_h(y; d\ell, dy') \le \int e^{q|\ell|} H_h(y; d\ell, dy') \le C$ uniformly in y, valid since $p' < 2 < q$, gives a factor C per edge and hence $C^{\ell - j}$ over the finite grid, uniformly in z; the independent Gaussian factors contribute $\prod_h e^{p'(p'-1)\alpha_h/2}$, and the atomic kernels inherit the bound by Lemma 4.1. For (3), both returns are residual laws convolved with $\gamma_{\alpha_{\ell-1}}$, so $\|\partial^i f\|_\infty \le \|\varphi^{(i)}_{\alpha_{\ell-1}}\|_\infty$ and Lemma 6.2 gives the convergence at every order. Hypothesis (4) is Assumption 6.3. Applying the same lemma to target continuations started from $y_n \to y$ and coupled through (15) gives continuity of the target jet map; the sequential claim follows by the triangle inequality. □

7 Projected dynamics and the density theorem

The first-edge factorization reduces both projections to covariance ratios over the residual kernel.

For the target first edge, Assumption 3.1 gives

$$X_1 - X_0 = L_0 + G_0, \qquad G_0 \perp (L_0, Y_1).$$

Homogeneity makes the continuation jet $A_m^\ell = a_m^{1,\ell}(Y_1)$ a function of the successor mark. Consequently $\mathrm{Cov}(A_m^\ell, X_1 - X_0) = \mathrm{Cov}_{H_0(y_0)}(a_m^{1,\ell}(Y_1), L_0)$ and $\mathrm{Var}(X_1 - X_0) = \alpha_0 + \mathrm{Var}_{H_0(y_0)}(L_0)$, so

$$d_m^\ell = \frac{\mathrm{Cov}_{H_0(y_0)}(a_m^{1,\ell}(Y_1), L_0)}{\alpha_0 + \mathrm{Var}_{H_0(y_0)}(L_0)}. \tag{28}$$

For the finite chain write its first-edge atoms as (λ_r^n, y_r^n) with weights p_r^n, and put

$$A_{m,r}^n = a_{m,n}^{1,\ell}(y_r^n), \qquad \bar{A}_m^n = \sum_r p_r^n A_{m,r}^n, \qquad \bar{\lambda}^n = \sum_r p_r^n \lambda_r^n.$$

Then

$$d_{m,n}^\ell = \frac{\sum_r p_r^n (A_{m,r}^n - \bar{A}_m^n)(\lambda_r^n - \bar{\lambda}^n)}{\alpha_0 + \sum_r p_r^n (\lambda_r^n - \bar{\lambda}^n)^2}. \tag{29}$$

Lemma 7.1 (Projected rows from converging continuation jets)**.** *Let (R_n, ζ_n) and (R, ζ) be coupled first-edge log returns and successor labels, ζ_n taking values in a finite set $\mathcal{Z}_n$ and ζ in a metric label space, with $R_n \to R$ and $\zeta_n \to \zeta$ in L^p for some $p > 2$. Let A_n on $\mathcal{Z}_n$ and A on the label space be bounded uniformly in n, with A uniformly continuous and $\max_{z \in \mathcal{Z}_n} |A_n(z) - A(z)| \to 0$. If $\mathrm{Var}(R) > 0$, then*

$$\frac{\mathrm{Cov}(A_n(\zeta_n), R_n)}{\mathrm{Var}(R_n)} \longrightarrow \frac{\mathrm{Cov}(A(\zeta), R)}{\mathrm{Var}(R)}.$$

Proof. Since $|A_n(\zeta_n) - A(\zeta)| \le \max_{z\in\mathcal{Z}_n} |A_n(z) - A(z)| + |A(\zeta_n) - A(\zeta)|$, and the second term tends to zero in probability by uniform continuity, $A_n(\zeta_n) \to A(\zeta)$ in probability, and the uniform bound upgrades this to $L^{p/(p-1)}$. Hölder's inequality against $R_n \to R$ in L^p gives convergence of $\mathbb{E}[A_n(\zeta_n)R_n]$ and of the means, and $p > 2$ gives $\mathrm{Var}(R_n) \to \mathrm{Var}(R) > 0$. □

Lemma 7.2 (Projected readout continuity)**.** *For every finite m and every $\ell \ge 2$,*

$$d^{\ell}_{m,n} \longrightarrow d^{\ell}_{m}.$$

Proof. Apply Lemma 7.1 under the first-edge coupling of Theorem 5.2, with $R_n = X^n_1 - x_0$, $R = X_1 - x_0$, $\zeta_n = Y^n_1$, $\zeta = Y_1$, $A_n = a^{1,\ell}_{m,n}$ and $A = a^{1,\ell}_{m}$. Proposition 6.5 supplies the hypotheses on A_n and A: continuity on the compact E is uniform and bounds A, and uniform convergence then bounds the A_n uniformly in n. Finally $\mathrm{Var}(X_1 - X_0) \ge \alpha_0 > 0$. The two sides are (29) and (28). □

Let $\mathfrak{T}_{\mathrm{SV}}$ be the class of homogeneous rooted targets on the fixed grid and state space that satisfy Assumptions 3.1, 3.2, and 6.3. Let $\mathfrak{G}_{\mathrm{SV}}$ be the finite marked GM chains of Definition 5.1 whose finite labels lie in the homogeneous mark space E, with the stated log and spot moments and well-defined raw readouts, and set

$$\mathfrak{M}_{\mathrm{SV}} := \mathfrak{T}_{\mathrm{SV}} \cup \mathfrak{G}_{\mathrm{SV}}, \qquad \tau^{\mathrm{SV}}_{\mathrm{mov}} := \tau_{\mathrm{mov}}(\mathfrak{M}_{\mathrm{SV}}).$$

Theorem 7.3 (Homogeneous-SV density at every finite projected order)**.** *Let $P \in \mathfrak{T}_{\mathrm{SV}}$ and let $\mathcal{L} \subseteq \{2, \dots, J\}$ be finite. Then the finite marked GM chains P^n of Theorem 5.2 satisfy:*

1. *every P^n is an exact asset martingale;*
2. *for every $j \ge 1$, the log-price marginal $\mathrm{Law}(X^n_j)$ is a finite Gaussian mixture and converges to $\mathrm{Law}(X_j)$ in W_p, while the corresponding finite lognormal spot marginal converges in W_1; the $j = 0$ marginal remains the fixed root δ_{x_0};*
3. *for every finite N and every $\ell \in \mathcal{L}$, the static jets through a^{ℓ}_{N+1} and the projected rows through d^{ℓ}_{N} converge;*
4. *for every finite N and each $\ell \in \mathcal{L}$ with the regular pivot $a^{\ell}_1 \neq 0$,*

$$v^{\ell}_{m,n} \longrightarrow v^{\ell}_{m}, \qquad 0 \le m \le N.$$

Consequently

$$\mathfrak{T}_{\mathrm{SV}} \subseteq \overline{\mathfrak{G}_{\mathrm{SV}}}^{\,\tau^{\mathrm{SV}}_{\mathrm{mov}}},$$

and $\mathfrak{G}_{\mathrm{SV}}$ is dense in $\mathfrak{M}_{\mathrm{SV}}$ both for the raw-readout topology and for $\tau^{\mathrm{SV}}_{\mathrm{mov}}$. Equivalently,

$$k_*(P; \mathfrak{G}_{\mathrm{SV}}) = \infty.$$

Proof. Items 1 and 2 are Theorem 5.2 and the paragraph following its proof. Proposition 6.5 with $j = 0$, whose only label is the root mark, gives all finite static jets, and Lemma 7.2 gives all finite projected rows. The standard inversion (8) is a continuous triangular rational map near a target with $a^{\ell}_1 \neq 0$, so the velocity jets converge. The sequence does not depend on N or $\mathcal{L}$, and every finite collection of raw readout coordinates converges along it, so Lemma 2.3 gives density in the initial topology. The marginal convergence in item 2 gives density in $\tau^{\mathrm{SV}}_{\mathrm{mov}}$. □

Proposition 7.4 (The density does not extend to the wings)**.** *For a chain P and a maturity t_ℓ let $\beta_R(P) := \limsup_{k\to\infty} w_{0,\ell}(k)/k$ be the right-wing slope of the root smile, and $\tilde p(P) := \sup\{u \geq 0 : \mathbb{E}_P[e^{(1+u)X_\ell}] < \infty\}$ the right moment index. If $\mathrm{Law}_Q(X_\ell)$ is a finite Gaussian mixture — in particular for every chain in $\mathfrak{G}_{\mathrm{SV}}$ or $\mathfrak{G}_{\mathrm{lift}}$ — then $\beta_R(Q) = 0$. If $\tilde p(P) < \infty$, then*

$$\beta_R(P) = 2 - 4\Big(\sqrt{\tilde p(P)^2 + \tilde p(P)} - \tilde p(P)\Big) > 0.$$

Consequently, once β_R is adjoined to the readout, no target with a finite right moment index lies in the closure of either finite class; the left wing and the left moment index behave alike.

Proof. A finite Gaussian mixture has exponential moments of every order, so $\tilde p(Q) = \infty$, and Lee's moment formula [33] gives $\beta_R(Q) = 0$. For P the same formula gives the displayed value, which is positive because $\sqrt{\tilde p^2 + \tilde p} < \tilde p + \frac{1}{2}$ for every finite $\tilde p \geq 0$. The set $\{Q : |\beta_R(Q) - \beta_R(P)| < \frac{1}{2}\beta_R(P)\}$ is then a neighbourhood of P in the augmented initial topology containing no chain of either class, the finite-mixture property of their marginals being Proposition 2.5. □

Proposition 7.5 (Fixed marginals cap the dynamics)**.** *Let Q be a chain as in Proposition 2.5 whose date-1 return marginal $\mathrm{Law}(X_1 - x_0)$ is*

$$\mu_1 = \sum_{i=1}^{n} p_i \mathcal{N}\big(y_i - \tfrac{1}{2}s,\ s\big), \qquad y_1 < \cdots < y_n, \quad p_i > 0, \quad s > 0.$$

Then its first edge has variance s, its first-edge locations take the value y_i with probability p_i, and every projected row is

$$d^\ell_m = \frac{1}{\mathrm{Var}(\mu_1)} \sum_{i=1}^{n} p_i\,(y_i - \bar y)\,\bar A^\ell_m(i), \qquad \bar y := \sum_i p_i y_i, \tag{30}$$

where $\bar A^\ell_m(i)$ is the average of the continuation jet A^ℓ_m over the first-edge branches located at y_i. In particular, if $n = 1$, every chain in either finite class whose first marginal is Gaussian has $d^\ell_m = 0$ for all m and ℓ.

Proof. By Proposition 2.5, $\mu_1 = \nu_1 * \mathcal{N}(0, \alpha_0)$ with ν_1 finitely supported and α_0 the first-edge variance, while $\mu_1 = \big(\sum_i p_i \delta_{y_i - s/2}\big) * \mathcal{N}(0, s)$. If $\alpha_0 < s$, then ν_1 is a Gaussian convolution, which is never finitely supported. If $\alpha_0 > s$, then $\widehat{\nu_1}(\xi) = e^{(\alpha_0 - s)\xi^2/2} \sum_k p_k e^{\mathrm{i}\xi(y_k - s/2)}$, which is unbounded, unlike a characteristic function, because the almost periodic sum returns arbitrarily close to its value 1 at the origin. Hence $\alpha_0 = s$, and uniqueness gives $\nu_1 = \sum_i p_i \delta_{y_i - s/2}$. Writing Λ for the first-edge location, $X_1 - x_0 = \Lambda - \frac{1}{2}s + G_0$ with G_0 independent of the label chain, so the numerator of (6) is $\mathrm{Cov}(A^\ell_m, \Lambda)$; grouping the branches by location, and using $\sum_i p_i(y_i - \bar y) = 0$, gives (30), the denominator being $\mathrm{Var}(\mu_1)$. If $n = 1$, Λ is constant. □

Remark 7.6 (How much the cap binds)**.** A general martingale chain with the same two marginals is not capped, since its continuation may depend on where X_1 falls inside a component. With lognormal marginals at the SPX one- and four-month at-the-money volatilities of Section 10.2 ($n = 1$), every finite chain has $d_0 = 0$, while martingale couplings of grid discretizations of the same marginals reach $d_0 \in [-0.251, 0.243]$, stably under refinement. At the resolution of the SPX fit ($n = 36$) the cap is mild but strict: since the readout averages jets after the convex at-the-money inversion, branches sharing a location can move it, and a piecewise-linear relaxation

over such branches certifies $d_0 \in [-0.7722, 0.1210]$ for the whole finite class, attained to within $3 \cdot 10^{-4}$, against $[-0.838, 0.156]$ for grid couplings. Both contain the value $d_0 = -0.057$ ($v_0 = 1.34$) of [18]. Refining the atoms is thus what lets the finite class express dynamics, and the bands of Section 10.2 are computed inside the capped set.

8 A convergence rate

Theorem 7.3 produces a sequence, not a budget. Under the same assumptions the construction admits an explicit rate in the number of atoms per node, linear in the quantization error rather than square-root: the exponential barycentre is exact at every resolution, so no accuracy is spent restoring martingality, and the fixed-variance factor makes the readout a Lipschitz functional of the residual law. Throughout, $E \subset \mathbb{R}^d$ has diameter D, L is the constant in (15), $q > p > 2$ and C are those of (16), and $\alpha := \min_j \alpha_j > 0$.

Lemma 8.1 (Quantization at a rate)**.** *Let H be as in Lemma 4.1, with $\int e^{q|\ell|} H \le C$. For every $\delta \in (0, e^{-1/2}]$ there is a finitely supported $H^\delta = \sum_r p_r \delta_{(\lambda_r, y_r)}$ with $y_r \in E$, with at most*

$$M \le c(p, q, d)\,(1 + D/\delta)^d\,\delta^{-1}\big(1 + \log(1/\delta)\big) \tag{31}$$

atoms, with exact barycentre $\sum_r p_r e^{\lambda_r} = 1$ and two-sided exponential bound $\sum_r p_r e^{q|\lambda_r|} \le 2C$, such that

$$W_p(H^\delta, H) \le c(p, q, C)\,\delta\,\big(1 + \log(1/\delta)\big). \tag{32}$$

Equivalently, in terms of the atom count,

$$\varepsilon(M) := W_p(H^{\delta(M)}, H) \le c\,M^{-1/(d+1)} (\log M)^{(d+2)/(d+1)}. \tag{33}$$

Proof. Let $(L, Y) \sim H$ and put $R := (2p/q)\log(1/\delta)$, so that $qR \ge p$. Partition $\mathbb{R}$ into $\lceil 2R/\delta \rceil$ intervals of length at most δ covering $[-R, R]$ and the two tails $\{\ell > R\}$, $\{\ell < -R\}$, and E into at most $c_d(1 + D/\delta)^d$ sets of diameter at most δ; the cells are all products of the two partitions, tails included, and number at most (31). On a cell A with $H(A) > 0$ set $\lambda_A = \log \mathbb{E}[e^L \mid A]$ and $y_A = \mathbb{E}[Y \mid A] \in E$. The map $(L, Y) \mapsto (\lambda_A, y_A)$ on A is a coupling of H with $H^\delta := \mathrm{Law}(\lambda_A, y_A)$; the tower property gives the exact barycentre and the conditional Jensen estimates of Lemma 4.1 give the exponential bound.

For the cost, $u \mapsto \log \mathbb{E}[e^u]$ is monotone, so λ_A lies between the essential infimum and supremum of L on A; hence $|L - \lambda_A| \le \delta$ on every core cell, and $|Y - y_A| \le \delta$ on every cell. The core contribution is therefore at most $(2\delta)^p$.

It remains to sum over the tail cells, rather than estimate their number. From $x^p \le c_{p,q} e^{qx/2}$ for $x \ge 0$ and $e^{q|L|/2} \mathbf{1}_{\{|L| > R\}} \le e^{-qR/2} e^{q|L|}$,

$$\sum_{A \subset \{|L| > R\}} \mathbb{E}\big[|L|^p \mathbf{1}_A\big] = \mathbb{E}\big[|L|^p \mathbf{1}_{\{|L| > R\}}\big] \le c_{p,q} C e^{-qR/2}. \tag{34}$$

On a lower-tail cell, $L < -R$ and Jensen gives $-\lambda_A \le \mathbb{E}[-L \mid A]$, so another application of Jensen and then summation give

$$\sum_{A \subset \{L < -R\}} H(A) |\lambda_A|^p \le \mathbb{E}\big[|L|^p \mathbf{1}_{\{L < -R\}}\big].$$

On an upper-tail cell, $\lambda_A \geq R$ and conditional Jensen gives

$$H(A)e^{q\lambda_A} = H(A)\big(\mathbb{E}[e^L \mid A]\big)^q \leq \mathbb{E}[e^{qL}\mathbf{1}_A].$$

Consequently, $x^p \leq c_{p,q}e^{qx/2}$ and Cauchy–Schwarz across the upper-tail cells yield

$$\begin{aligned}\sum_{A\subset\{L>R\}} H(A)|\lambda_A|^p &\leq c_{p,q}\sum_{A\subset\{L>R\}} H(A)e^{q\lambda_A/2}\\ &\leq c_{p,q}\,H(L>R)^{1/2}\mathbb{E}[e^{qL}\mathbf{1}_{\{L>R\}}]^{1/2} \leq c_{p,q,C}e^{-qR/2},\end{aligned}$$

where the last inequality uses $H(L>R)\leq Ce^{-qR}$. By $|L-\lambda_A|^p \leq 2^{p-1}(|L|^p+|\lambda_A|^p)$, the total tail cost is therefore $O(e^{-qR/2}) = O(\delta^p)$. Together with the Y-cost this gives $W_p(H^\delta, H)\leq c\delta$, and hence the weaker displayed bound (32). Finally, choosing $\delta \asymp (\log M/M)^{1/(d+1)}$ in (31) and substituting into (32) gives (33) for all sufficiently large M; enlarging the constant and using the one-atom exponential-barycentric approximation covers the remaining $M \geq 2$. □

The second ingredient replaces the splitting at a scale δ in the proof of Lemma 6.4, which would cost a square root if optimized: with the factor present, the ATM tail probability is the expectation of a Lipschitz function of the residual.

Lemma 8.2 (Linear readout response)**.** *Fix $\alpha > 0$ and let $R = Z + G$ and $\widetilde{R} = \widetilde{Z} + G$ with $G \sim \gamma_\alpha$ independent of Z and $\widetilde{Z}$. Write $C(k) = \mathbb{E}[(e^R - e^k)^+]$ and $\widetilde{C}(k) = \mathbb{E}[(e^{\widetilde{R}} - e^k)^+]$, and let ε be the L^p cost of a coupling of Z and $\widetilde{Z}$, $p > 2$. If $\mathbb{E}[e^{p'Z}] \vee \mathbb{E}[e^{p'\widetilde{Z}}] \leq K$ with $p' = p/(p-1)$, then*

$$\big|\widetilde{C}(0) - C(0)\big| \leq 2K^{1/p'}\varepsilon, \qquad \big|\widetilde{C}^{(i)}(0) - C^{(i)}(0)\big| \leq c_i\left(\alpha^{-1/2}\vee\alpha^{-i/2}\right)\varepsilon \quad (i\geq 1),$$

with c_i absolute.

Proof. The level bound is $|e^u - e^v| \leq |u-v|(e^u+e^v)$ and Hölder, as in Lemma 6.4. For $i = 1$, $C'(k) = -e^k\mathbb{P}(R>k)$, so

$$C'(0) = -\mathbb{P}(Z+G>0) = -\mathbb{E}\left[\Phi\left(\frac{Z-\alpha/2}{\sqrt{\alpha}}\right)\right], \tag{35}$$

because $G \sim \mathcal{N}(-\alpha/2, \alpha)$. The map $z \mapsto \Phi((z-\alpha/2)/\sqrt{\alpha})$ is Lipschitz with constant $(2\pi\alpha)^{-1/2}$, so the difference of the two expectations is at most $(2\pi\alpha)^{-1/2}W_1(\widetilde{Z}, Z) \leq (2\pi\alpha)^{-1/2}\varepsilon$. For $i \geq 2$, the identity $C^{(i)} = C^{(i-1)} + \partial_k^{i-2}(e^k f(k))$ expresses $C^{(i)}(0)$ through $C'(0)$ and the values $f^{(j)}(0)$, $j \leq i-2$, where f is the density of R; the same holds for $\widetilde{C}$. Each return density is a residual law convolved with φ_α, so Lemma 6.2 gives $\|\widetilde{f}^{(j)} - f^{(j)}\|_\infty \leq \|\varphi_\alpha^{(j+1)}\|_\infty\varepsilon = c_j\alpha^{-(j+2)/2}\varepsilon$; with the $\alpha^{-1/2}$ of $C'(0)$ the largest power is $\alpha^{-1/2}\vee\alpha^{-i/2}$, attained at $j = i-2$ when $\alpha \leq 1$. □

Theorem 8.3 (Rate for the homogeneous class)**.** *Let $P \in \mathfrak{T}_{\mathrm{SV}}$, let N be finite and $\mathcal{L} \subseteq \{2,\dots,J\}$ finite. For $M \geq 2$ let P^M be the finite marked GM chain of Theorem 5.2 built from Lemma 8.1 with at most M atoms per node. Then every P^M is an exact asset martingale,* $\mathrm{Law}(X_j^M)$ *is a Gaussian mixture of at most M^j components, and*

$$\big\|\mathcal{R}^{\mathrm{raw}}_{N,\mathcal{L}}(P^M) - \mathcal{R}^{\mathrm{raw}}_{N,\mathcal{L}}(P)\big\|_\infty \leq C_*\,M^{-1/(d+1)}(\log M)^{(d+2)/(d+1)}, \tag{36}$$

where

$$C_* = c\big(N,p,q,C,D,d,\underline{w},\overline{w}\big)\,(1+L)^J\big(1+\alpha^{-(N+1)/2}\big)\big(1+\alpha_0^{-1}\big).$$

On a chart with $a_1^\ell \neq 0$ the velocity jets satisfy the same bound, with a constant depending in addition on $|a_1^\ell|^{-1}$.

Proof. Apply Lemma 8.1 at every selected node with the common budget M, giving a per-edge error $\varepsilon = \varepsilon(M)$ as in (33). The coupling recursion (25) gives $\max_{j\le J} e_j \le 2\varepsilon\big((1+L)^J - 1\big)/L$, and $2J\varepsilon$ when $L = 0$; the same recursion started from a common label bounds the continuation residuals of Lemma 6.1 by $C_{\mathrm{tree}}\varepsilon$, uniformly over labels. The order-N readout needs $C^{(i)}(0)$ only for $i \le N+1$, through a^{ℓ}_{N+1}, and Lemma 8.2 converts this into $\max_{i\le N+1} |C_n^{(i)}(0) - C^{(i)}(0)| \le c\,(1+\alpha^{-(N+1)/2})C_{\mathrm{tree}}\varepsilon$, uniformly over reachable states, the exponential moments being uniform by the tower-property estimate in Proposition 6.5. By Assumption 6.3 the target's ATM variances lie in a compact subset $[\underline{w}, \overline{w}]$ of $(0,\infty)$. The approximants' are at least α, by Jensen's inequality in the independent factor, and bounded above through the exponential moments, so for every M all of them lie in one compact window, on which the jet map of Lemma 2.1 is C^∞ with denominator bounded away from zero; it is therefore Lipschitz in the call-derivative vector, which gives (36) for the static jets. For the rows, the quantization coupling is explicit, so $\|a^{1,\ell}_{m,n}(Y_1^n) - a^{1,\ell}_{m}(Y_1)\|_{L^{p'}}$ is bounded by the uniform level error plus $C_{\mathrm{tree}}\varepsilon$ through the Lipschitz label dependence of Lemma 6.1; Hölder against $L_0^n - L_0$ and $\mathrm{Var}(X_1 - X_0) \ge \alpha_0$ then give the same bound for $d^{\ell}_{m,n} - d^{\ell}_{m}$ through (29). The triangular inversion (8) is smooth near a target with $a^{\ell}_1 \ne 0$, which transfers the rate to the velocity jets. □

Proposition 8.4 (Lower bounds)**.** *Fix $p \ge 1$ and use on $\mathbb{R}\times E$ the metric $|\ell - \bar\ell| + |y - \bar y|$.*

1. *If $H \in \mathcal{P}(\mathbb{R}\times E)$ has a density bounded below by $\rho > 0$ on a cube of side s, then every law $\widetilde H$ with at most M atoms satisfies $W_p(\widetilde H, H) \ge c\,M^{-1/(d+1)}$, with $c = c(\rho, s, d, p) > 0$.*

2. *If $\mathrm{Law}_P(Y_1)$ has a density bounded below by $\rho > 0$ on a cube of side s in E, then every chain Q whose date-1 mark takes at most M values satisfies*

$$\mathcal{AW}_p(P, Q) \ge W_p\big(\mathrm{Law}_P(Y_1), \mathrm{Law}_Q(Y_1)\big) \ge c\,M^{-1/d}.$$

Proof. Both parts are one volume argument in dimension $d' = d+1$, respectively $d' = d$. Balls of radius r about the M atoms cover at most half of the cube when $M v_{d'} r^{d'} \le \frac{1}{2} s^{d'}$, where $v_{d'}$ is the volume of the unit ball; the uncovered half carries mass at least $\frac{1}{2}\rho s^{d'}$ at distance at least r from every atom, and every coupling must move that mass onto the atoms, so $W_p^p \ge \frac{1}{2}\rho s^{d'} r^p$. Take $r = s(2v_{d'}M)^{-1/d'}$. For (2), the adapted distance dominates the Wasserstein distance of the date-1 marginals, and the mark marginal of Q has at most M atoms. □

Remark 8.5 (What the rate does and does not say)**.** By Proposition 8.4(1) the quantization exponent $1/(d+1)$ is optimal up to the logarithm, so exact martingality costs nothing in it. When the date-1 mark law has a density bounded below on a cube, part (2), the recursion (25) and (33) give

$$c\,M^{-1/d} \le \mathcal{AW}_p(P^M, P) \le C\,M^{-1/(d+1)}(\log M)^{(d+2)/(d+1)},$$

the lower bound holding for every finite chain, martingale or not; the exponents differ by one because the return coordinate is smoothed rather than quantized, and whether a fixed factor closes the gap is open. Neither bound concerns the readout alone: Proposition 10.1 matches the order-N readout of one edge exactly with at most $3N+8$ components, so a readout lower bound would have to come from the multi-step recursion, and we have none. The constant of Theorem 8.3 degrades like $\alpha^{-(N+1)/2}$, so a high-order readout needs a factor that is not small relative to the residual scale, and the bound is per node, the date-j marginal having up to M^j components.

9 Local stochastic volatility

Local volatility rests on mimicking [28, 15]; the targets here are grid skeletons of local-stochastic-volatility models, whose return kernel depends on the log price and which supply no Gaussian factor. A finite proxy log price records where the local kernel is sampled, while the traded log price receives independent mean-one Gaussian innovations of vanishing variance; the class contains the homogeneous one (Remark 9.14).

9.1 Local-SV kernels and readout regularity

A motivating diffusion is

$$dX_t = -\tfrac{1}{2}\sigma(t, X_t, Y_t)^2 dt + \sigma(t, X_t, Y_t)\left(\rho(t, Y_t)^\top dW_t + \sqrt{1 - \|\rho(t, Y_t)\|^2}\, dB_t\right), \tag{37}$$

with Y as in (18), $\sigma > 0$, $\|\rho\| < 1$, and B independent of W. Conditioning on the W-path now leaves a state-dependent diffusion, so the argument of Proposition 3.3 does not apply, and we work directly with grid kernels.

Put $\mathsf{Z} := \mathbb{R} \times E$ with $d_{\mathsf{Z}}((x, y), (\bar{x}, \bar{y})) := |x - \bar{x}| + |y - \bar{y}|$, and for $z = (x, y)$ let $H_j(z; du, dy') := \mathrm{Law}_z(X_{j+1} - X_j, Y_{j+1})$.

Assumption 9.1 (Local-SV Markov kernels)**.** Fix $p > 2$ and $q > p$. The state space $E \subset \mathbb{R}^d$ is compact and convex, the root $z_0 = (x_0, y_0)$ is known, and $(X_j, Y_j)_{j=0}^J$ is a time-inhomogeneous Markov chain with kernels H_j. There are $L, C < \infty$ such that, for every $j < J$ and $z, \bar{z} \in \mathsf{Z}$,

$$\int e^u H_j(z; du, dy') = 1, \tag{38}$$

$$W_p(H_j(z), H_j(\bar{z})) \le L d_{\mathsf{Z}}(z, \bar{z}), \tag{39}$$

$$\sup_{z \in \mathsf{Z}} \int e^{q|u|} H_j(z; du, dy') \le C. \tag{40}$$

Assumption 9.2 (All-order terminal-kernel regularity)**.** For every $j < J$, the return marginal has a density

$$H_j(z; du, E) = q_j(u; z)\, du. \tag{41}$$

For each integer $r \ge 0$, $(u, z) \mapsto \partial_u^r q_j(u; z)$ is bounded and uniformly continuous on $\mathbb{R} \times \mathsf{Z}$. Equivalently, for suitable constants $M_{j,r} < \infty$ and moduli $\omega_{j,r}(s) \downarrow 0$ as $s \downarrow 0$,

$$\sup_{u,z} |\partial_u^r q_j(u; z)| \le M_{j,r}, \tag{42}$$

$$|\partial_u^r q_j(u; z) - \partial_u^r q_j(\bar{u}; \bar{z})| \le \omega_{j,r}(|u - \bar{u}| + d_{\mathsf{Z}}(z, \bar{z})). \tag{43}$$

Assumption 9.3 (Local-SV readout window)**.** For every date pair used by the readout, the target ATM implied total variance satisfies

$$0 < \underline{w} \le w_{j,\ell}(0; z) \le \overline{w} < \infty \qquad (z \in \mathsf{Z}), \tag{44}$$

and

$$\mathrm{Var}_{z_0}(X_1 - X_0) > 0. \tag{45}$$

For (37), bounded volatility gives the exponential moments and Lipschitz coefficients give (39); Assumption 9.2 holds for the following class.

Proposition 9.4 (A class satisfying Assumptions 9.1–9.3)**.** *Fix $\nu > 0$, let B' be a Brownian motion independent of (W, B), and write $\eta(t) := t_j$ for $t \in [t_j, t_{j+1})$. Consider the local-SV model with an independent variance floor and edge-frozen leverage,*

$$dX_t = -\tfrac{1}{2}\big(\sigma_t^2 + \nu^2\big)dt + \sigma_t\big(\rho(t, Y_t)^\top dW_t + \sqrt{1 - \|\rho\|^2}\, dB_t\big) + \nu\, dB'_t, \qquad \sigma_t = \sigma(t, X_{\eta(t)}, Y_t), \tag{46}$$

with Y as in (18) *and valued in a compact convex E (for instance by the boundary conditions of Proposition 3.4), and σ, ρ, b, Γ bounded and globally Lipschitz with $\|\rho\| \le \bar\rho < 1$. Then the grid chain (X_j, Y_j) is Markov and satisfies Assumptions 9.1, 9.2 and 9.3, with $\Delta_j = t_{j+1} - t_j$ and*

$$M_{j,r} = \big\|\varphi^{(r)}_{\nu^2\Delta_j}\big\|_\infty, \qquad \omega_{j,r}(s) = \big\|\varphi^{(r+1)}_{\nu^2\Delta_j}\big\|_\infty (1 + L)s,$$

where L is any constant realizing the synchronous-coupling estimate (39)*.*

Proof. On each edge the leverage reads the frozen spot X_{t_j} and the W-driven mark path, so from the current state $z = (x, y)$ the increment splits as $X_{j+1} - X_j = R_j + N_j$ with

$$\begin{aligned} N_j &:= -\tfrac{1}{2}\nu^2\Delta_j + \nu\,(B'_{t_{j+1}} - B'_{t_j}) \sim \gamma_{\nu^2\Delta_j}, \\ R_j &:= -\tfrac{1}{2}\int_{t_j}^{t_{j+1}} \sigma_t^2\, dt + \int_{t_j}^{t_{j+1}} \sigma_t\big(\rho^\top dW_t + \sqrt{1 - \|\rho\|^2}\, dB_t\big). \end{aligned}$$

With the leverage frozen at x, R_j depends on z and the (W, B) increments on the edge, and N_j on the B' increment only. Post-t_j increments of (W, B) and of B' are independent of each other and of $\mathcal{F}_{t_j}$, so $R_j \perp N_j$ given the state and

$$H_j(z; du, E) = \mathrm{Law}_z(R_j) * \gamma_{\nu^2\Delta_j}, \qquad \partial_u^r q_j(\cdot\,; z) = \mathrm{Law}_z(R_j) * \varphi^{(r)}_{\nu^2\Delta_j},$$

which gives (42). For the modulus, the mean value theorem bounds the u-increment by $\|\varphi^{(r+1)}_{\nu^2\Delta_j}\|_\infty |u - \bar u|$, and Lemma 6.2 gives

$$|\partial_u^r q_j(u; z) - \partial_u^r q_j(u; \bar z)| \le \|\varphi^{(r+1)}_{\nu^2\Delta_j}\|_\infty W_1\big(\mathrm{Law}_z(R_j), \mathrm{Law}_{\bar z}(R_j)\big).$$

Driving copies from $z, \bar z$ by the same (W, B, B') makes the N_j equal, and Gronwall, with $\sqrt{1 - \|\rho\|^2}$ Lipschitz because $\|\rho\| \le \bar\rho < 1$, bounds the L^p distance of the return–mark pairs by a multiple of $d_{\mathsf{Z}}(z, \bar z)$; this gives (39) and, adding the two moduli, (43). Since σ is bounded, Novikov's condition gives $\mathbb{E}_z[e^{R_j}] = 1$, so independence gives (38), and bounded coefficients give (40).

For Assumption 9.3, with $\bar\sigma := \sup \sigma$ the instantaneous variance lies in $[\nu^2, \bar\sigma^2 + \nu^2]$; since a call is convex, [21] places the conditional call price between the Black–Scholes prices at these volatilities, and monotonicity of $B(0, \cdot)$ places the ATM total variance in $[\nu^2(t_\ell - t_j), (\bar\sigma^2 + \nu^2)(t_\ell - t_j)]$, and $\mathrm{Var}_{z_0}(X_1 - X_0) \ge \mathrm{Var}(N_0) = \nu^2\Delta_0 > 0$. □

Remark 9.5 (Why the leverage is frozen)**.** Read continuously, $\sigma_t = \sigma(t, X_t, Y_t)$ depends on the past of B' within the edge, so R_j and N_j need not be independent and the convolution identity is lost; Lipschitz coefficients alone need not give bounded density derivatives of every order. The alternative is parabolic: smooth coefficients and a uniformly elliptic diffusion matrix for (X, Y) give Assumption 9.2 by [22, 32], but a chain invariant on a compact mark space degenerates at ∂E, so we rely on Proposition 9.4 or on Assumption 9.2 directly.

For the LSV target the smiles, jets and rows are those of Section 2 with $\zeta_j = (X_j, Y_j)$: $C_{j,\ell}(k; x, y) = \mathbb{E}_{x,y}[(e^{X_\ell - X_j} - e^k)^+]$, $a^\ell_m = a^{0,\ell}_m(x_0, y_0)$, $A^\ell_m = a^{1,\ell}_m(X_1, Y_1)$, and d^ℓ_m is the covariance ratio (6).

9.2 Finite proxy-GM chains

Definition 9.6 (Finite proxy-GM chain)**.** At date t_j let $\mathcal{Q}_j^n \subset \mathsf{Z}$ be finite and write a proxy label as $\zeta = (\xi, y)$. At each $\zeta \in \mathcal{Q}_j^n$ choose positive weights and successor data

$$p_{j,\zeta,r}^n, \qquad (\lambda_{j,\zeta,r}^n, y_{j,\zeta,r}^{n,+}) \in \mathbb{R} \times E, \qquad \sum_r p_{j,\zeta,r}^n = 1,$$

such that

$$\sum_r p_{j,\zeta,r}^n e^{\lambda_{j,\zeta,r}^n} = 1. \tag{47}$$

The successor label is

$$\zeta_{j,\zeta,r}^{n,+} := (\xi + \lambda_{j,\zeta,r}^n, y_{j,\zeta,r}^{n,+}) \in \mathcal{Q}_{j+1}^n.$$

For an edge variance $\beta_{j,n} > 0$, the transition from the full approximating state $(\widehat{x}, \zeta)$ is

$$\begin{aligned} &K_{j,n}^{\mathrm{pr}}((\widehat{x}, \zeta), d\widehat{x}', d\zeta') \\ &\quad = \sum_r p_{j,\zeta,r}^n \mathcal{N}\left(\widehat{x} + \lambda_{j,\zeta,r}^n - \frac{\beta_{j,n}}{2}, \beta_{j,n}\right)(d\widehat{x}') \delta_{\zeta_{j,\zeta,r}^{n,+}}(d\zeta'). \end{aligned} \tag{48}$$

The root is $(\widehat{X}_0^n, \Xi_0^n, \widehat{Y}_0^n) = (x_0, x_0, y_0)$. The proxy coordinate Ξ_j^n is part of the finite continuation label; the traded log price is $\widehat{X}_j^n$. Let $\widehat{\mathcal{F}}_j$ be the natural filtration of $(\widehat{X}_h^n, \Xi_h^n, \widehat{Y}_h^n)_{h \le j}$.

The normalized future return law depends only on the label ζ, so the approximating smiles and jets are written $C_{j,\ell,n}(k;\zeta)$ and $a_{m,n}^{j,\ell}(\zeta)$. Equivalently, after choosing branch r at $\zeta = (\xi, y)$, draw $G_{j,n} \sim \gamma_{\beta_{j,n}}$, independently of the complete proxy-label chain and of the other Gaussian innovations, and set

$$\Xi_{j+1}^n = \Xi_j^n + \lambda_{j,\zeta,r}^n, \quad \widehat{Y}_{j+1}^n = y_{j,\zeta,r}^{n,+}, \quad \widehat{X}_{j+1}^n = \widehat{X}_j^n + \lambda_{j,\zeta,r}^n + G_{j,n}. \tag{49}$$

Lemma 9.7 (Exact martingality and finite marginals)**.** *Every finite proxy-GM chain is an exact asset martingale. Put $B_{j,n} := \sum_{h<j} \beta_{h,n}$, with the convention $\gamma_0 = \delta_0$. Then*

$$\widehat{X}_j^n - \Xi_j^n = \sum_{h<j} G_{h,n} \sim \gamma_{B_{j,n}}, \tag{50}$$

independently of $(\Xi_j^n, \widehat{Y}_j^n)$. Consequently

$$\mathrm{Law}(\widehat{X}_j^n) = \sum_{(\xi,y) \in \mathcal{Q}_j^n} \mathbb{P}((\Xi_j^n, \widehat{Y}_j^n) = (\xi, y)) \mathcal{N}\left(\xi - \frac{B_{j,n}}{2}, B_{j,n}\right). \tag{51}$$

Thus every positive-date log-price marginal is a finite Gaussian mixture, with at most $|\mathcal{Q}_j^n|$ components.

Proof. Conditioning on the current state and using (47),

$$\mathbb{E}[e^{\widehat{X}_{j+1}^n} \mid \widehat{\mathcal{F}}_j] = e^{\widehat{X}_j^n} \sum_r p_{j,\zeta,r}^n e^{\lambda_{j,\zeta,r}^n} \mathbb{E}[e^{G_{j,n}}] = e^{\widehat{X}_j^n}.$$

Subtracting the proxy recursion from the traded-price recursion in (49) gives (50). By the independent product construction in (49), the complete proxy-label chain is independent of $(G_{h,n})_{h<J}$. Conditioning on the finitely many proxy labels gives (51). □

9.3 Diagonal quantization and lifted adapted convergence

Let φ_β denote the density of γ_β.

Lemma 9.8 (Diagonal martingale smoothing)**.** *Let $f_z(u)\,du$, $z \in \mathsf{Z}$, be probability laws whose derivatives through order R are bounded and uniformly continuous jointly in (u, z), and suppose $\int e^u f_z(u)\,du = 1$. Let $\Theta_n \subset \mathsf{Z}$ be finite and let*

$$\eta_{n,z} = \sum_r p_{n,z,r}\delta_{\lambda_{n,z,r}}, \qquad z \in \Theta_n,$$

satisfy $\sum_r p_{n,z,r}e^{\lambda_{n,z,r}} = 1$ and

$$\sup_{z\in\Theta_n} W_1(\eta_{n,z}, f_z(u)\,du) \le \varepsilon_n.$$

If $\beta_n \downarrow 0$ and

$$\varepsilon_n \max_{0\le i\le R} \|\varphi_{\beta_n}^{(i+1)}\|_\infty \longrightarrow 0, \tag{52}$$

then $g_{n,z} := \eta_{n,z} * \varphi_{\beta_n}$ has unit exponential mean and

$$\max_{0\le i\le R} \sup_{z\in\Theta_n} \|\partial_u^i g_{n,z} - \partial_u^i f_z\|_\infty \longrightarrow 0. \tag{53}$$

Proof. The exponential mean factorizes:

$$\int e^u g_{n,z}(u)\,du = \left(\sum_r p_{n,z,r}e^{\lambda_{n,z,r}}\right)\mathbb{E}[e^{G_{\beta_n}}] = 1.$$

Lemma 6.2 gives, for $i \le R$,

$$\|\partial_u^i(\eta_{n,z} * \varphi_{\beta_n} - f_z * \varphi_{\beta_n})\|_\infty \le \|\varphi_{\beta_n}^{(i+1)}\|_\infty \varepsilon_n.$$

Moreover $(f_z * \varphi_{\beta_n})^{(i)} = (\partial_u^i f_z) * \varphi_{\beta_n}$, and the common modulus of continuity gives

$$\sup_z \|(\partial_u^i f_z) * \varphi_{\beta_n} - \partial_u^i f_z\|_\infty \le \mathbb{E}[\omega_i(|G_{\beta_n}|)] \longrightarrow 0.$$

The limit follows by bounded convergence, the modulus being bounded. Combining the two bounds proves (53). Since β denotes variance, $\|\varphi_\beta^{(i+1)}\|_\infty = c_i\beta^{-(i+2)/2}$; thus $\varepsilon_n = o(\beta_n^{(R+2)/2})$ is a sufficient explicit rate. □

Equip the lifted state space $\mathbb{R} \times (\mathbb{R} \times E)$ with the metric $d^\uparrow((x,(\xi,y)),(\bar{x},(\bar{\xi},\bar{y}))) := |x-\bar{x}| + |\xi-\bar{\xi}| + |y-\bar{y}|$, and use it in the corresponding adapted Wasserstein distance.

Theorem 9.9 (Lifted finite-tree approximation)**.** *Under Assumptions 9.1 and 9.2, there are integers $r_n \uparrow \infty$, variances $\beta_n \downarrow 0$, and finite proxy-GM chains P^n with $\beta_{j,n} = \beta_n$ on every edge such that:*

1. *every transition is an exact asset-martingale finite-GM transition and every* $\mathrm{Law}(\widehat{X}_j^n)$, $j \ge 1$, *is a finite Gaussian mixture;*

2. *after lifting the target and approximating states to the common space*

$$Z_j^{\uparrow} = (X_j, (X_j, Y_j)), \qquad \widehat{Z}_j^n = (\widehat{X}_j^n, (\Xi_j^n, \widehat{Y}_j^n)),$$

their path laws converge in $\mathcal{AW}_p$;

3. *if $g_{j,n}(\cdot\,;\zeta)$ is the one-edge return density of* (48)*, then*

$$\max_{j<J} \max_{\zeta\in\mathcal{Q}_j^n} \max_{0\le i\le r_n} \|\partial_u^i g_{j,n}(\cdot\,;\zeta) - \partial_u^i q_j(\cdot\,;\zeta)\|_\infty \longrightarrow 0. \tag{54}$$

At each date,

$$W_p(\mathrm{Law}(\widehat{X}_j^n, \widehat{Y}_j^n), \mathrm{Law}(X_j, Y_j)) \longrightarrow 0, \tag{55}$$

and the spot marginals converge in W_1.

Proof. Choose any $r_n \uparrow \infty$. By Assumption 9.2 and the approximate-identity estimate in Lemma 9.8, one may choose $\beta_n \downarrow 0$ with

$$\max_{j<J} \max_{0\le i\le r_n} \sup_{z\in\mathsf{Z}} \|(\partial_u^i q_j(\cdot\,;z)) * \varphi_{\beta_n} - \partial_u^i q_j(\cdot\,;z)\|_\infty \le n^{-1}. \tag{56}$$

Put $D_n := 1 + \max_{0\le i\le r_n} \|\varphi_{\beta_n}^{(i+1)}\|_\infty$ and choose $0 < \varepsilon_n \le (nD_n)^{-1}$.

Start with $\mathcal{Q}_0^n = \{(x_0, y_0)\}$. Suppose $\mathcal{Q}_j^n$ is finite. For every $\zeta = (\xi, y) \in \mathcal{Q}_j^n$, apply Lemma 4.1 to the full joint law $H_j(\xi, y)$. Choose

$$H_{j,n}^{\mathrm{at}}(\zeta) = \sum_r p_{j,\zeta,r}^n \delta_{(\lambda_{j,\zeta,r}^n, y_{j,\zeta,r}^{n,+})} \tag{57}$$

with $W_p(H_{j,n}^{\mathrm{at}}(\zeta), H_j(\zeta)) \le \varepsilon_n$ and exact exponential barycentre (47). The conditional-Jensen estimate in Lemma 4.1 gives the uniform q-exponential bound for the selected return atoms. Define $\mathcal{Q}_{j+1}^n$ as the finite set of successor labels $(\xi + \lambda_{j,\zeta,r}^n, y_{j,\zeta,r}^{n,+})$. Recursion gives one coherent chain over the entire grid, including all option maturities.

Since $W_1 \le W_p$ and projection onto the return is 1-Lipschitz, the return marginal of (57) is within ε_n in W_1 of $q_j(u;\zeta)\,du$. Equations (56) and the definition of D_n therefore give (54) through the two estimates in the proof of Lemma 9.8, used directly because the order r_n grows: the quantization term is $\max_{i\le r_n} \|\varphi_{\beta_n}^{(i+1)}\|_\infty \varepsilon_n \le D_n\varepsilon_n \le n^{-1}$ by the choice of ε_n, and the approximate-identity term is at most n^{-1} by (56), both already uniform in $i \le r_n$, in j, and in the node. Exact martingality and finite-GM marginals follow from Lemma 9.7.

It remains to prove the adapted convergence. Given a current target state (X_j, Y_j) and proxy label $(\Xi_j^n, \widehat{Y}_j^n)$, the triangle inequality and (39) give

$$W_p\big(H_j(X_j, Y_j), H_{j,n}^{\mathrm{at}}(\Xi_j^n, \widehat{Y}_j^n)\big) \le L\big(|X_j - \Xi_j^n| + |Y_j - \widehat{Y}_j^n|\big) + \varepsilon_n. \tag{58}$$

Choose a Borel measurable ε_n-optimal coupling kernel, which exists by measurable selection of optimal plans [36, Corollary 5.22], the map $z \mapsto H_j(z)$ being continuous into $\mathcal{P}_p$ by (39). Denote the coupled target return–mark by (U_j, Y_{j+1}) and the proxy atom by $(\Lambda_j^n, \widehat{Y}_{j+1}^n)$. Set

$$X_{j+1} = X_j + U_j, \qquad \Xi_{j+1}^n = \Xi_j^n + \Lambda_j^n,$$

and independently add $G_{j,n} \sim \gamma_{\beta_n}$ to obtain $\widehat{X}^n_{j+1}$ as in (49). The joint one-step kernel has first marginal $H_j(X_j, Y_j)$ and second marginal $K^{\mathrm{pr}}_{j,n}((\widehat{X}^n_j, (\Xi^n_j, \widehat{Y}^n_j)), \cdot)$, each depending only on its own current state. Recursion therefore gives a bicausal coupling.

With $e_{j,n} := \big\| |X_j - \Xi^n_j| + |Y_j - \widehat{Y}^n_j| \big\|_{L^p}$, the argument of Theorem 5.2 with (58) in place of (24) gives the recursion (25) from $e_{0,n} = 0$, the roots coinciding, hence $\max_{j \le J} e_{j,n} \le C_{J,L}\varepsilon_n$. By (50),

$$\|\widehat{X}^n_j - \Xi^n_j\|_{L^p} \le c_p\sqrt{J\beta_n} + \frac{1}{2}J\beta_n.$$

For the lifted-state distance,

$$|\widehat{X}^n_j - X_j| + |\Xi^n_j - X_j| + |\widehat{Y}^n_j - Y_j| \le 2\big(|X_j - \Xi^n_j| + |Y_j - \widehat{Y}^n_j|\big) + |\widehat{X}^n_j - \Xi^n_j|.$$

Summing over the finite grid gives the quantitative bound

$$\mathcal{AW}_p(\mathrm{Law}(\widehat{Z}^n), \mathrm{Law}(Z^{\uparrow})) \le C_{p,J,L}(\varepsilon_n + \sqrt{J\beta_n} + J\beta_n), \tag{59}$$

and hence adapted convergence. Projection by $(x, (\xi, y)) \mapsto (x, y)$ is 1-Lipschitz for $d^{\uparrow}$, which proves (55). In particular $\widehat{X}^n_j \to X_j$ in law, so $e^{\widehat{X}^n_j} \Rightarrow e^{X_j}$; both have first moment e^{x_0}, by exact martingality, and for nonnegative laws weak convergence with converging first moments is W_1 convergence. □

Since normalized proxy returns depend only on the label, the same recursion can be started from any $\zeta = (\xi, y) \in \mathcal{Q}^n_j$ with both log prices normalized to ξ; it gives for the coupled continuations $(X^{\zeta}_h, Y^{\zeta}_h)$ and $(\widehat{X}^{n,\zeta}_h, \Xi^{n,\zeta}_h, \widehat{Y}^{n,\zeta}_h)$ the uniform bound

$$\sup_{\zeta \in \mathcal{Q}^n_j} \left(\mathbb{E}_{\zeta} \left[\sum_{h=j}^{J} \Big(|X^{\zeta}_h - \widehat{X}^{n,\zeta}_h| + |X^{\zeta}_h - \Xi^{n,\zeta}_h| + |Y^{\zeta}_h - \widehat{Y}^{n,\zeta}_h| \Big)^p \right] \right)^{1/p} \le C_{j,J}(\varepsilon_n + \sqrt{\beta_n} + \beta_n). \tag{60}$$

9.4 Continuation densities and smile jets

Fix $j < \ell$. Under the target continuation from $z \in \mathsf{Z}$, write $V := X_{\ell-1} - X_j$ and $Z_{\ell-1} := (X_{\ell-1}, Y_{\ell-1})$. Conditioning at $t_{\ell-1}$ gives the density

$$f_{j,\ell}(u; z) = \mathbb{E}_z[q_{\ell-1}(u - V; Z_{\ell-1})]. \tag{61}$$

For the proxy continuation from $\zeta = (\xi, y)$, normalize the traded starting log price to ξ and write $\widehat{V}^n = \widehat{X}^n_{\ell-1} - \widehat{X}^n_j$. Its terminal density is

$$f^n_{j,\ell}(u; \zeta) = \mathbb{E}_{\zeta}[g_{\ell-1,n}(u - \widehat{V}^n; \Xi^n_{\ell-1}, \widehat{Y}^n_{\ell-1})]. \tag{62}$$

Lemma 9.10 (Terminal-kernel transfer)**.** *For each fixed $R < \infty$ and $j < \ell$,*

$$\max_{0 \le i \le R} \max_{\zeta \in \mathcal{Q}^n_j} \|\partial^i_u f^n_{j,\ell}(\cdot\,; \zeta) - \partial^i_u f_{j,\ell}(\cdot\,; \zeta)\|_{\infty} \longrightarrow 0. \tag{63}$$

For each i, $(u, z) \mapsto \partial^i_u f_{j,\ell}(u; z)$ is bounded and uniformly continuous.

Proof. For large n, $r_n \geq R$; couple the preterminal continuations by (60). Differentiation under the expectation is justified by dominated convergence, with the bounds $M_{\ell-1,i}$ of (42) and, for large n, $M_{\ell-1,i}+1$ for $g_{\ell-1,n}$ by (54). Add and subtract $\partial_u^i q_{\ell-1}(u-\widehat{V}^n;\Xi^n_{\ell-1},\widehat{Y}^n_{\ell-1})$. The first difference is bounded uniformly by (54). With $D_{n,\zeta} := |\widehat{V}^n - V| + |\Xi^n_{\ell-1} - X_{\ell-1}| + |\widehat{Y}^n_{\ell-1} - Y_{\ell-1}|$, the second is bounded, uniformly in u, by $\mathbb{E}_\zeta[\omega_{\ell-1,i}(D_{n,\zeta})]$. Choose the modulus bounded by $2M_{\ell-1,i}$. For every $\delta > 0$,

$$\sup_{\zeta\in\mathcal{Q}^n_j} \mathbb{E}_\zeta[\omega_{\ell-1,i}(D_{n,\zeta})] \leq \omega_{\ell-1,i}(\delta) + 2M_{\ell-1,i}\delta^{-p} \sup_{\zeta\in\mathcal{Q}^n_j} \mathbb{E}_\zeta[D^p_{n,\zeta}].$$

The second term tends to zero by (60); letting $\delta \downarrow 0$ proves (63). The same calculation for two target continuations, via (39), gives the uniform continuity. □

Proposition 9.11 (Local-SV conditional static jets)**.** *Under Assumptions 9.1–9.3, for every $j < \ell$ and every fixed m,*

$$\max_{\zeta\in\mathcal{Q}^n_j} |a^{j,\ell}_{m,n}(\zeta) - a^{j,\ell}_m(\zeta)| \longrightarrow 0. \tag{64}$$

The target map $z \mapsto a^{j,\ell}_m(z)$ is bounded and uniformly continuous on Z.

Proof. Apply Lemma 6.4 with $\mathcal{Z}_n = \mathcal{Q}^n_j$ and the coupled normalized continuation returns. Hypothesis (1) is (60); (2) follows from (40) by the edgewise tower argument in the proof of Proposition 6.5, the Gaussian innovations contributing a factor bounded uniformly in n and the atomic kernels inheriting the bound by Lemma 4.1; (3) is Lemma 9.10, the target bounds coming from (61) and (42); and (4) is Assumption 9.3. The same lemma applied to two target continuations coupled through (39), with the uniform continuity of Lemma 9.10, gives a common modulus for the jets, hence bounded and uniformly continuous $z \mapsto a^{j,\ell}_m(z)$. □

Lemma 9.12 (Local-SV projected rows)**.** *For every fixed m and $\ell \geq 2$,*

$$d^\ell_{m,n} \longrightarrow d^\ell_m. \tag{65}$$

Proof. Apply Lemma 7.1 under the first-edge coupling of Theorem 9.9, with $R_n = \widehat{X}^n_1 - x_0$, $R = X_1 - x_0$, $\zeta_n = (\Xi^n_1, \widehat{Y}^n_1)$ and $\zeta = (X_1, Y_1)$. The coupling gives $\zeta_n \to \zeta$ in L^p, and $R_n \to R$ in L^p because the proxy gap (50) vanishes with β_n. Proposition 9.11 supplies the jets, bounded and uniformly continuous, and their uniform convergence bounds the A_n uniformly in n; (45) gives the variance. □

9.5 The second density theorem

Let $\mathfrak{T}_{\mathrm{LSV}}$ be the rooted Markov chains on the grid, with state space $\mathbb{R}\times E$ and known root, whose kernels satisfy Assumptions 9.1–9.3, such as the chains of Proposition 9.4. Let $\mathfrak{G}_{\mathrm{lift}}$ be the finite proxy-GM chains of Definition 9.6 with the stated moments and well-defined raw readouts, and put

$$\mathfrak{M}_{\mathrm{LSV}} := \mathfrak{T}_{\mathrm{LSV}} \cup \mathfrak{G}_{\mathrm{lift}}, \qquad \tau^{\mathrm{LSV}}_{\mathrm{mov}} := \tau_{\mathrm{mov}}(\mathfrak{M}_{\mathrm{LSV}}).$$

Theorem 9.13 (Local-SV density at every finite projected order)**.** *Let $P \in \mathfrak{T}_{\mathrm{LSV}}$ and let $\mathcal{L} \subseteq \{2,\dots,J\}$ be finite. The single diagonal sequence P^n of Theorem 9.9 satisfies items 1–4 of Theorem 7.3, with $\widehat{X}^n$ in place of X^n. Consequently*

$$\mathfrak{T}_{\mathrm{LSV}} \subseteq \overline{\mathfrak{G}_{\mathrm{lift}}}^{\tau^{\mathrm{LSV}}_{\mathrm{mov}}}, \qquad k_*(P;\mathfrak{G}_{\mathrm{lift}}) = \infty.$$

Proof. Items 1 and 2 are Theorem 9.9 and Lemma 9.7. Lemma 9.10 and Proposition 9.11 give every fixed finite collection of root and continuation static jets, the root jets being the case $j = 0$, whose only label is the root; the diagonal $r_n \uparrow \infty$ makes one sequence sufficient for all finite orders. Lemma 9.12 gives the projected rows. On the regular chart, the triangular map (8) is continuous, which gives the velocity jets. Raw-coordinate and marginal convergence is convergence in every generator of (10), so $P^n \to P$ in $\tau_{\mathrm{mov}}^{\mathrm{LSV}}$; as P was arbitrary, the class inclusion follows. □

Remark 9.14 (Relation to the homogeneous theorem)**.** The hypotheses of Theorem 7.3 imply those of Theorem 9.13: under Assumption 3.1 the return density is $\mathrm{Law}(L_j) * \varphi_{\alpha_j}$, so $\|\partial_u^r q_j\|_\infty \le \|\varphi_{\alpha_j}^{(r)}\|_\infty$ and, by Lemma 6.2 and (15),

$$|\partial_u^r q_j(u; y) - \partial_u^r q_j(u; \bar{y})| \le L\|\varphi_{\alpha_j}^{(r+1)}\|_\infty |y - \bar{y}|,$$

which is Assumption 9.2; the remaining conditions transfer by independence of G_j, and $\mathrm{Var}(X_1 - X_0) \ge \alpha_0 > 0$. The inclusion is strict: in (46) with σ depending on x, the residual law depends on X_{t_j}. The homogeneous theorem is kept because its regularity follows from a checkable structural condition and its proof needs no diagonal: with a fixed factor one sequence serves every order, without the rate coupling of Lemma 9.8 or the component growth of Remark 9.16.

Remark 9.15 (Finite regularity)**.** If Assumption 9.2 holds only through order $R = \max\{N-1, 0\}$ on the terminal edges of a finite maturity set $\mathcal{L}$, take $r_n \equiv R$ and impose (56) on those edges only, the only ones at which (54) enters (Lemma 9.10); the same proof gives closure in $\tau_{\mathrm{mov}}^{N,\mathcal{L}}(\mathfrak{T}_{\mathrm{LSV}}^{N,\mathcal{L}} \cup \mathfrak{G}_{\mathrm{lift}})$ for this finite-regularity class $\mathfrak{T}_{\mathrm{LSV}}^{N,\mathcal{L}}$. The all-order assumption is what yields one sequence converging in the full topology.

Remark 9.16 (Component budget)**.** With $\|\varphi_\beta^{(i+1)}\|_\infty = c_i \beta^{-(i+2)/2}$ the proof needs $\varepsilon_n \le (nD_n)^{-1}$, $D_n = 1 + \max_{i \le r_n} c_i \beta_n^{-(i+2)/2}$. On the schedule $r_n = n$, $\beta_n = 10^{-n}$ this is below 10^{-25} at $n = 6$, and the component count in (51) grows accordingly — driven by the derivative order, unlike the homogeneous construction. At a fixed order ($r_n \equiv R$, Remark 9.15) the admissible error is $n^{-1}\beta_n^{(R+2)/2}$ up to a constant.

Corollary 9.17 (Rate at a fixed readout order in the local-SV class)**.** *Run the lifted construction with per-node quantization error ε and a common innovation variance β, and fix a readout order $m \le N$.*

1. *For $m = 0$ the error is $O(\varepsilon + \beta)$, so there is no balance to strike. For $m \ge 1$, if $\partial_u^{m-2} q_j$ is Hölder-γ in u uniformly in the label, $\gamma \in (0, 1]$ — at $m = 1$ read $\partial_u^{-1} q_j$ as the distribution function, Lipschitz because q_j is bounded, so $\gamma = 1$ — the errors in a_m^ℓ and d_m^ℓ are at most $c_m\big(\beta^{-m/2}\varepsilon + M_m \beta^{\gamma/2}\big)$, optimized at $\beta \asymp \varepsilon^{2/(m+\gamma)}$, giving $O\big(\varepsilon^{\gamma/(m+\gamma)}\big)$.*

2. *If in addition the target return densities are C^m with bounded derivatives — as in Proposition 9.4, where $\|\partial_u^i q_j\|_\infty \le \|\varphi_{\nu^2\Delta_j}^{(i)}\|_\infty$ — the innovation bias is $O(\beta)$ and the balance improves to $\beta \asymp \varepsilon^{2/(m+2)}$, giving*

$$\big|d_{m,n}^\ell - d_m^\ell\big| \vee \big|a_{m,n}^\ell - a_m^\ell\big| = O\big(\varepsilon^{2/(m+2)}\big), \qquad \varepsilon = \varepsilon(M) \text{ as in (33).}$$

Proof. Compare three objects: the target; the *smoothed target*, whose return from any state over $h \ge 1$ edges is the target's return plus an independent $G \sim \gamma_{h\beta}$; and the approximant. The error

splits into a quantization part, approximant against smoothed target, and a bias, smoothed target against target.

Quantization part. From a label ζ, the approximant's return over h edges is the sum of its proxy increments plus the accumulated innovations, which by (50) have law $\gamma_{h\beta}$ and are independent of the proxy-label chain. Coupling the proxy increments with the target's returns from the point ζ, as in Theorem 9.9, costs at most $C\varepsilon$ in L^p, uniformly in ζ; the innovations do not enter the proxy recursion, so this cost does not depend on β. Lemma 8.2 with the common factor G and $\alpha := h\beta \geq \beta$ then bounds the difference of $C^{(i)}(0)$ by $c\,\varepsilon$ at $i = 0$ and by $c_i\beta^{-i/2}\varepsilon$ for $1 \leq i \leq m$. Only density derivatives of order at most $m-2$ enter, which is where this differs from a bound through $\|\varphi_\beta^{(m+1)}\|_\infty$. The exponential moments are uniform by the tower argument in the proof of Proposition 9.11, and on the window of Assumption 9.3, slightly enlarged, the jet map of Lemma 2.1 is Lipschitz, as in the proof of Theorem 8.3. The static jets therefore differ by $c_m\beta^{-m/2}\varepsilon$, uniformly in the label. The same two lemmas, applied to smoothed target continuations from two states coupled through (39), show that the smoothed jets are Lipschitz in the state with constant $c_m\beta^{-m/2}$; no modulus of the target's own jets in the state is needed.

Rows. The first-edge innovation is independent of the date-1 label $\zeta_1 = (\Xi_1^n, \widehat{Y}_1^n)$, so the approximant's row has numerator $\mathrm{Cov}(A^\ell_{m,n}, \Lambda_0)$, with Λ_0 the first proxy increment, and denominator $\mathrm{Var}(\Lambda_0)+\beta$. Split $A^\ell_{m,n} - A^\ell_m$ into the approximant's jets against the smoothed jets at ζ_1, the smoothed jets at ζ_1 against those at $Z_1 = (X_1, Y_1)$, and the smoothed jets at Z_1 against the target's. The first term is the quantization part, the second is at most $c_m\beta^{-m/2}\mathbb{E}|\zeta_1 - Z_1| \leq c_m\beta^{-m/2}C\varepsilon$, and the third is the bias below, uniform in the state. Hölder's inequality against the first-edge returns, which are $C\varepsilon$ apart in L^p, and the denominators, which differ by at most $C\varepsilon + \beta$ and are bounded below by (45), give d^ℓ_m the same bound as the jets.

Bias. For a target return U from any state and an independent $G_\beta \sim \gamma_\beta$, the call function of $U + G_\beta$ is $C_\beta(k) = \mathbb{E}[e^{G_\beta}C(k - G_\beta)]$, so $C^{(i)}_\beta(k) = \mathbb{E}[e^{G_\beta}C^{(i)}(k - G_\beta)]$. Taylor's formula in G_β, whose first two moments are $-\beta/2$ and $\beta + \beta^2/4$, gives

$$C^{(i)}_\beta(0) - C^{(i)}(0) = \tfrac{\beta}{2}\big(C^{(i+2)}(0) - C^{(i+1)}(0)\big) + O(\beta^2) \tag{66}$$

when the return density has $i+2$ bounded derivatives, as in Proposition 9.4, and $C^{(i)}_\beta(0) - C^{(i)}(0) = O(\beta)$ already when it has i, the second-order remainder being $O(\mathbb{E}[G_\beta^2 e^{|G_\beta|}])$; over h edges, replace β by $h\beta$. Under (2) the bias is therefore $O(\beta)$, uniformly in the state, with a constant controlled by $\|\partial_u^m q_j\|_\infty$; at $m = 0$ it is $O(\beta)$ because the density is bounded. Under (1) only the modulus is available and the cruder $M_m\mathbb{E}[e^{G_\beta}|G_\beta|^\gamma] \leq cM_m\beta^{\gamma/2}$ is used.

Equating the two parts gives the stated balances; at $m = 0$ the quantization part carries no power of β, so no balance arises. □

One might keep the target's own floor instead: in Proposition 9.4 the increment splits as $R_j + N_j$ with $N_j \sim \gamma_{\nu^2\Delta_j}$ independent of R_j given the state, and quantizing R_j alone would give the fixed-factor situation of Section 8. But a finite chain samples its kernel at the finitely valued proxy label, the target at its own traded log price, and a fixed innovation keeps the two apart. The proxy enters through the approximant's filtration, not through the cost: let $\Pi_{\mathrm{bc}}(P, Q)$ be the couplings of P and a finite proxy-GM chain Q that are bicausal for the natural filtration of (X, Y) under P and for $\widehat{\mathcal{F}}$ under Q, and put

$$\mathcal{AW}_1^{\mathrm{proj}}(P, Q) := \inf_{\pi \in \Pi_{\mathrm{bc}}(P,Q)} \mathbb{E}_\pi\Big[\sum_{j=0}^{J}\big(|X_j - \widehat{X}_j| + |Y_j - \widehat{Y}_j|\big)\Big].$$

It charges only the traded price and the mark and is at most a multiple of the lifted distance of Theorem 9.9, so a lower bound on it is the stronger statement.

Proposition 9.18 (A non-vanishing innovation obstructs adapted approximation)**.** *Let $P \in \mathfrak{T}_{\mathrm{LSV}}$, and suppose some edge kernel depends on the log price at every mark: for some $1 \le j < J$ and some bounded 1-Lipschitz ψ, the map*

$$x \longmapsto \Psi_j(x,y) := \int \psi(u)\, H_j((x,y);du,E)$$

is non-constant for every $y \in E$. If Q_n are finite proxy-GM chains whose total innovation variance before date j, $B_{j,n} := \sum_{h<j} \beta_{h,n}$, satisfies $\inf_n B_{j,n} > 0$, then $\inf_n \mathcal{AW}_1^{\mathrm{proj}}(P,Q_n) > 0$.

Proof. Suppose not; then there are indices, relabelled n, and $\pi_n \in \Pi_{\mathrm{bc}}(P,Q_n)$ with cost tending to zero. Put $\zeta_j := (\Xi_j^n, \widehat{Y}_j^n)$ and $G := \widehat{X}_j^n - \Xi_j^n$, which by (50) has law $\gamma_{B_{j,n}}$ and is independent of ζ_j. The laws of $\widehat{X}_j^n = \Xi_j^n + G$ converge, which is impossible if $B_{j,n}$ is unbounded, since then $\sup_c \mathbb{P}(G \in [c-R, c+R]) \to 0$ for every R; so along a further subsequence $B_{j,n} \in [b, \bar{b}]$ with $b > 0$. Under π_n, conditionally on the joint past at date j, bicausality and the Markov property make $X_{j+1} - X_j$ distributed as $H_j((X_j,Y_j);\cdot,E)$ and $\widehat{X}_{j+1}^n - \widehat{X}_j^n$ as a law $\kappa_n(\zeta_j)$ that depends on the proxy label alone, by (48). With $k_n := \int \psi\, d\kappa_n$ and ψ 1-Lipschitz,

$$\mathbb{E}_{\pi_n}\big|\Psi_j(X_j,Y_j) - k_n(\zeta_j)\big| \le \mathbb{E}_{\pi_n}|X_{j+1} - \widehat{X}_{j+1}^n| + \mathbb{E}_{\pi_n}|X_j - \widehat{X}_j^n| \longrightarrow 0.$$

By (39), Ψ_j is L-Lipschitz, so (X_j,Y_j) may be replaced by $(\widehat{X}_j^n, \widehat{Y}_j^n) = (\Xi_j^n + G, \widehat{Y}_j^n)$ at a cost tending to zero, and $\mathbb{E}|\Psi_j(\Xi_j^n + G, \widehat{Y}_j^n) - k_n(\zeta_j)| \to 0$. For (ξ, y, B) put $D(\xi,y,B) := \inf_{c \in \mathbb{R}} \mathbb{E}|\Psi_j(\xi + G_B, y) - c|$ with $G_B \sim \gamma_B$. Since G is independent of ζ_j, the last expectation is at least $\mathbb{E}[D(\Xi_j^n, \widehat{Y}_j^n, B_{j,n})]$. Since Ψ_j is bounded and L-Lipschitz, c may be restricted to $[-\|\psi\|_\infty, \|\psi\|_\infty]$, where the infimum is attained, and $\mathbb{E}|\Psi_j(\xi + G_B, y) - c|$ is equicontinuous in (ξ, y, B) uniformly in c, the G_B being coupled through one standard normal; so D is continuous. It is positive because G_B has full support: $D = 0$ would make the continuous function $\Psi_j(\cdot, y)$ equal to the minimizing c everywhere. The $\Xi_j^n = \widehat{X}_j^n - G$ are tight; a compact K with $\mathbb{P}(\Xi_j^n \in K) \ge \frac{1}{2}$ for all n and $c_K := \min_{K \times E \times [b,\bar{b}]} D > 0$ give $\bar{\mathbb{E}}[D] \ge c_K/2$, a contradiction. □

Kernels that do not depend on the log price are homogeneous, and fall under Theorem 7.3 when they carry a factor (Assumptions 3.1 and 3.2). Keeping the floor of Proposition 9.4 gives $B_{j,n} = \nu^2(t_j - t_0)$, so under the hypothesis it cannot converge in $\mathcal{AW}_1^{\mathrm{proj}}$, and quantizing the floor instead removes the Gaussian component it was meant to supply (Proposition 2.5). The obstruction concerns adapted approximation: a fixed innovation can still match finitely many readout coordinates on one edge (Proposition 10.1), and the vanishing need not be tied to the derivative order (Proposition 9.20).

Proposition 9.19 (The innovation term is sharp)**.** *Let $P \in \mathfrak{T}_{\mathrm{LSV}}$, and suppose there are $1 \le j < J$, a bounded 1-Lipschitz ψ, constants $\kappa > 0$ and $s \in (0,1]$, and a Borel set $S \subseteq \mathbb{R} \times E$ with $\varpi := \mathbb{P}((X_j,Y_j) \in S) > 0$, such that whenever $(x,y) \in S$ and $|y' - y| \le s$ the map $\Psi_j(\cdot, y')$ is monotone on $[x-3s, x+3s]$ with $|\Psi_j(u',y') - \Psi_j(u,y')| \ge \kappa|u' - u|$ there. Then every finite proxy-GM chain Q with $B_j := \sum_{h<j} \beta_h \le b_0 := s^2/(8\log(8/\varpi))$ satisfies*

$$\mathcal{AW}_1^{\mathrm{proj}}(P,Q) \ge c\sqrt{B_j}, \qquad c := \varpi \min\Big\{\frac{1}{4}, \frac{p_0\kappa}{8(1+L)}\Big\}, \quad p_0 := \Phi(-\tfrac{1}{2}) - \Phi(-1).$$

For the chains of Theorem 9.9, $B_{j,n} = j\beta_n$, and with (59) this gives $c\sqrt{j\beta_n} \le \mathcal{AW}_1^{\mathrm{proj}}(P, P^n) \le C(\varepsilon_n + \sqrt{\beta_n})$ for large n: the innovation term of (59) is optimal. The construction has $\varepsilon_n \le (nD_n)^{-1}$ with $D_n \ge \|\varphi'_{\beta_n}\|_\infty = (2\pi e)^{-1/2}\beta_n^{-1}$, so $\varepsilon_n = o(\beta_n)$ and its adapted error is of exact order $\sqrt{\beta_n}$.

Proof. Let $\pi \in \Pi_{\mathrm{bc}}(P, Q)$ have cost C_π, and keep the notation of the proof of Proposition 9.18. The two estimates at its start give $\mathbb{E}_\pi|\Psi_j(\Xi_j + G, \widehat{Y}_j) - k(\zeta_j)| \le (1 + L)C_\pi$. If $C_\pi \ge \varpi s/4$, then $C_\pi \ge \frac{\varpi}{4}\sqrt{B_j}$ because $B_j \le s^2$. Otherwise Markov's inequality gives $\mathbb{P}(|\widehat{X}_j - X_j| > s) + \mathbb{P}(|\widehat{Y}_j - Y_j| > s) < \frac{\varpi}{2}$, and $\mathbb{P}(|G| > s) \le 2e^{-s^2/(8B_j)} \le \frac{\varpi}{4}$; on the intersection of $\{(X_j, Y_j) \in S\}$ with the complements, $|\Xi_j - X_j| \le 2s$ and $|\widehat{Y}_j - Y_j| \le s$, so on this event, of probability at least $\varpi/4$, ζ_j lies in the set S' of labels (ξ, y') within $2s$ and s of a point of S. For $(\xi, y') \in S'$ the map $\Psi_j(\cdot, y')$ is κ-monotone on $[\xi - s, \xi + s]$, which contains $\xi + G$ on the events $\{G + B_j/2 \in [-\sqrt{B_j}, -\frac{1}{2}\sqrt{B_j}]\}$ and $\{G + B_j/2 \in [\frac{1}{2}\sqrt{B_j}, \sqrt{B_j}]\}$, each of probability p_0; on them $\Psi_j(\xi + G, y')$ lies respectively below and above two values $\kappa\sqrt{B_j}$ apart, so every constant is at distance at least $\frac{1}{2}\kappa\sqrt{B_j}$ from one of them and $D(\xi, y', B_j) \ge \frac{1}{2}p_0\kappa\sqrt{B_j}$. Independence of G and ζ_j then gives $(1 + L)C_\pi \ge \mathbb{E}[D(\zeta_j, B_j)] \ge \frac{\varpi}{8}p_0\kappa\sqrt{B_j}$. Take the infimum over π. □

Proposition 9.20 (Linear rate when no evaluation follows the readout)**.** *Let the target be in the class of Proposition 9.4 with final-edge floor $\alpha := \nu^2\Delta_{J-1}$, and let $\mathcal{L} = \{J\}$. Run the construction of Theorem 9.9 with edge variances $\beta_{j,n} = \varepsilon_n^2$ for $j < J - 1$ and $\beta_{J-1,n} = \alpha$, the final-edge atoms quantizing the residual R_{J-1} alone. Then for every finite N,*

$$\left\|\mathcal{R}^{\mathrm{raw}}_{N,\{J\}}(P^n) - \mathcal{R}^{\mathrm{raw}}_{N,\{J\}}(P)\right\|_\infty \le C\,\varepsilon_n,$$

with $C = c(N, p, q, C, D, d, \underline{w}, \overline{w})\,(1 + L)^J\big(1 + \alpha^{-(N+1)/2}\big)\big(1 + (\nu^2\Delta_0)^{-1}\big)$, and with no diagonal and no order-dependent component budget.

Proof. The static jets at t_J, root and continuation, are functionals of return laws ending at J, each containing $N_{J-1} \sim \gamma_\alpha$ common to target and approximant, so Lemma 8.2 gives $\max_{i \le N+1}|C_n^{(i)}(0) - C^{(i)}(0)| \le c\,(1 + \alpha^{-(N+1)/2})\varepsilon_n$ once the parts before that factor are $O(\varepsilon_n)$ apart in L^p. They are: the kernels are evaluated at the proxy, which (25) keeps within $O(\varepsilon_n)$ of the target's log price (on the final edge for the residual, Lipschitz in the state with the same L since the synchronous coupling of Proposition 9.4 makes the floors equal), and the innovations before the final edge, the only ones preceding a kernel evaluation, have total variance at most $J\varepsilon_n^2$. For the rows, Lemma 8.2 on two synchronously coupled target continuations makes the continuation jets Lipschitz in the state, and the proof of Theorem 8.3 applies with denominator at least $\nu^2\Delta_0$. Definition 9.6 permits a different variance on each edge, and the jet map is Lipschitz on the window as in Theorem 8.3. □

Remark 9.21 (Why the two classes separate quantitatively)**.** Theorem 8.3 keeps α fixed, so the readout error is the quantization error times a constant; Corollary 9.17 must send $\beta \downarrow 0$ and pays $\varepsilon^{2/(m+2)}$ even in case (2). With weights 2^{-m} on the order-m coordinates, a metric compatible with the readout topology, the error is of order ε^θ, $\theta = \min\{1, 2\log 2/\log(1/\alpha)\}$, with a fixed factor and only $\exp(-c\sqrt{\log(1/\varepsilon)})$ with vanishing innovations. Neither exponent is intrinsic — polynomially decaying weights turn both into logarithms — but the gap is, being present coordinate by coordinate. All of these are upper bounds; Section 10.1 reports where they are loose.

10 Closed-form readout, cubature and numerics

The readout of a finite chain is available in closed form, which drives the discrete model of [29] and the computations below; one edge with fixed continuation admits exact finite-order matching by cubature.

If a normalized log return has the finite mixture law $R \sim \sum_{i=1}^{M} p_i \mathcal{N}(m_i, s_i^2)$, with $s_i > 0$ and $\sum_i p_i e^{m_i + s_i^2/2} = 1$, its call price is, as in [13],

$$C(k) = \sum_i p_i \left[e^{m_i + s_i^2/2} \Phi\left(\frac{m_i + s_i^2 - k}{s_i} \right) - e^k \Phi\left(\frac{m_i - k}{s_i} \right) \right]. \tag{67}$$

All strike derivatives are finite sums, so Lemma 2.1 gives the ATM jets by one scalar inversion and implicit differentiation, and the projected rows follow from the mixture formulas of Section 2.3, which for branch-constant continuation jets give (29) with variance α_0 (homogeneous) or $\beta_{0,n}$ (proxy).

Proposition 10.1 (One-step finite-order attainment)**.** *Fix a current state, one maturity, a fixed continuation, and the standard pivot $a_1 \neq 0$. Suppose the target one-step marked return has the common-variance representation*

$$\int \left[\mathcal{N}(\lambda(\theta) - \alpha/2, \alpha) \otimes \delta_{y(\theta)} \right] \Lambda(d\theta),$$

where $\alpha > 0$ and $\int e^{\lambda(\theta)} \Lambda(d\theta) = 1$. Write $m(\theta) = \lambda(\theta) - \alpha/2$, let $A_j(\theta)$ be the fixed-continuation jet, and define

$$C_\theta(k) := \mathbb{E}\left[\left(e^{\lambda(\theta) + G + R_\theta^{\mathrm{cont}}} - e^k \right)^+ \right], \qquad G \sim \gamma_\alpha, \quad G \perp R_\theta^{\mathrm{cont}},$$

where R_θ^{cont} is the continuation log return from $y(\theta)$. Assume every coordinate of

$$\left(1, e^\lambda, m, m^2 + \alpha, A_j, A_j m : 0 \le j \le N, \partial_k^h C_\theta(0) : 0 \le h \le N + 1 \right)$$

is in $L^1(\Lambda)$. For every N, there is a finite mixture with at most $3N + 8$ components which is an exact asset martingale and matches exactly the raw static and projected inputs needed to recover $v_0, \ldots, v_N$.

Proof. Use the displayed L^1 vector as the cubature test vector. The common positive Gaussian variance justifies differentiation under the Λ-integral: Gaussian derivative bounds dominate the strike derivatives, while the exponential-moment coordinate controls the call value. Including the constant, the test vector has

$$1 + 1 + 2 + 2(N + 1) + (N + 2) = 3N + 8$$

coordinates. Tchakaloff's theorem [6], applied to the law of the test vector with linear test functions, gives at most that many nodes in its support and positive weights p_i matching every coordinate. The coordinates depend on θ only through $(\lambda(\theta), y(\theta))$ and are continuous there, the continuation jets by Propositions 6.5 and 9.11; as m bounds λ along convergent sequences and E is compact, the range on $\operatorname{supp} \Lambda$ is closed, so every node is the value at some θ_i. A martingale version, with support inside that of the original law, is in [9]. Matching the martingale coordinate

Table 1: Theorem 8.3 at $d = 1$: readout error against the atom count $M = K^2$ with the factor variance held at $\alpha = 10^{-3}$. Last row: log-log slopes in K over the four finest resolutions, against the predicted $M^{-1/(d+1)} = K^{-1}$ up to logarithms.

		$\lvert a_m^M - a_m\rvert$			$\lvert d_m^M - d_m\rvert$		
K	M	$m=0$	$m=1$	$m=2$	$m=0$	$m=1$	$m=2$
6	36	$1.4\cdot10^{-3}$	$2.9\cdot10^{-3}$	$1.2\cdot10^{-2}$	$8.1\cdot10^{-4}$	$4.0\cdot10^{-3}$	$8.3\cdot10^{-3}$
13	169	$3.8\cdot10^{-4}$	$8.3\cdot10^{-4}$	$8.1\cdot10^{-4}$	$5.0\cdot10^{-4}$	$1.8\cdot10^{-3}$	$3.7\cdot10^{-3}$
28	784	$1.1\cdot10^{-4}$	$2.6\cdot10^{-4}$	$4.4\cdot10^{-4}$	$2.7\cdot10^{-4}$	$7.6\cdot10^{-4}$	$1.5\cdot10^{-3}$
58	3364	$3.4\cdot10^{-5}$	$9.3\cdot10^{-5}$	$2.5\cdot10^{-4}$	$1.3\cdot10^{-4}$	$3.2\cdot10^{-4}$	$6.3\cdot10^{-4}$
121	14641	$1.1\cdot10^{-5}$	$3.4\cdot10^{-5}$	$1.1\cdot10^{-4}$	$5.9\cdot10^{-5}$	$1.2\cdot10^{-4}$	$2.4\cdot10^{-4}$
175	30625	$6.5\cdot10^{-6}$	$2.1\cdot10^{-5}$	$7.3\cdot10^{-5}$	$3.7\cdot10^{-5}$	$7.4\cdot10^{-5}$	$1.4\cdot10^{-4}$
slope in K		-1.49	-1.35	-1.13	-1.15	-1.32	-1.34

gives $\sum_i p_i e^{\lambda_i} = 1$. Matching the first two moments and the A_j cross-moments gives the exact regression numerators and denominator, including the within-component variances. Matching the call derivatives gives the exact static jets through the order needed by the triangular inversion. Hence the recovered projected velocity jets agree. □

Proposition 10.1 concerns one edge with fixed continuation and variance, not the constructions, whose continuations are endogenous.

10.1 Two checks of the rates

Take two edges, $t_0 < t_1 < t_2$, and a latent mark Y_1 on $E = [0.10, 0.40]$ with a smooth density. Conditionally on $Y_1 = y$,

$$L_0 \mid \{Y_1 = y\} \sim \mathcal{N}\big(\mu(y) - \tfrac{1}{2}s(y)^2,\ s(y)^2\big)\,, \qquad s(y) = y\sqrt{\Delta_0}, \qquad \mu(y) = -\kappa(y - \bar{y}) + c,$$

with c fixed by $\mathbb{E}[e^{L_0}] = 1$, and from y the continuation return is a two-component Gaussian mixture with skewed means and y-proportional scales, renormalized to unit exponential mean. Both conditional laws are finite mixtures, so the maturity-t_2 jets $A_m = a_m^{1,2}(Y_1)$ follow from (67) and Lemma 2.1, and the target readout $d_m := d_m^2$ of (6) is a one-dimensional quadrature in y.

The approximant quantizes the joint law of (L_0, Y_1) into $M = K^2$ cells, K mark bands times K conditional return quantiles, so $d = 1$. Cell r carries mass p_r, mark $y_r = \mathbb{E}[Y_1 \mid r]$ and log-barycentre $\lambda_r = \log \mathbb{E}[e^{L_0} \mid r]$, so $\sum_r p_r e^{\lambda_r} = 1$ at every resolution; the first-edge return is λ_r plus the independent factor, and, the target being homogeneous, the label is the mark alone.

Theorem 8.3 assumes a factor of fixed variance, so we split $\alpha = 10^{-3}$ off both edges and hold it fixed, with $s(y)^2 = y^2\Delta_0 - \alpha > 0$ on E, and vary only the resolution. With $\kappa = 2$ and $\Delta_0 = \Delta_1 = 0.25$ its readout is

$$(a_0, a_1, a_2) = (0.074497,\ -0.133267,\ 0.012041),$$
$$(d_0, d_1, d_2) = (-0.117275,\ +0.216807,\ -0.418134),$$

and $\mathrm{Var}(X_1 - X_0) = 0.031458$. The continuation is kept exact at the cell's mark barycentre, which isolates the quantization error of Lemma 8.1 and its propagation.

In Table 1, $\sum_r p_r e^{\lambda_r}$ equals one to $7\cdot10^{-14}$ at every resolution, so martingality is exact rather than asymptotic, and every measured slope is at least 1 in modulus against a predicted $M^{-1/2}$ up

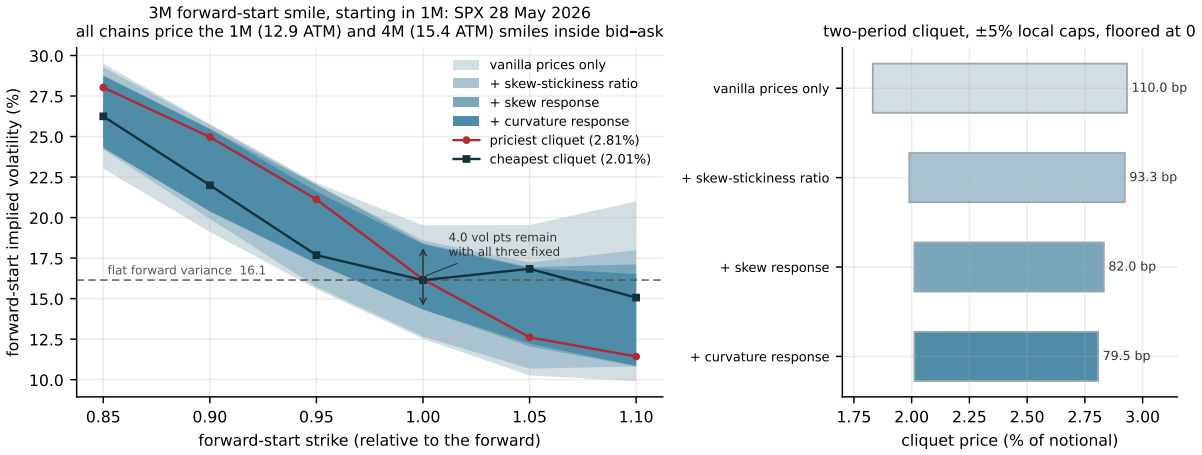


Figure 1: Forward-start and cliquet values of exact-martingale finite-GM chains that price the same SPX vanillas inside bid–ask, 28 May 2026. Left: the three-month forward-start smile starting in one month; the outer band is the finite-grid coupling range for the fixed fitted marginals, and the inner bands are the corresponding ranges under chains that also match v_0, then v_1, then v_2, pinned at the three-month values of [18]. For each constrained problem the shading uses dual outer endpoints; the explicit feasible-chain endpoints lie within 0.01 volatility points for the forward-start smile and within 0.04 basis points for the cliquet, so the inner and outer bands coincide at the plotted precision. The two lines are the chains at the cliquet endpoints under all three constraints. Right: the same for a two-period cliquet with ±5% local caps floored at zero.

to logarithms, the ratio to $M^{-1/2}(\log M)^{3/2}$ falling monotonically over the last four rows. The bound therefore holds with room; the projected rows, which carry the covariance against the quantized edge, track it most tightly.

The same setup shows where Corollary 9.17 is loose. For a floored return $U = S + N$, $N \sim \gamma_A$ with $A = 10^{-2}$ and S a two-component mixture, quantized into K equal-probability log-barycentric cells and smoothed by γ_β as in Section 9 (worst case over the phase of the cell grid relative to $k = 0$), the innovation bias is linear in β — over $\beta \in [2.8 \cdot 10^{-4}, 2.8 \cdot 10^{-3}]$ the error grows by factors 3.07 to 3.16 against a β ratio of 3.12, as (66) predicts — and at $m = 2$ the measured error exponent in ε is 0.72 at the worst phase and 0.92 at the median, against 1/2 from Corollary 9.17(2). The exponent is right in kind but conservative: equal-probability cells make the distribution function accurate to $O(\varepsilon)$ before any smoothing, which a bound through W_1 cannot see.

10.2 A constrained band on an SPX surface

Whether pinning a finite readout says anything about prices is an empirical question, and this section examines it on one surface in detail and then on sixty month-ends.

Data and fit. SPX end-of-day chains for 28 May 2026 at $\tau_1 = 0.090$ and $\tau_2 = 0.342$ years, with 227 and 353 two-sided quotes and forwards from put–call parity. A two-date chain with 36 and 47 atoms prices all 580 quotes inside their bid–ask spreads, with implied-volatility errors of 0.027 and 0.078 points root mean square, and is a martingale to $6 \cdot 10^{-17}$.

Experiment. Hold both fitted marginals fixed, so that every quote stays inside bid–ask, and search over the finite exact-martingale chains that join them, with the fitted atoms and component

variances: a chain is a finite set of first-date branches, each at a fitted location y_i with its own transition law ρ over the second-date components, $\sum_k \rho_k e^{z_k - y_i} = 1$, and total masses p_i per location and r_k per component as fitted. Several branches may share a location — a latent state the price does not reveal — and the readout then averages their jets (Proposition 7.5). Maximize and minimize the forward-start implied volatility for $[t_1, t_2]$ at strikes $\kappa \in \{0.85, 0.90, \dots, 1.10\}$, first under these constraints alone and then adding the velocity coordinates one at a time, each pinned to the three-month value of [18], $(v_0, v_1, v_2) = (1.336, 0.901, 37.26)$, within $(0.005, 0.01, 0.2)$.

Computation. Without the readout these are linear programs in the joint weights, the martingale optimal transport problem of [7, 26] on a finite grid. With it, each branch still enters linearly — through its mass, its law and its jets $J(C\rho)$, with $C\rho \in \mathbb{R}^3$ its continuation call price and first two strike derivatives at the money and J the map of Lemma 2.1 — and so does the velocity box, since with the date-2 marginal fixed the velocities are a fixed triangular image of the rows (8). Each program is thus a linear program over infinitely many candidate branches, solved by column generation; the resulting explicit chain is checked against every bid–ask interval and the prescribed readout box, so its value is an attained inner bound. Conversely, weak duality bounds the largest price by $\nu \cdot r + \sigma(\mu) + \sum_i p_i g_i$ for any multipliers ν on the second-date weights and μ on the velocity box, with σ the support function of the box and g_i the largest reduced value of one branch at location i; each g_i is nonconvex only through the three coordinates of $C\rho$ and is bounded numerically by branch-and-bound using analytic interval enclosures and explicit floating-point margins. Thus every true optimum is numerically bracketed between a feasible attained value and a weak-duality outer bound; the gaps are at most 0.01 volatility points at every reported endpoint. Code is available from the authors; the ORATS option data are proprietary.

Result. For the fixed fitted marginals, the finite-grid coupling range leaves the forward-start smile undetermined by 6.5 to 11.1 volatility points across the six strikes, 7.0 at the money (Figure 1). Matching the three readout coordinates as well narrows the range at every strike, to between 4.0 and 5.7 points and to 4.0 at the money, so that one half to three quarters of the vanilla range remains; at the money the skew response v_1 does most of the work and the curvature response v_2 the least. A two-period cliquet with ±5% local caps floored at zero, resetting at the two quoted expiries so that its periods are 33 and 92 days and paid at t_2, ranges over 110 basis points on a premium between 1.83 and 2.93% of notional, and over 79.5 with the three coordinates matched. For a desk this is a measure of residual model risk, not a model.

Across dates. Run on the last trading day of each of the sixty months from July 2021 to June 2026, with the two expiries chosen automatically near one and four months and the atom grids scaled to each surface, the same construction passes this fit screen on 55 of them — four of the other five fail the in-band threshold and one fails the root-mean-square-error threshold — and gives the same picture on every one (Figure 2): the velocity box is reachable, and the explicit attained values and dual outer bounds differ at the money by at most 0.0021 volatility points; the three coordinates remove a median 39% of the at-the-money vanilla range, quartiles 35 and 43% and never below 29%, a median 7.6 points narrowing to 4.4; for the cliquet the median removed is 25%. The staged picture repeats too: at the money v_0, v_1 and v_2 remove a median 32, 53 and 14% of the total narrowing, and v_1 is the largest single step on 52 of the 55 dates, its interval clear of the other two at every one of them. The pattern across strikes does not: on 47 of the 55 it is the 0.85 wing that is narrowed by the largest fraction, while the 1.10 wing is largest on the other eight; the median narrowing is 55% at 0.85 against 39% at the money, whereas this surface has it the other way about. The forward-start figure does not track the level of volatility (correlation −0.12 with the one-month at-the-money volatility, which ranges over 9.9 to 28.3%), whereas the cliquet figure does (−0.56), its local caps binding harder when volatility is high. Holding the

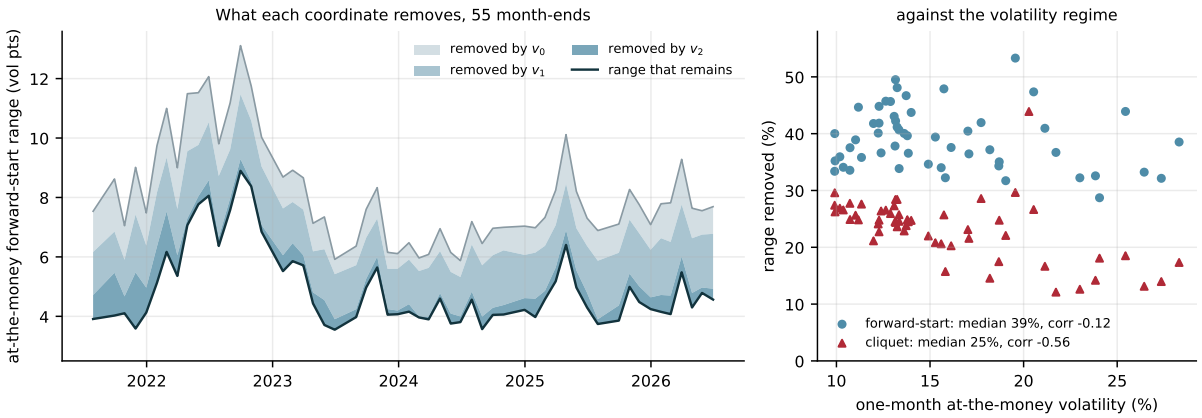


Figure 2: The same experiment on the last trading day of each of the sixty months from July 2021 to June 2026, on the 55 surfaces passing the fit screen: at least 95% of each expiry's quotes inside bid–ask and implied-volatility root-mean-square error at most 0.5 volatility points. Left: the at-the-money finite-grid coupling range for the fixed fitted marginals, divided into what each coordinate removes as it is added and what remains with all three pinned at the values of [18], shaded as in Figure 1; each true endpoint is bracketed by an explicit feasible-chain value and a dual outer bound, with gaps below 0.01 volatility points. Right: the fraction removed against the level of volatility, for both targets. The forward-start fraction does not track that level; the cliquet fraction does, its ±5% local caps binding harder when volatility is high.

component volatility at the 5% of this section instead of a quarter of the at-the-money volatility fits 51 of the sixty and removes more, a median 50%, with the same staged ordering.

What is and is not claimed. Every reported optimum over the class searched lies between its attained inner endpoint and its dual outer endpoint, whose gap is at most 0.01 volatility points. Every value between the two attained endpoints is attained, since mixing two chains branch by branch mixes prices and readouts. Chains with a single branch per location — continuation a function of the price location alone — can only have narrower constrained bands; sequential linear programming over them reaches 14.78 to 18.26 at the money, against 14.33 to 18.35 for the hidden-branch class at the displayed precision. Letting the marginals move within the spreads could only widen every band. The closest listed forward-starting claim, a VIX future, does not test the pinned coordinates: its square is a log contract over the whole smile rather than an at-the-money quantity, and on the VIX calendar of this surface chains matching the three coordinates still attain 0.715 to 0.994 of $\sqrt{\mathbb{E}[\mathrm{VIX}^2]}$, against 0.671 to 1 without them, while strike truncation in the variance-swap anchors pins the market ratio only to 0.97–1.03; the listed VIX prices therefore neither confirm nor contradict the readout. And the coefficients of [18] are physical-measure regressions of realized daily moves, whereas d_m^ℓ is a risk-neutral finite-step projection: they fix the readout at a plausible magnitude to demonstrate capacity, not to calibrate. The narrowing is not very sensitive over the stated one-coordinate-at-a-time sweeps: varying v_0 from 1 to 2, or v_1 and v_2 each from zero to twice the values used, with the other two held, keeps the at-the-money range removed between 41 and 47% and the cliquet's between 27 and 31%. The placebo comparisons suggest that the spot-response structure matters: the sampled white-noise and low-degree Hermite contrasts across first-date states, pinned as tightly relative to their spread, remove at most 16% of the at-the-money range, whereas the sampled spot-dependent

moments of the two returns remove about as much as the readout. Within these experiments, the narrowing is driven by pinning how the continuation law responds to the spot, which the readout records in quoted implied volatilities.

11 Scope and extensions

Coverage. Imperfect correlation and Lipschitz dependence on the mark are mild; the implied-variance window and $\mathrm{Var}(X_1 - X_0) > 0$ need total volatility bounded above and away from zero; the variance floor of Assumption 3.1 and a finite-dimensional Markov mark are structural, whereas compactness of E and uniform exponential moments are technical and should yield to localization. Finitely many known roots are handled by a disjoint union of trees.

target	status	reason
bounded SV with a variance floor	Theorem 7.3	by construction
square-root variance, floored and capped	Theorem 7.3	Proposition 3.4
square-root variance, unregularized	outside as stated	volatility degenerates
LSV with an independent variance floor	Theorem 9.13	Proposition 9.4
LSV, kernel with R derivatives	orders up to $R+1$	Remark 9.15
N-factor forward-variance models	Theorem 7.3	finite Markov state, truncated
curve-valued and rough volatility	outside as stated	E assumed finite-dimensional

Square-root variance. An unregularized square-root variance fails both theorems through degeneracy, not only noncompactness: localizing to $[\underline{v}, \overline{v}]$ gives a factor of variance $\theta\underline{v}(1-\bar{\rho}^2)\Delta_j \to 0$ as $\underline{v} \downarrow 0$, and one-step density derivatives of order $V^{-(r+1)/2}$. The regularizations that restore the hypotheses (Propositions 3.4 and 9.4) are of the kind implementations commonly apply.

Mark space. An N-factor forward-variance model, truncated as in Proposition 3.4, is covered. A curve-valued or rough mark is not ($E \subset \mathbb{R}^d$), although the proofs use only that E is compact and convex, as is any sup-norm-closed set of uniformly bounded curves with a common modulus of continuity.

Fixed marginals. For generic targets the moving-marginal formulation is the strongest one available, not a weakening (Section 2.6), and with the marginals held at prescribed finite-mixture laws the finite class is capped at the resolution of the first marginal (Proposition 7.5). A fixed-marginal density theorem can therefore hold only in a larger class, whose mixing weights depend on the continuous state. That route rests on stability of martingale couplings, which holds on the line [8, 37, 4] and fails on $\mathbb{R}^2$ [14]; whether it can be carried out with finitely many components per fibre is open.

Open directions. (i) *Instantaneous limit:* d_m is a regression over a fixed step; an instantaneous version needs a diagonal in the mesh. (ii) *Noncompact states:* Heston-type targets call for Lyapunov localization, as mimicking already handles degenerate covariance [15]. (iii) *Budgets:* the bound $3N + 8$ of Proposition 10.1 is one-step; a fixed budget yields a capacity profile across orders. (iv) *Other readouts:* fixed-strike implied volatilities can be adjoined by (26); for targets with a finite moment index wing slopes cannot (Proposition 7.4).

12 Conclusion

In the topology generated by the dynamics characteristics themselves, for the two target classes,

$$k_*(P;\mathfrak{G}_{\mathrm{SV}})=\infty \quad (P\in\mathfrak{T}_{\mathrm{SV}}), \qquad k_*(P;\mathfrak{G}_{\mathrm{lift}})=\infty \quad (P\in\mathfrak{T}_{\mathrm{LSV}})\,:$$

every finite order of projected at-the-money smile dynamics is matched to arbitrary accuracy by exact-martingale finite Gaussian-mixture chains with converging marginals, using the target's Gaussian factor (SV) or one manufactured at vanishing variance (LSV). The kernel that generates the dynamics is never observed; what these statements provide is a licence to search inside a closed-form class, in the coordinates that are observed, without excluding any finite-order behaviour these models produce.

Three qualifications travel with that licence. The rate is $M^{-1/(d+1)}$ up to logarithms, with only the latent dimension in the exponent (Theorem 8.3), and our bound degrades when the Gaussian innovations must vanish, which adapted approximation forces for innovations preceding a price-dependent kernel (Proposition 9.18); we do not prove the degradation itself necessary. The licence is at the money: finite mixtures have zero implied-variance wing slopes, so it does not extend to a readout containing the wings (Proposition 7.4). And it is a capacity statement, not a calibration procedure — though in the fixed-fitted-marginal finite-chain experiment on one SPX surface the coupling range for the fitted marginals is 7.0 points at the money, and fixing the first three characteristics narrows it to 4.0 (Section 10.2).

References

[1] Eduardo Abi Jaber and Shaun (Xiaoyuan) Li, *Capturing smile dynamics with the quintic volatility model: SPX, skew-stickiness ratio and VIX*, arXiv:2503.14158 (2025; revised 2026).

[2] Athanassia Bacharoglou, *Approximation of probability distributions by convex mixtures of Gaussian measures*, Proceedings of the American Mathematical Society 138 (2010), no. 7, 2619–2628, doi:10.1090/S0002-9939-10-10340-2.

[3] Julio Backhoff-Veraguas, Daniel Bartl, Mathias Beiglböck, and Manu Eder, *Adapted Wasserstein distances and stability in mathematical finance*, Finance and Stochastics 24 (2020), 601–632; arXiv:1901.07450, doi:10.1007/s00780-020-00426-3.

[4] Julio Backhoff-Veraguas and Gudmund Pammer, *Stability of martingale optimal transport and weak optimal transport*, Annals of Applied Probability 32 (2022), no. 1, 721–752, doi:10.1214/21-AAP1694.

[5] Daniel Bartl, Mathias Beiglböck, and Gudmund Pammer, *The Wasserstein space of stochastic processes*, Journal of the European Mathematical Society 28 (2026), no. 1, 393–454; arXiv:2104.14245.

[6] Christian Bayer and Josef Teichmann, *The proof of Tchakaloff's theorem*, Proceedings of the American Mathematical Society 134 (2006), 3035–3040, doi:10.1090/S0002-9939-06-08249-9.

[7] Mathias Beiglböck, Pierre Henry-Labordère, and Friedrich Penkner, *Model-independent bounds for option prices—a mass transport approach*, Finance and Stochastics 17 (2013), 477–501, doi:10.1007/s00780-013-0205-8.

[8] Mathias Beiglböck, Benjamin Jourdain, William Margheriti, and Gudmund Pammer, *Approximation of martingale couplings on the line in the adapted weak topology*, Probability Theory and Related Fields 183 (2022), no. 1–2, 359–413, doi:10.1007/s00440-021-01103-y.

[9] Mathias Beiglböck and Marcel Nutz, *Martingale inequalities and deterministic counterparts*, Electronic Journal of Probability 19 (2014), no. 95, 1–15, doi:10.1214/EJP.v19-3270.

[10] Lorenzo Bergomi, *Smile dynamics IV*, Risk Magazine, December 2009, 94–100.

[11] Lorenzo Bergomi, *Stochastic Volatility Modeling*, Chapman & Hall/CRC, 2016.

[12] Lorenzo Bergomi and Julien Guyon, *Stochastic volatility's orderly smiles*, Risk, May 2012, 60–66.

[13] Damiano Brigo and Fabio Mercurio, *Lognormal-mixture dynamics and calibration to market volatility smiles*, Int. J. Theor. Appl. Finance 5 (2002), no. 4, 427–446, doi:10.1142/S0219024902001511.

[14] Martin Brückerhoff and Nicolas Juillet, *Instability of martingale optimal transport in dimension $d \geq 2$*, Electronic Communications in Probability 27 (2022), paper no. 24, doi:10.1214/22-ECP463.

[15] Gerard Brunick and Steven Shreve, *Mimicking an Itô process by a solution of a stochastic differential equation*, Annals of Applied Probability 23 (2013), no. 4, 1584–1628, doi:10.1214/12-AAP881.

[16] René Carmona and Sergey Nadtochiy, *Local volatility dynamic models*, Finance and Stochastics 13 (2009), no. 1, 1–48.

[17] René Carmona and Sergey Nadtochiy, *Tangent models as a mathematical framework for dynamic calibration*, International Journal of Theoretical and Applied Finance 14 (2011), no. 1, 107–135.

[18] Charlie Che and Pradeepta Das, *Beyond the skew-stickiness ratio: transport geometry of spot-driven variance surface dynamics*, arXiv:2608.12493 (2026).

[19] Rama Cont and Peter Tankov, *Financial Modelling with Jump Processes*, Chapman & Hall/CRC, Boca Raton, 2004.

[20] Valdo Durrleman, *From implied to spot volatilities*, Finance and Stochastics 14 (2010), no. 2, 157–177.

[21] Nicole El Karoui, Monique Jeanblanc-Picqué, and Steven E. Shreve, *Robustness of the Black and Scholes formula*, Mathematical Finance 8 (1998), no. 2, 93–126, doi:10.1111/1467-9965.00047.

[22] Avner Friedman, *Partial Differential Equations of Parabolic Type*, Prentice-Hall, 1964.

[23] Masaaki Fukasawa, *Martingale expansion for stochastic volatility*, SIAM Journal on Financial Mathematics 17 (2026), no. 2, SC1–SC12, doi:10.1137/26M1841318; arXiv:2601.09324.

[24] Masaaki Fukasawa, *On the skew stickiness ratio*, arXiv:2602.05241 (2026).

[25] Siegfried Graf and Harald Luschgy, *Foundations of Quantization for Probability Distributions*, Lecture Notes in Mathematics 1730, Springer, Berlin, 2000.

[26] Alfred Galichon, Pierre Henry-Labordère, and Nizar Touzi, *A stochastic control approach to no-arbitrage bounds given marginals, with an application to lookback options*, The Annals of Applied Probability 24 (2014), no. 1, 312–336, doi:10.1214/13-AAP925.

[27] Julien Guyon and Jordan Lekeufack, *Volatility is (mostly) path-dependent*, Quantitative Finance 23 (2023), no. 9, 1221–1258.

[28] István Gyöngy, *Mimicking the one-dimensional marginal distributions of processes having an Itô differential*, Probability Theory and Related Fields 71 (1986), 501–516, doi:10.1007/BF00699039.

[29] Shaosai Huang, *SANOS-Evolve: a discrete stochastic-local-volatility model for European-option smile dynamics*, SSRN working paper no. 7151258 (2026); ssrn.com/abstract=7151258.

[30] Benjamin Jourdain and Gilles Pagès, *Quantization and martingale couplings*, ALEA, Latin American Journal of Probability and Mathematical Statistics 19 (2022), 1–22; arXiv:2012.10370, doi:10.30757/ALEA.v19-01.

[31] Ioannis Karatzas and Steven E. Shreve, *Brownian Motion and Stochastic Calculus*, 2nd ed., Graduate Texts in Mathematics 113, Springer, New York, 1991.

[32] Olga A. Ladyzhenskaya, Vsevolod A. Solonnikov, and Nina N. Ural'tseva, *Linear and Quasi-linear Equations of Parabolic Type*, American Mathematical Society, Providence, RI, 1968.

[33] Roger W. Lee, *The moment formula for implied volatility at extreme strikes*, Mathematical Finance 14 (2004), no. 3, 469–480, doi:10.1111/j.0960-1627.2004.00200.x.

[34] Philipp J. Schönbucher, *A market model for stochastic implied volatility*, Philosophical Transactions of the Royal Society of London A 357 (1999), no. 1758, 2071–2092.

[35] Martin Schweizer and Johannes Wissel, *Term structures of implied volatilities: absence of arbitrage and existence results*, Mathematical Finance 18 (2008), no. 1, 77–114.

[36] Cédric Villani, *Optimal Transport: Old and New*, Grundlehren Math. Wiss. 338, Springer, 2009.

[37] Johannes Wiesel, *Continuity of the martingale optimal transport problem on the real line*, Annals of Applied Probability 33 (2023), no. 6A, 4645–4692, doi:10.1214/22-AAP1928.