LOCAL-VOLATILITY CALIBRATION WITH HARD-CONSTRAINED DIFFERENTIABLE GALERKIN LAYERS

Anonymous authors
Paper under double-blind review

ABSTRACT

Local-volatility calibration is an ill-posed inverse parabolic problem, and standard pipelines typically separate surface interpolation from the pricing operator used downstream. We instead learn only the variance-rate coefficient $\kappa = \sigma_{\mathrm{loc}}^2/2$ and obtain prices from the exact log-moneyness Dupire equation through differentiable weighted-Galerkin layers. A positive parameterisation makes coefficient admissibility architectural; the weighted discretisation makes every implicit state matrix symmetric positive-definite (SPD); and reverse-mode gradients reuse the same sparse solves as the forward pass. Smoothness remains a statistical regulariser rather than a surrogate for either positivity or the PDE. We prove that finite soft penalties cannot certify membership in the admissible ellipticity set and test the distinction through synthetic recovery, a mechanism-level hard-versus-soft comparison, and out-of-sample calibration on an S&P 500 option snapshot. The results show that structural guarantees can be retained without claiming universal coefficient accuracy or a complete static-arbitrage certificate.

1 INTRODUCTION

Local-volatility calibration is an inverse parabolic problem: given an arbitrage-screened surface of European option prices, recover a maturity- and state-dependent diffusion coefficient whose pricing operator reproduces the observations. The classical Dupire construction provides the governing equation Dupire (1994); Derman & Kani (1994), but direct inversion differentiates noisy, irregular quotes in strike and maturity. Practical pipelines therefore fit a smooth implied-volatility surface—often SVI or SSVI—and differentiate the fit afterwards Gatheral (2004); Gatheral & Jacquier (2014). This is effective for interpolation, but it separates the inferred coefficient from the numerical operator subsequently used for pricing and risk.

Physics-informed neural networks, variational residual methods, and Deep Ritz models encode a differential equation through a finite training objective Raissi et al. (2019); Kharazmi et al. (2021); Yu et al. (2018). By contrast, classical finite elements and PDE-constrained optimisation provide coercive state solves and parameter-to-state derivatives Brenner & Scott (2008); Hinze et al. (2008). We combine these ideas in a deliberately narrow architecture: a trainable field specifies only the coefficient, while the state is always produced by an implicit Galerkin solve. Thus coefficient positivity and state solvability hold for every iterate, whereas data fit and inverse regularisation retain their statistical roles. Unlike a learned surrogate operator such as DeepONet or FNO Lu et al. (2021); Li et al. (2020), the resulting map is a differentiable numerical operator whose discretisation remains explicit.

Our contributions are threefold. First, we express the exact log-moneyness Dupire equation as a weighted symmetric problem and discretise it as an SPD implicit layer in which the learned coefficient enters a positive mass matrix. Second, we differentiate the complete coefficient-to-price map and formalise the distinction between hard admissibility and any finite soft penalty. Third, we evaluate the resulting inverse solver on synthetic coefficient recovery, a hard-versus-soft mechanism study, and maturity- and strike-block holdouts on S&P 500 options. Additional density, noise, obstacle, neural-operator, and uncertainty studies are reported in the appendices.

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

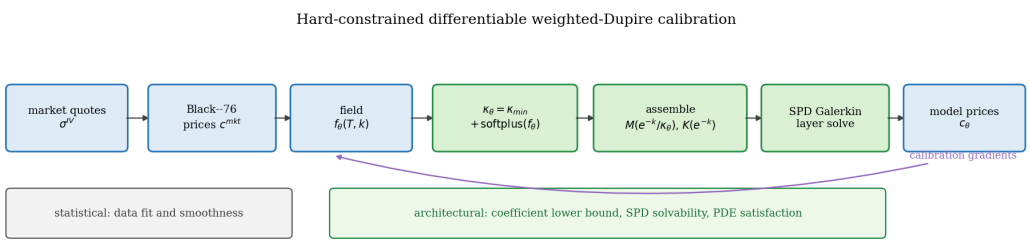

Figure 1: Hard-constrained exact-Dupire calibration. The learned positive coefficient enters the weighted mass matrix of $\mathcal{Q}_D$; fixed weighted stiffness and SPD solves produce prices, while Black–76 prices and smoothness regularisation define the statistical objective.

2 Related work

Local-volatility inversion. Dupire and Derman–Kani introduced the local-volatility representation Dupire (1994); Derman & Kani (1994); regularised coefficient recovery addresses its ill-posedness Crepey (2003). SVI/SSVI provide effective arbitrage-aware surface parameterisations Gatheral (2004); Gatheral & Jacquier (2014). Our distinction is architectural: coefficient regularisation is coupled to a differentiable variational solver.

Variational and structure-preserving SciML. PINNs, hp-VPINNs and Deep Ritz place differential or variational information in training objectives Raissi et al. (2019); Kharazmi et al. (2021); Yu et al. (2018). Hamiltonian, Lagrangian and port-Hamiltonian networks hard-code dynamical structure Greydanus et al. (2019); Lutter et al. (2019); Desai et al. (2021). Classical FEM theory already explains why a positive symmetric elliptic coefficient yields SPD; PDE-constrained optimisation already supplies parameter-to-state differentiation Brenner & Scott (2008); Hinze et al. (2008).

Neural operators. DeepONet, FNO and the graph kernel operator family motivate amortisation across problem instances, and they formalise the notion of discretisation invariance Lu et al. (2021); Li et al. (2020); Kovachki et al. (2023). This paper contributes a paired-seed comparison under shift in which the learned operators are more accurate in distribution and structurally unsafe outside it, and shows that one post-hoc σ-Q/SPD correction reduces targeted violations without restoring either accuracy or sampled-grid monotonicity.

Score-based diffusion. Anderson's reverse-time theorem underlies score-based SDEs and denoising diffusion probabilistic models Anderson (1982); Song et al. (2020); Ho et al. (2020); implicit and denoising score matching are the two standard estimators of the score Hyvärinen & Dayan (2005); Vincent (2011). Our contribution is to show that the Dupire dual supplies the same forward Fokker–Planck object exactly (8.5).

Bayesian deep learning and UQ for scientific ML. Bayesian inverse problems in function space, and their sampling algorithms, provide the reference framework Stuart (2010); Kaipio & Somersalo (2005); Cotter et al. (2013). Bayesian PINNs, Laplace approximations for deep networks and SciML-UQ surveys provide the modern comparators Yang et al. (2021); Daxberger et al. (2021); Psaros et al. (2023). Appendix D follows the standard validation protocol for such approximations, namely a parametric bootstrap Tibshirani & Efron (1993) and MCMC with split-$\hat{R}$ and effective-sample-size diagnostics Gelman & Rubin (1992); Geyer (1992).

3 Exact Weighted-Dupire Formulation

Let the risk-neutral asset price satisfy

$$\mathrm{d}S_t = (r - q)S_t\,\mathrm{d}t + \sigma_{\mathrm{loc}}(t, S_t)S_t\,\mathrm{d}W_t. \tag{1}$$

For discount factor $D(0,T)$, forward F_T, and log-moneyness $k = \log(K/F_T)$, define the normalised forward call $c(T,k) = C(T, F_T e^k)/(D(0,T)F_T)$. With deterministic rates and dividends, the exact

Dupire equation is

$$\partial_T c = \kappa(T,k)(\partial_{kk} c - \partial_k c), \qquad \kappa = \tfrac{1}{2}\sigma_{\mathrm{loc}}^2, \tag{2}$$

with $c(0,k) = (1-e^k)_+$. The first derivative is essential: a numeraire change alone does not turn (2) into a divergence-form diffusion with coefficient κ.

Set

$$m_\kappa(T,k) = \frac{e^{-k}}{\kappa(T,k)}, \qquad w(k) = e^{-k}. \tag{3}$$

Multiplying (2) by m_κ yields the exact weighted form

$$m_\kappa \partial_T c = \partial_k (w \partial_k c). \tag{4}$$

After subtracting a Dirichlet lift, let $\{\phi_i\}$ be the interior P_1 basis and define

$$[\mathbf{M}(m)]_{ij} = \int_\Omega m \phi_i \phi_j \,\mathrm{d}k, \qquad [\mathbf{K}(w)]_{ij} = \int_\Omega w \phi_i' \phi_j' \,\mathrm{d}k. \tag{5}$$

Backward Euler gives

$$\mathbf{A}_n(\kappa) c^n = \mathbf{M}(m_\kappa^n) c^{n-1} + b^n, \qquad \mathbf{A}_n(\kappa) = \mathbf{M}(m_\kappa^n) + \Delta T_n \mathbf{K}(w), \tag{6}$$

where b^n contains the boundary lift.

Proposition 1 (SPD exact-Dupire layer)**.** *If $0 < \kappa_{\min} \le \kappa(T,k) \le \kappa_{\max} < \infty$ on the computational domain, then $\mathbf{M}(m_\kappa^n) \succ 0$ and, after Dirichlet lifting, $\mathbf{K}(w) \succ 0$. Consequently $\mathbf{A}_n(\kappa) \succ 0$ at every time layer.*

The proof is given in Section 4. The lower bound is the hard constraint used by the estimator. A finite upper bound is needed for uniform ellipticity estimates and may be imposed by a scaled sigmoid; the experiments use softplus and monitor the attained range. SPD gives unique discrete states, but it does not by itself identify the coefficient or certify all forms of static no-arbitrage.

4 Mathematical Details for Exact Weighted Dupire

4.1 Weighted transformation and well-posedness

Because $m_\kappa \kappa = e^{-k} = w$, multiplying (2) by m_κ gives

$$m_\kappa \partial_T c = e^{-k}(c_{kk} - c_k) = \partial_k (e^{-k} c_k) = \partial_k (w c_k).$$

After a regular Dirichlet lift, the weak problem is

$$\int_\Omega m_\kappa \partial_T c \, v \,\mathrm{d}k + \int_\Omega w c_k v_k \,\mathrm{d}k = 0.$$

If $0 < \kappa_{\min} \le \kappa \le \kappa_{\max} < \infty$ on a bounded domain, then m_κ is bounded above and below by positive constants and $w = e^{-k}$ is positive and bounded. Standard non-autonomous coercive-form theory therefore gives a unique weak solution after the lift, with continuous dependence on the data.

4.2 Proof of Proposition 1

For every nonzero coefficient vector z representing an interior finite-element function z_h,

$$z^\top \mathbf{M}(m_\kappa^n) z = \int_\Omega m_\kappa^n z_h^2 \,\mathrm{d}k > 0, \qquad z^\top \mathbf{K}(w) z = \int_\Omega w (z_h')^2 \,\mathrm{d}k > 0.$$

The second inequality uses the homogeneous boundary conditions after lifting; a function with zero derivative and zero trace is zero. Hence both matrices are SPD and so is their positive sum $\mathbf{A}_n$. With boundary conditions for which $\mathbf{K}$ is merely positive semidefinite, the positive mass matrix still makes $\mathbf{A}_n$ SPD.

4.3 Proof of Proposition 3

Stack $C = (c^1, \dots, c^{N_T})$ and define

$$F_n(\kappa, C) = \mathbf{A}_n(\kappa)c^n - B_n(\kappa)c^{n-1} - b^n(\kappa).$$

The Jacobian $\partial_C F$ is block lower bidiagonal with SPD diagonal blocks $\mathbf{A}_n$ and subdiagonal blocks $-B_n$. It is therefore invertible. The finite-dimensional implicit-function theorem gives a locally unique C^1 map $C(\kappa)$; differentiating $F = 0$ yields (10). For a scalar loss ℓ, the adjoint recursion is

$$\mathbf{A}_n^\top p^n = \partial_{c^n}\ell + B_{n+1}^\top p^{n+1}, \qquad p^{N_T+1} = 0.$$

Since $\mathbf{A}_n^\top = \mathbf{A}_n$, reverse mode reuses the forward matrix factorisations. The dependence $m_\kappa = e^{-k}/\kappa$ is smooth on every set bounded away from zero; no finite-difference derivative is required.

4.4 Proof of Proposition 2

By the boundary-crossing assumption, there is a sequence $\theta_j \notin \mathcal{A}$ with $\theta_j \to \theta_0$. Continuity and $P(\theta_0) = 0$ imply $P(\theta_j) \to 0$, while continuity of the data loss gives $L_{\text{data}}(\theta_j) \to L_{\text{data}}(\theta_0)$. Thus, for every fixed finite λ, the penalised losses converge to $L_{\text{data}}(\theta_0)$ and eventually lie in the stated sublevel set. Each θ_j is outside $\mathcal{A}$, so some interface coefficient is below $\kappa_{\min}$. Under (8), no such sequence exists in parameter space.

5 Symmetric Benchmark Theory

Auxiliary mechanism studies use

$$\partial_\tau u = \partial_k(\kappa \partial_k u), \qquad a_\kappa(u, v) = \int_\Omega \kappa u_k v_k \, \mathrm{d}k. \tag{7}$$

On a uniform grid, the P_1 stiffness has interface weights $\kappa_{i+1/2} = (\kappa_i + \kappa_{i+1})/2$ and quadratic form

$$v^\top \mathbf{K}(\kappa) v = \sum_i \frac{\kappa_{i+1/2}}{h}(v_{i+1} - v_i)^2.$$

Positive interfaces therefore make $\mathbf{K}$ SPD after Dirichlet lifting. The arithmetic rule makes the assembled stiffness affine in nodal coefficients, so its Jacobian is constant and can be evaluated exactly. Other smooth interface rules can also be differentiable; affinity, not differentiability itself, is the distinguishing convenience.

With lumped mass $\mathbf{M} = hI$, the layer $\mathbf{A}_n = hI + \Delta\tau_n \mathbf{K}$ may remain SPD even when some interfaces are negative, because

$$\mathbf{A}_n \succ 0 \quad \Longleftrightarrow \quad \lambda_{\min}(\mathbf{K}) > -h/\Delta\tau_n.$$

Hence layer SPD is weaker than coefficient admissibility, which motivates separate coefficient-admissibility and layer-spectrum audits.

6 Hard-Constrained Differentiable Calibration

6.1 Positive coefficient and coefficient-to-price map

Let $f_\theta(T, k)$ be a neural field or a free nodal tensor and set

$$\kappa_\theta(T, k) = \kappa_{\min} + \mathrm{softplus}(f_\theta(T, k)), \qquad \kappa_{\min} = \tfrac{1}{2}\sigma_{\min}^2. \tag{8}$$

This makes $\kappa_\theta \geq \kappa_{\min}$ for every parameter value. We define the discrete weighted-Dupire map

$$\mathcal{Q}_D : \kappa \longmapsto (c^1, \dots, c^{N_T}) \tag{9}$$

by the sequence of solves in (6). The state is not an independent network output: it is the solution of the assembled PDE layer.

Property	Enforcement
$\kappa_\theta \geq \kappa_{\min}$	hard parameterisation (8)
SPD state layer	weighted Galerkin construction
discrete PDE satisfaction	implicit solve (6)
coefficient smoothness	statistical regularisation in (11)
density and static-arbitrage shape	sampled-grid diagnostics unless proved

Table 1: Architectural guarantees and statistical or empirical properties are kept separate.

6.2 Why finite soft penalties are different

Proposition 2 (Finite penalties do not certify admissibility)**.** *Let $\theta \mapsto \kappa_\theta$ be continuous in its interface values and let*

$$\mathcal{A} = \{\theta : \kappa_{\theta,i+1/2} \geq \kappa_{\min} \text{ for every } i\}.$$

Suppose $P \geq 0$ is continuous with $P^{-1}(0) = \mathcal{A}$, the data loss is continuous, and the parameterisation can cross the boundary of $\mathcal{A}$ at some $\theta_0 \in \partial\mathcal{A}$. Then, for every finite λ and every $\varepsilon > 0$, the sublevel set

$$\{\theta : L_{\text{data}}(\theta) + \lambda P(\theta) \leq L_{\text{data}}(\theta_0) + \varepsilon\}$$

contains a parameter outside $\mathcal{A}$, hence one with $\min_i \kappa_{\theta,i+1/2} < \kappa_{\min}$. In contrast, (8) excludes such parameters from the model class.

A finite penalty can make violations costly but cannot remove them from parameter space. The proposition does *not* claim that every boundary crossing makes a coefficient negative: it guarantees violation of the stated ellipticity threshold. The hard-versus-soft study records the stronger empirical event that several soft fits cross zero.

6.3 Differentiation and adjoints

Write $B_n(\kappa) = \mathbf{M}(m_\kappa^n)$ and $\mathbf{A}_n = B_n + \Delta T_n \mathbf{K}(w)$. For a coefficient perturbation $\delta\kappa$, differentiating the nth layer gives

$$\mathbf{A}_n Dc^n[\delta\kappa] = DB_n[\delta\kappa](c^{n-1} - c^n) + B_n Dc^{n-1}[\delta\kappa] + Db^n[\delta\kappa]. \tag{10}$$

Proposition 3 (Discrete differentiability)**.** *On any open coefficient set where $\kappa > 0$ and the layer data depend continuously differentiably on κ, the map $\mathcal{Q}_D$ is continuously Fréchet differentiable. Its directional derivative is obtained from (10); reverse mode solves the transposed layer systems in reverse time and therefore reuses the same SPD factorisations as the forward pass.*

Appendix 4 proves the statement using the block lower-triangular state Jacobian. Invertibility guarantees a well-defined derivative; gradient conditioning still depends on the mesh, maturity steps, and coefficient contrast.

7 Calibration Objective

Market implied volatilities are first converted to option prices with the market's Black–76 convention and then normalised as above. For observed nodes (T_j, k_j), we minimise

$$L(\theta) = \underbrace{\sum_j \omega_j \left(c_\theta(T_j, k_j) - c_j^{\text{mkt}}\right)^2}_{L_{\text{data}}} + \mu \underbrace{\left(\|\partial_{kk}\kappa_\theta\|_{L^2}^2 + \eta_T \|\partial_{TT}\kappa_\theta\|_{L^2}^2\right)}_{R(\kappa_\theta)}. \tag{11}$$

Weights ω_j reduce the influence of illiquid deep-wing quotes. Regularisation controls the bias–variance trade-off of the inverse problem; it does not enforce the PDE. Energy identities and linear-system residuals are reported only as solver diagnostics.

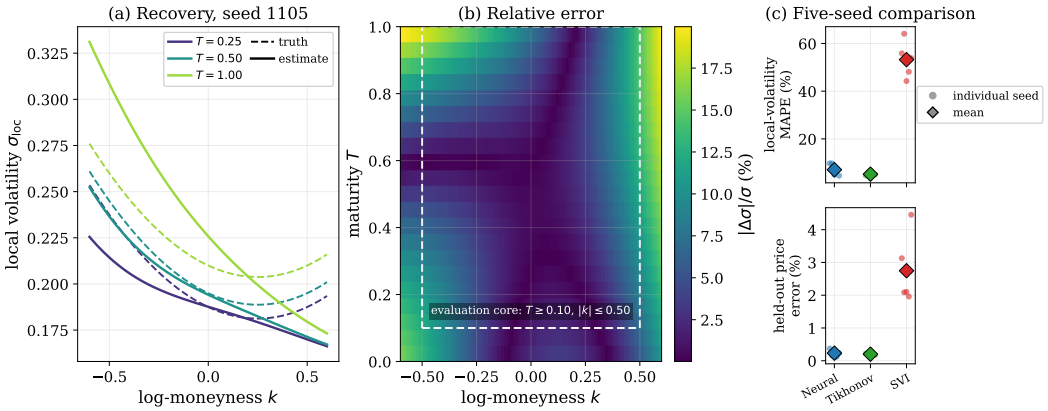

Figure 2: Synthetic coefficient recovery with exact weighted Dupire. Left: recovered and true local-volatility slices at three maturities; centre: relative coefficient-error map; right: total-loss and data-misfit optimisation histories. Aggregate statistics use all five stored noise realisations.

8 Experiments

Scope and protocol. The experiments test distinct claims and should not be pooled into a global ranking. Synthetic coefficient recovery evaluates a correctly specified inverse problem. The study of hard constraints versus soft physics penalties isolates the operational consequences of architectural and finite-penalty enforcement, but its soft baselines also use independently learned state networks. Out-of-sample S&P 500 calibration evaluates market prediction. The synthetic recovery and market-calibration studies use the exact weighted-Dupire map $\mathcal{Q}_D$; the hard-versus-soft study uses the symmetric benchmark.

8.1 Synthetic coefficient recovery

The clean surface is generated on a 243×83 reference grid and subsampled onto a distinct 81×21 inversion grid, avoiding an identical-grid inverse crime. We add 1% noise and use an 80/20 observation split. Across five fixed noise realisations, the neural field attains coefficient MAPE 7.1% with bootstrap interval $[5.1\%, 9.1\%]$ over 209 observations (152 training and 57 held out), and held-out price relative RMSE 0.23%. The neural parameterisation does not dominate classical regularisation: a nodal Tikhonov estimator achieves 5.2% coefficient MAPE and 0.20% held-out price error. SVI followed by numerical Dupire differentiation yields 53.2% mean coefficient MAPE. The synthetic recovery study therefore validates the exact-Dupire differentiation and recovery pipeline without claiming neural superiority.

8.2 Hard constraints versus soft physics penalties

All methods use the same coefficient-network family, observations, optimiser, and step budget. The hard branch maps its positive coefficient through the Galerkin solver. PINN, Deep Ritz, and VPINN baselines additionally learn a state network and use, respectively, strong-form, incremental variational, and weak-form physics objectives. Each soft physics weight is selected on an independent pilot seed and frozen before evaluation. This is a mechanism-level comparison, not a controlled ablation of penalty placement alone.

Across 5 evaluation seeds, the hard model keeps $\min \kappa \geq 1.35 \times 10^{-2}$ with no negative interfaces. The soft fits cross zero: their mean minima are -1.7×10^{-3} (PINN), -2.0×10^{-3} (Deep Ritz), and -9.0×10^{-4} (VPINN), with negative-interface fractions 6.9%, 55.4%, and 8.8%. The theorem only predicts possible violation of $\kappa_{\min}$; the observed negative values are a stronger empirical failure.

Price fit does not diagnose coefficient admissibility. The soft methods obtain relative price RMSE 1.4%, 1.4%, and 1.4%, compared with 0.58% for the hard method, yet their coefficient MAPEs are 40.3%, 80.9%, and 50.5%, compared with 4.3%. A separate layerwise profile finds 5 indefinite layers

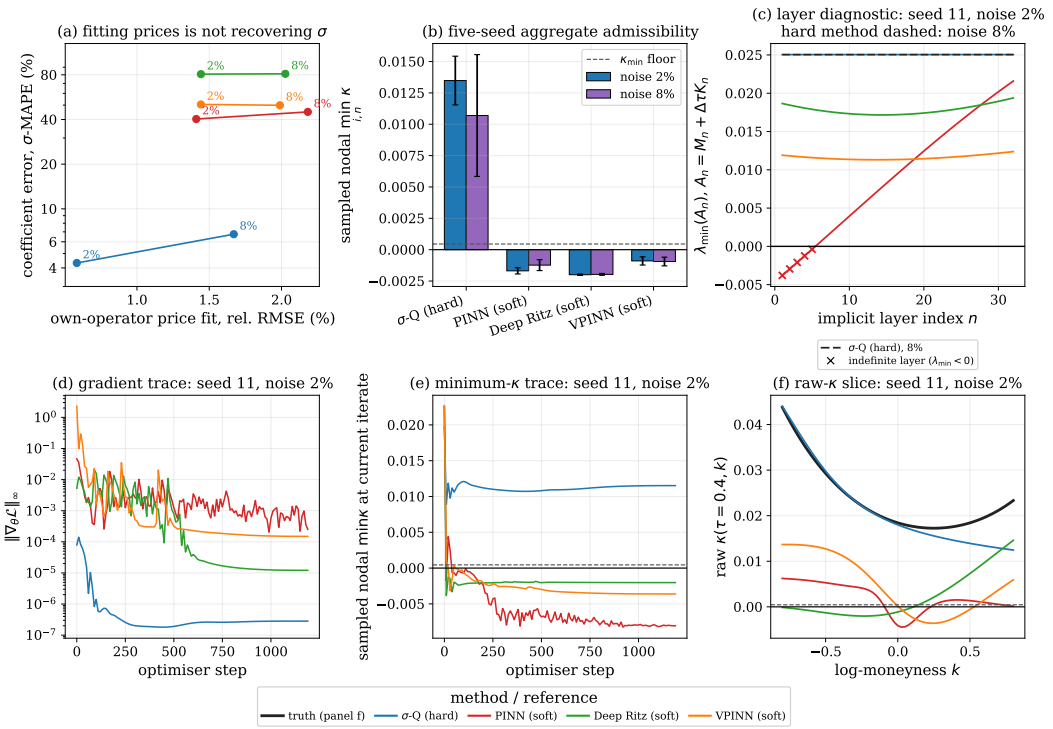

Figure 3: Extended hard-versus-soft diagnostics. (a) Own-operator price fit versus coefficient recovery; (b) minimum recovered coefficient; (c) smallest eigenvalue across implicit layers, with crosses marking indefinite layers; (d) gradient magnitude during optimisation; (e) minimum coefficient along the optimisation path; (f) a recovered volatility slice against truth. Aggregate coefficient statistics use five seeds, whereas the layer spectrum, gradient trace, optimisation path, and slice are seed-level stress diagnostics.

out of 32 for one PINN fit, with worst eigenvalue -3.8×10^{-3}; other sign-violating fits remain SPD because the positive mass term masks modest coefficient violations. We therefore audit coefficient admissibility and layer spectra separately.

As an operational transfer test, each learned coefficient is inserted into the same Galerkin propagator. The hard model's transfer error equals its own fit error by construction. At 8% noise, PINN and VPINN transfer errors reach 6.0×10^{-2} and 6.0×10^{-2}, while Deep Ritz reaches 28.1% already at 2% noise. Agreement between the differentiable tridiagonal layer and a dense factorisation is 2.8×10^{-16} in relative sup-norm.

8.3 Out-of-sample S&P 500 calibration

The delayed CBOE snapshot contains 16,880 clean out-of-the-money quotes across 34 expiries, aggregated to 1,534 price nodes after grid snapping. We compare exact Dupire with monotone SSVI under held-out maturity blocks and contiguous held-out strike blocks. Elementary squared errors and counts are pooled before taking the root; the two split families are reported separately because they test different extrapolation tasks.

A final all-data fit is used only for shape inspection. Its liquid-core IV RMSE is 0.64 volatility points; sampled-grid flux-density negativity and calendar-violation fractions are 0.0% and 0.0%. These diagnostics are not used for model selection and are not a complete static-arbitrage certificate.

8.4 Neural-Operator Generalisation

The neural-operator study compares the symmetric Galerkin map $\mathcal{Q}_S$ with FNO, DeepONet, and a graph operator on smooth random coefficients. In distribution, FNO is more accurate (1.93×10^{-3}

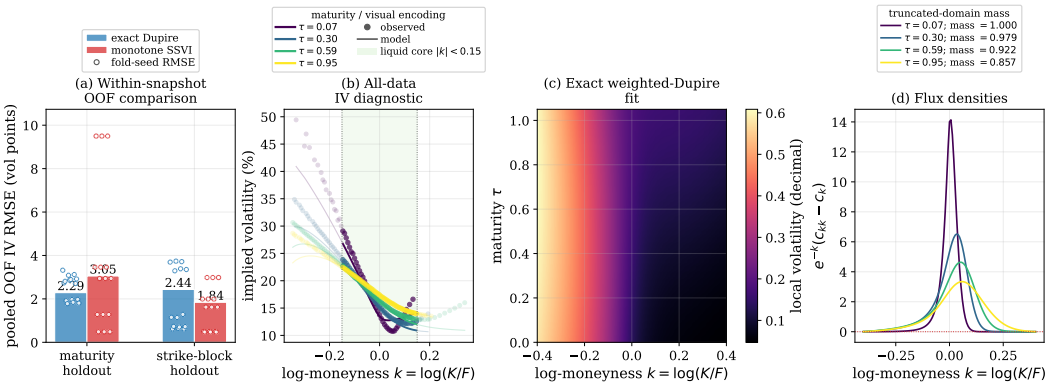

Figure 4: Out-of-sample S&P 500 calibration. Top left: fitted local-volatility surface; top right: all-data implied-volatility fit shown diagnostically; bottom left: recovered flux densities; bottom right: leakage-safe maturity and strike-block holdout errors for exact Dupire and monotone SSVI.

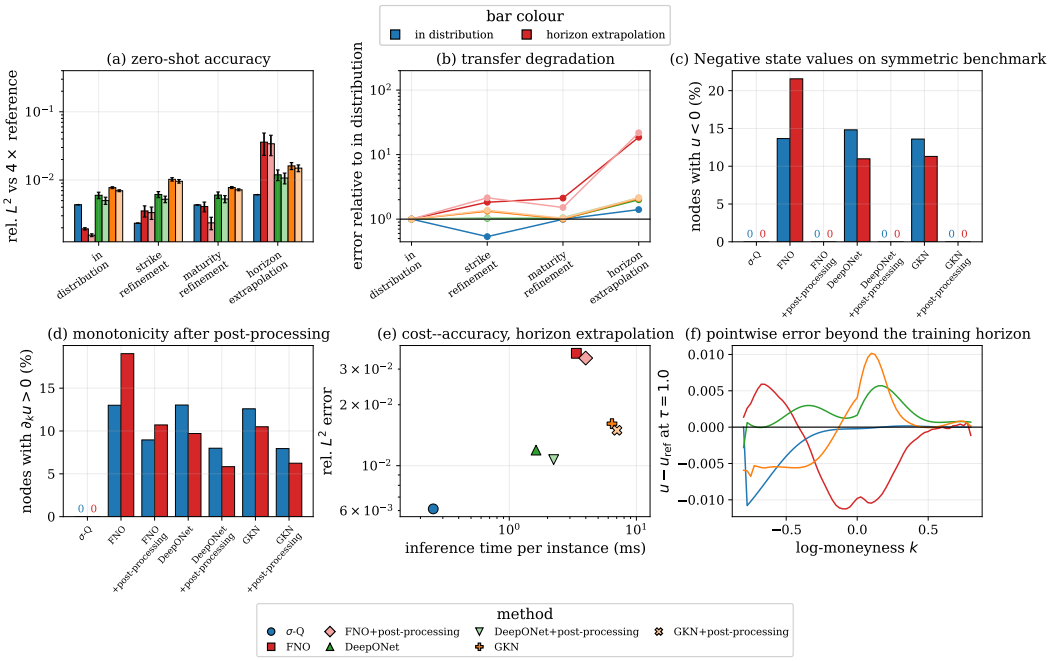

Figure 5: Neural-operator generalisation. (a) Zero-shot relative error across in-distribution, strike-refinement, maturity-refinement, and horizon-shift regimes; (b) degradation relative to the in-distribution regime; (c) negative state values; (d) sampled-grid monotonicity violations after optional SPD repair; (e) horizon-shift cost–accuracy trade-off; (f) pointwise error beyond the training horizon.

versus 4.36×10^{-3} relative L^2 error), consistent with learning a correction to coarse-grid discretisation error. Under horizon shift, the ordering reverses: the Galerkin error is 6.12×10^{-3} and the FNO error is 3.59×10^{-2}, while the FNO is negative on 21.6% of nodes and violates sampled-grid monotonicity on 19.0%. The comparison supports robustness under this declared shift, not universal superiority or equal-compute convergence.

8.5 Exact-Dupire dual-density and score study

For a sufficiently regular exact-Dupire solution, define

$$q_T(k) = e^{-k}(c_{kk} - c_k) = \partial_k(e^{-k}c_k). \tag{12}$$

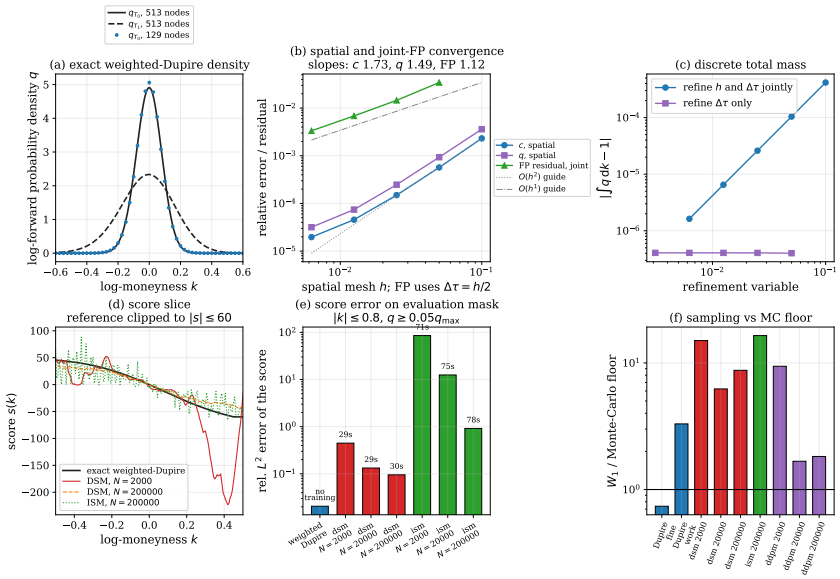

Figure 6: Exact-Dupire dual-density and score study using $\mathcal{Q}_D$. (a) Dual densities on working and reference grids; (b) price, density, and Fokker–Planck convergence; (c) discrete mass error; (d) operator-derived and learned scores; (e) score error and training cost; (f) reverse-time sampling quality relative to the Monte Carlo floor. Convergence includes constant-κ analytic checks, while the Fokker–Planck, score, and sampling experiments use variable κ.

Breeden–Litzenberger implies that q_T is the density of $X_T = \log(S_T/F_T)$ under the T-forward measure. Under the usual tail conditions, $\int_{\mathbb{R}} q_T \, \mathrm{d}k = 1$. Differentiating (12) and using (2) gives

$$\partial_T q = \partial_k(\kappa q) + \partial_{kk}(\kappa q), \tag{13}$$

the Fokker–Planck equation of

$$\mathrm{d}X_T = -\kappa(T, X_T)\,\mathrm{d}T + \sqrt{2\kappa(T, X_T)}\,\mathrm{d}W_T. \tag{14}$$

The implementation uses the exact weighted-Dupire operator $\mathcal{Q}_D$. Its convergence study includes constant-κ checks against the closed-form Black–76 price and Gaussian log-forward density, with separate spatial and temporal refinements. The variable-κ experiments test the Fokker–Planck residual in (13), compare the operator-derived score with denoising and implicit score matching, and assess forward and reverse-time sampling together with a DDPM comparator.

9 Conclusion

We formulated local-volatility calibration as a differentiable weighted-Galerkin inverse problem in which coefficient admissibility and SPD solvability are properties of the architecture rather than finite penalties. The coefficient-to-price map supports end-to-end adjoint calibration while leaving smoothness as a separate statistical regulariser. Synthetic, hard-versus-soft, and S&P 500 experiments show that these guarantees can be retained without assuming that they imply universal recovery accuracy or full static no-arbitrage.

LLM USE DISCLOSURE

We used generative AI tools, including ChatGPT and Qoder(`https://qoder.com/`), for research implementation, mathematical review, and language editing. The authors reviewed all AI-assisted material and take responsibility for the final mathematical claims, experimental design, numerical results, software, and manuscript.

Research implementation and execution. Qoder was used to assist with code completion, refactoring, debugging, and the coarse-to-fine implementation of the numerical experiments reported in Section 8 and in the additional experimental appendices. In particular, AI-assisted software development was used in implementing and checking components of the exact weighted-Dupire solver $\mathcal{Q}_D$, the auxiliary symmetric solver $\mathcal{Q}_S$, and the experiment pipelines used for synthetic recovery, hard-versus-soft comparisons, neural-operator comparisons, density and score experiments, and uncertainty quantification. The authors designed the experimental protocols, selected the mathematical models, baselines, data splits, evaluation metrics, and validation procedures, and implemented and checked the initial proof-of-concept methodology. AI tools did not generate the reported numerical results.

Mathematical claims and proofs. ChatGPT was used as an auxiliary mathematical-review and drafting tool for arguments in Section 3, Section 6, and the corresponding mathematical appendices. This included checking the weighted-Dupire transformation, the SPD-layer argument, the discrete parameter-to-state differentiation argument, and the finite-penalty admissibility result. In particular, AI-assisted review identified an over-strong earlier formulation of Proposition 2: crossing the boundary of the admissible set guarantees violation of the prescribed lower ellipticity threshold $\kappa_{\min}$, but does not by itself imply that the recovered coefficient becomes negative. The theorem statement and proof were consequently revised to distinguish these two claims. AI tools were also used to improve the exposition and organisation of the proofs. All final propositions, assumptions, derivations, and proofs were independently checked and accepted by the authors.

Data and reported results. Generative AI tools were not used to generate the synthetic datasets or market data used in this work. Synthetic observations were generated numerically from the mathematical models described in the paper, using either the exact weighted-Dupire operator $\mathcal{Q}_D$ or the explicitly identified symmetric benchmark $\mathcal{Q}_S$. The market-calibration study uses CBOE S&P 500 option quotes processed according to the data pipeline described in the Market Data and Reproducibility appendix. AI tools were not used to fabricate, modify, or replace stored experimental measurements. AI-assisted interpretations and manuscript descriptions of experimental results were checked against the corresponding numerical outputs by the authors.

We reviewed all AI-assisted mathematical text, manuscript text, and software changes. We take responsibility for the final content of this work, including all text, claims, proofs, numerical results, software, and artifacts produced with the aid of generative AI.

ETHICS STATEMENT

This work studies numerical and machine-learning methods for local-volatility calibration and does not involve human-subject experiments, interventions, surveys, or individual-level personal data. The synthetic experiments use observations generated numerically from the mathematical models specified in the paper. The empirical market study uses delayed CBOE S&P 500 option quotes and reports only aggregate market quantities; the associated filtering, forward estimation, normalisation, and train/holdout procedures are described in the Market Data and Reproducibility appendix.

The proposed methodology is intended for research on derivative-pricing and inverse-PDE problems. If deployed in financial applications, calibration error, model misspecification, unstable extrapolation, or misuse of the reported structural guarantees could lead to inaccurate valuation or risk estimates. We therefore distinguish coefficient admissibility and SPD solvability from coefficient identifiability, predictive accuracy, and complete static no-arbitrage. Out-of-sample market errors and sampled-grid shape diagnostics are reported separately, and the paper does not present the method or its empirical results as investment, portfolio, or trading advice.

REFERENCES

Brian DO Anderson. Reverse-time diffusion equation models. *Stochastic Processes and their Applications*, 12(3):313–326, 1982.

Douglas T Breeden and Robert H Litzenberger. Prices of state-contingent claims implicit in option prices. *Journal of business*, pp. 621–651, 1978.

Susanne C Brenner and L Ridgway Scott. *The mathematical theory of finite element methods*. Springer, 2008.

Simon L Cotter, Gareth O Roberts, Andrew M Stuart, and David White. Mcmc methods for functions: modifying old algorithms to make them faster. *Statistical Science*, pp. 424–446, 2013.

Stephane Crepey. Calibration of the local volatility in a generalized black–scholes model using tikhonov regularization. *SIAM Journal on Mathematical Analysis*, 34(5):1183–1206, 2003.

Erik Daxberger, Agustinus Kristiadi, Alexander Immer, Runa Eschenhagen, Matthias Bauer, and Philipp Hennig. Laplace redux-effortless bayesian deep learning. volume 34, pp. 20089–20103, 2021.

Emanuel Derman and Iraj Kani. Riding on a smile. *Risk*, 7(2):32–39, 1994.

Shaan A Desai, Marios Mattheakis, David Sondak, Pavlos Protopapas, and Stephen J Roberts. Port-hamiltonian neural networks for learning explicit time-dependent dynamical systems. *Physical Review E*, 104(3):034312, 2021.

Bruno Dupire. Pricing with a smile. *Risk*, 7(1):18–20, January 1994.

Jim Gatheral. A parsimonious arbitrage-free implied volatility parameterization with application to the valuation of volatility derivatives, 2004.

Jim Gatheral and Antoine Jacquier. Arbitrage-free svi volatility surfaces. *Quantitative Finance*, 14 (1):59–71, 2014.

Andrew Gelman and Donald B Rubin. Inference from iterative simulation using multiple sequences. *Statistical science*, 7(4):457–472, 1992.

Charles J Geyer. Practical markov chain monte carlo. *Statistical science*, pp. 473–483, 1992.

Samuel Greydanus, Misko Dzamba, and Jason Yosinski. Hamiltonian neural networks. volume 32, 2019.

Michael Hinze, René Pinnau, Michael Ulbrich, and Stefan Ulbrich. *Optimization with PDE constraints*. Springer Science & Business Media, 2008.

Jonathan Ho, Ajay Jain, and Pieter Abbeel. Denoising diffusion probabilistic models. volume 33, pp. 6840–6851, 2020.

Aapo Hyvärinen and Peter Dayan. Estimation of non-normalized statistical models by score matching. *Journal of Machine Learning Research*, 6(4):695, 2005.

Jari P Kaipio and Erkki Somersalo. *Statistical and computational inverse problems*. Springer, 2005.

Ehsan Kharazmi, Zhongqiang Zhang, and George Em Karniadakis. hp-vpinns: Variational physics-informed neural networks with domain decomposition. *Computer Methods in Applied Mechanics and Engineering*, 374:113547, 2021.

Nikola Kovachki, Zongyi Li, Burigede Liu, Kamyar Azizzadenesheli, Kaushik Bhattacharya, Andrew Stuart, and Anima Anandkumar. Neural operator: Learning maps between function spaces with applications to pdes. *Journal of Machine Learning Research*, 24(89):1–97, 2023.

Zongyi Li, Nikola Kovachki, Kamyar Azizzadenesheli, Burigede Liu, Kaushik Bhattacharya, Andrew Stuart, and Anima Anandkumar. Fourier neural operator for parametric partial differential equations. 2020.

Lu Lu, Pengzhan Jin, Guofei Pang, Zhongqiang Zhang, and George Em Karniadakis. Learning nonlinear operators via deeponet based on the universal approximation theorem of operators. *Nature machine intelligence*, 3(3):218–229, 2021.

Michael Lutter, Christian Ritter, and Jan Peters. Deep lagrangian networks: Using physics as model prior for deep learning. 2019.

Apostolos F Psaros, Xuhui Meng, Zongren Zou, Ling Guo, and George Em Karniadakis. Uncertainty quantification in scientific machine learning: Methods, metrics, and comparisons. *Journal of Computational Physics*, 477:111902, 2023.

Maziar Raissi, Paris Perdikaris, and George E Karniadakis. Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations. *Journal of Computational physics*, 378:686–707, 2019.

Yang Song, Jascha Sohl-Dickstein, Diederik P Kingma, Abhishek Kumar, Stefano Ermon, and Ben Poole. Score-based generative modeling through stochastic differential equations. 2020.

Andrew M Stuart. Inverse problems: a bayesian perspective. *Acta numerica*, 19:451–559, 2010.

Robert J Tibshirani and Bradley Efron. *An introduction to the bootstrap*, volume 57. 1993.

Pascal Vincent. A connection between score matching and denoising autoencoders. *Neural computation*, 23(7):1661–1674, 2011.

Liu Yang, Xuhui Meng, and George Em Karniadakis. B-pinns: Bayesian physics-informed neural networks for forward and inverse pde problems with noisy data. *Journal of Computational Physics*, 425:109913, 2021.

Bing Yu et al. The deep ritz method: a deep learning-based numerical algorithm for solving variational problems. *Communications in Mathematics and Statistics*, 6(1):1–12, 2018.

A Mathematical Details for Exact Weighted Dupire

A.1 Continuous setting and weighted form

Let the risk-neutral asset satisfy

$$\mathrm{d}S_t = (r-q)S_t\,\mathrm{d}t + \sigma_{\mathrm{loc}}(t,S_t)S_t\,\mathrm{d}W_t. \tag{15}$$

The change to normalized forward calls gives (2); the first-derivative term is essential and is not removed by a numeraire change. Multiplication by $m_\kappa = e^{-k}/\kappa$ gives (4), whose spatial bilinear form is $a(c,v) = \int_\Omega e^{-k}c_k v_k\,\mathrm{d}k$ after a Dirichlet lift.

Proposition 4 (Parabolic well-posedness). *Suppose $\kappa \in W^{1,\infty}(0,T;L^\infty(\Omega))$ and $0 < \kappa_{\min} \le \kappa \le \kappa_{\max} < \infty$. After a sufficiently regular boundary lift, the weighted mass operator is uniformly positive and Lipschitz in time, while the spatial form is continuous and coercive on $H_0^1(\Omega)$. Standard non-autonomous parabolic theory therefore gives a unique weak solution in $L^2(0,T;H_0^1(\Omega)) \cap H^1(0,T;H^{-1}(\Omega))$, continuously dependent on the data.*

A.2 Proof of weighted-layer SPD

For any nonzero interior vector v,

$$v^\top \mathbf{A}_n v = v^\top \mathbf{M}(m_\kappa^n)v + \Delta T_n \int_\Omega e^{-k}|(v_h)_k|^2\,\mathrm{d}k > 0.$$

The weighted mass is positive-definite because $m_\kappa^n > 0$, and the stiffness is positive semidefinite (positive-definite after homogeneous Dirichlet restriction). Symmetry follows from both bilinear forms.

A.3 Full differentiability proof

Stack $C = (c^1,\dots,c^{N_T})$ and write $F_n(\kappa,C) = \mathbf{A}_n(\kappa)c^n - b_n(\kappa,c^{n-1})$. The state Jacobian $\partial_C F$ is block lower-bidiagonal with SPD diagonal blocks $\mathbf{A}_n$ and is therefore invertible. The finite-dimensional implicit function theorem gives

$$DC[\delta\kappa] = -(\partial_C F)^{-1}\partial_\kappa F[\delta\kappa],$$

whose forward substitution is (10). For a scalar loss ℓ, the reverse recursion is

$$\mathbf{A}_n^\top p^n = \partial_{c^n}\ell + \left(\partial_{c^n} b_{n+1}\right)^\top p^{n+1}, \qquad p^{N_T+1} = 0.$$

Because $\mathbf{A}_n^\top = \mathbf{A}_n$, the adjoint uses the same SPD factorisations. Here derivatives of the weighted mass include $Dm_\kappa[\delta\kappa] = -e^{-k}\delta\kappa/\kappa^2$; this is an exact analytic Jacobian, not a finite-difference approximation. Continuous parameter-to-state differentiability requires the usual function-space assumptions Hinze et al. (2008).

A.4 Dual density, Fokker–Planck flow, and reverse time

Breeden–Litzenberger gives $\partial_{KK}C(T,K) = D(0,T)f_{S_T}(K)$ in strike coordinates Breeden & Litzenberger (1978). For the normalised forward call $c(T,k) = C(T,F_Te^k)/(D(0,T)F_T)$ defined before (2), the corresponding object is the density of the log-forward return, and it satisfies a conservative forward equation.

Proposition 5 (Dual density, flux form, and the forward equation). *Let $c(T,\cdot) \in C^2$ solve the exact log-moneyness Dupire equation (2) with $c(T,k) \to 0$ and $\partial_k c(T,k) \to 0$ as $k \to +\infty$, and $c(T,k) - (1-e^k) \to 0$ as $k \to -\infty$. Define*

$$q_T(k) = e^{-k}\big(\partial_{kk}c(T,k) - \partial_k c(T,k)\big). \tag{16}$$

Then

(i) q_T is the density of $X_T = \log(S_T/F_T)$ under the T-forward measure, and it admits the divergence form $q_T = \partial_k\big(e^{-k}\partial_k c\big)$;

(ii) $\int_{\mathbb{R}} q_T(k)\,\mathrm{d}k = 1$ *for every* T*;*

(iii) q *solves the conservative forward equation*

$$\partial_T q = \partial_k(\kappa q) + \partial_{kk}(\kappa q), \tag{17}$$

which is exactly the Fokker–Planck equation of the log-forward diffusion

$$\mathrm{d}X_T = -\kappa(T, X_T)\,\mathrm{d}T + \sqrt{2\,\kappa(T, X_T)}\,\mathrm{d}W_T. \tag{18}$$

Proof. (i) With $K = F_T e^k$ and $c = C/(DF_T)$ one has $\partial_K = F_T^{-1} e^{-k}\partial_k$, hence

$$\partial_{KK} C = D(0,T)\,F_T^{-1} e^{-2k}\big(\partial_{kk} c - \partial_k c\big).$$

Since the density of X_T is $f_{S_T}(K)\,\mathrm{d}K/\mathrm{d}k = F_T e^k f_{S_T}(K)$, combining with Breeden–Litzenberger gives precisely (16). The divergence form follows from $\partial_k(e^{-k}\partial_k c) = e^{-k}(\partial_{kk}c - \partial_k c)$.

(ii) By (i), $\int_{-R}^{R} q_T = \big[e^{-k}\partial_k c\big]_{-R}^{R}$. As $k \to +\infty$, $\partial_k c \to 0$ faster than e^k grows for a call with finite price, so the upper term vanishes; as $k \to -\infty$, $c \to 1 - e^k$ gives $\partial_k c \to -e^k$ and $e^{-k}\partial_k c \to -1$. Hence the bracket tends to $0 - (-1) = 1$.

(iii) Differentiate (16) in T and substitute (2). Writing $\mathcal{L}c = \partial_{kk}c - \partial_k c$ and noting that $\mathcal{L}$ commutes with ∂_T,

$$\partial_T q = e^{-k}\mathcal{L}(\partial_T c) = e^{-k}\mathcal{L}(\kappa\,\mathcal{L}c) = e^{-k}\mathcal{L}\big(e^k \kappa q\big).$$

Put $g = \kappa q$. Then $e^{-k}\mathcal{L}(e^k g) = e^{-k}\big(\partial_{kk}(e^k g) - \partial_k(e^k g)\big) = e^{-k}\big(e^k g + 2e^k\partial_k g + e^k\partial_{kk} g - e^k g - e^k\partial_k g\big) = \partial_k g + \partial_{kk} g$, which is (17). Finally, the Fokker–Planck equation of $\mathrm{d}X = b\,\mathrm{d}T + \varsigma\,\mathrm{d}W$ is $\partial_T q = -\partial_k(bq) + \frac{1}{2}\partial_{kk}(\varsigma^2 q)$; matching $b = -\kappa$ and $\varsigma^2 = 2\kappa$ gives (18). The drift $-\kappa = -\frac{1}{2}\sigma_{\mathrm{loc}}^2$ is the Itô correction that keeps $\mathbb{E}[e^{X_T}] = 1$, i.e. the forward-martingale property. □

A P_1 finite-element function has no classical piecewise second derivative: its distributional second derivative is concentrated at nodes. The density recovery study in Section C.1 therefore recovers the financial density through the exact weighted-Dupire flux $q = e^{-k}(c_{kk} - c_k)$ on the output grid rather than interpreting $\partial_{kk} c_h$ as a classical finite-element field. Under input-matched noise, the correctly specified PDE fit has empirical mesh slope 0.64 with bootstrap interval $[0.60, 0.67]$, whereas applying the same differentiation directly to noisy quotes amplifies noise. We report this as a controlled empirical mesh study. A generic $O(h^2)$ theorem would require additional regularity and a dedicated mixed, flux-recovery, or dual Fokker–Planck discretisation. Non-negativity of the recovered density remains a diagnostic; SPD of the energy operator alone does not prove convexity preservation or complete static no-arbitrage.

B Hard Constraints versus Soft Physics Penalties and Symmetric Benchmark Theory

B.1 Auxiliary symmetric operator

The symmetric benchmark is

$$\partial_\tau u = \partial_k(\kappa(\tau, k)\partial_k u), \tag{19}$$

with bilinear form

$$a_\kappa(u, v) = \int_\Omega \kappa u_k v_k\,\mathrm{d}k. \tag{20}$$

On a uniform P_1 mesh,

$$\mathbf{K}(\kappa)_{ii} = \frac{\kappa_{i-1/2} + \kappa_{i+1/2}}{h}, \qquad \mathbf{K}(\kappa)_{i,i\pm1} = -\frac{\kappa_{i\pm1/2}}{h}, \tag{21}$$

with arithmetic interfaces $\kappa_{i+1/2} = \frac{1}{2}(\kappa_i + \kappa_{i+1})$. This choice makes the assembled stiffness affine in nodal coefficients, so its Jacobian is exact and constant. It is not the only differentiable interface rule. Backward Euler defines

$$(\mathbf{M} + \Delta\tau_n \mathbf{K}(\kappa^n))u^n = \mathbf{M}u^{n-1} + (\text{Dirichlet lift}), \tag{22}$$

and the corresponding map is denoted $\mathcal{Q}_S$, not $\mathcal{Q}_D$.

Proposition 6 (Positive interfaces imply SPD). *If $\kappa_{i+1/2} \geq \kappa_{\min} > 0$, then the interior $\mathbf{K}(\kappa)$ is SPD.*

Proof. For a nonzero interior vector extended by homogeneous boundary values,

$$v^\top \mathbf{K} v = \sum_i \frac{\kappa_{i+1/2}}{h}(v_{i+1} - v_i)^2 > 0.$$

□

B.2 Proof and interpretation of the finite-penalty no-go

Let $\theta_0 \in \partial\mathcal{A}$. Local boundary crossing supplies $\theta_j \to \theta_0$ with $\min_i \kappa_{i+1/2}(\theta_j) < \kappa_{\min}$; it need not supply a negative coefficient. Continuity gives $P(\theta_j) \to 0$ and $L_{\text{data}}(\theta_j) \to L_{\text{data}}(\theta_0)$, hence $L_\lambda(\theta_j) \to L_\lambda(\theta_0)$ for every finite λ. Thus arbitrary objective-value neighbourhoods of the boundary contain points outside $\mathcal{A}$. In contrast, (8) maps every parameter into $\mathcal{A}$. This proves Proposition 2; it says that finite penalties do not certify the lower bound, not that all violating fits are negative.

Remark 1. A soft penalty may be useful statistically, but admissibility must be audited after training. The hard-versus-soft study therefore reports the coefficient minimum, per-layer spectra, factorisation outcomes, and sampled-grid diagnostics separately.

Proposition 7 (Quantitative symmetric-layer spectrum). *Let $\mathbf{M} = h\mathbf{I}$ and $\mathbf{L} = \text{tridiag}(-1, 2, -1)$. If $\kappa_{\min} \leq \kappa_{i+1/2} \leq \kappa_{\max}$, then*

$$h + \frac{\Delta\tau_n \kappa_{\min}}{h}\lambda_{\min}(\mathbf{L}) \leq \lambda_{\min}(\mathbf{A}_n), \qquad \lambda_{\max}(\mathbf{A}_n) \leq h + \frac{4\Delta\tau_n \kappa_{\max}}{h}, \tag{23}$$

and

$$\text{cond}(\mathbf{A}_n) \leq 1 + \frac{4\Delta\tau_n \kappa_{\max}}{h^2}. \tag{24}$$

Proof. The quadratic form is bounded above and below by $\kappa_{\max} h^{-1} v^\top \mathbf{L} v$ and $\kappa_{\min} h^{-1} v^\top \mathbf{L} v$. The eigenvalues of $\mathbf{L}$ lie between $4\sin^2(\pi/(2N-2))$ and $4\cos^2(\pi/(2N-2)) < 4$; adding $h\mathbf{I}$ proves the bounds. □

Corollary 1 (Layer SPD is weaker than coefficient positivity). *For a fixed coefficient vector, $\mathbf{A}_n = h\mathbf{I} + \Delta\tau_n \mathbf{K}$ is SPD iff $\lambda_{\min}(\mathbf{K}) > -h/\Delta\tau_n$. A layer may therefore remain SPD even when some interface coefficient violates its lower bound.*

B.3 Energy audit and obstacle extension

For homogeneous lifted data, the symmetric benchmark satisfies

$$\tfrac{1}{2}\frac{\mathrm{d}}{\mathrm{d}\tau}\|u\|_{L^2}^2 + u^\top \mathbf{K}(\kappa) u = 0, \tag{25}$$

and backward Euler satisfies

$$\frac{\|u^n\|_{\mathbf{M}}^2 - \|u^{n-1}\|_{\mathbf{M}}^2 + \|u^n - u^{n-1}\|_{\mathbf{M}}^2}{2\Delta\tau_n} + (u^n)^\top \mathbf{K}(\kappa^n) u^n = 0. \tag{26}$$

Its residual is an audit, never a training term. American options lead to $\min_{u \geq \Phi} \frac{1}{2} u^\top \mathbf{A} u - b^\top u$ with $\mathbf{A} = \mathbf{M} + \Delta\tau(\mathbf{K} + r\mathbf{M})$; SPD makes this obstacle problem strictly convex. Fixed-active-set derivatives are local and require strict complementarity.

C Additional Calibration Experiments

C.1 Density recovery

For the exact-Dupire solution, the log-forward density is $q = e^{-k}(c_{kk} - c_k)$. Under identical additive price noise 10^{-4} across nested grids, the fitted Dupire estimator reaches finest-grid core RMSE 2.39×10^{-3} and empirical mesh slope 0.64 with interval $[0.60, 0.67]$. Tikhonov smoothing reaches 3.23×10^{-2}; raw finite differences deteriorate to 8.55 and are negative on 45.5% of core nodes. This is a controlled correctly specified study, not a generic second-derivative theorem for P_1 elements.

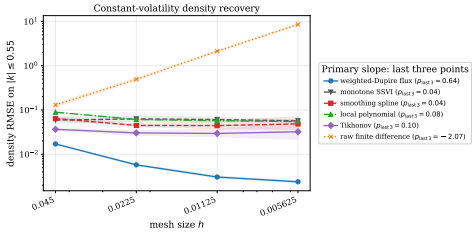

Figure 7: Density recovery from identically noise-perturbed option prices. Core-density RMSE is plotted against mesh size for exact weighted-Dupire flux recovery, monotone SSVI, smoothing splines, local polynomials, Tikhonov smoothing, and raw finite differences; shaded bands show uncertainty across realisations and legend entries report empirical slopes.

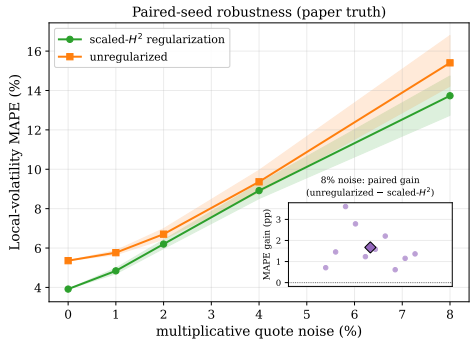

Figure 8: Noise robustness on the symmetric benchmark. Coefficient MAPE is shown against multiplicative quote noise for regularised and unregularised fits; curves compare paired-seed means and shaded bands show their uncertainty.

C.2 Noise robustness

On the symmetric benchmark, scaled-H^2 regularisation reduces coefficient MAPE at 8% quote noise from 15.4% to 13.7% over paired seeds.

C.3 Validation-selected regularisation

A separate 60/20/20 split selects $\mu = 3.0 \times 10^{-6}$ by validation price error and gives 4.7% held-out test error. Coefficient-truth and L-curve choices are retained only as diagnostics; only the validation choice is deployable.

C.4 American obstacle extension

For payoff Φ and reaction r, backward Euler gives the strictly convex quadratic programme

$$\min_{u \geq \Phi} \tfrac{1}{2} u^\top \mathbf{A} u - b^\top u, \qquad \mathbf{A} = \mathbf{M} + \Delta\tau(\mathbf{K} + r\mathbf{M}).$$

Howard policy iteration yields frontier RMSE 2.13×10^{-3} against a refined reference solve and maximum KKT residual 4.19×10^{-15}. This extension uses the symmetric benchmark and is not needed for the local-volatility calibration claim.

C.5 E6: Nested-sample scaling

Under a fixed independent test set and nested training prefixes, 20 repetitions give coefficient-RMSE slope -0.42 with bootstrap interval $[-0.53, -0.31]$ and price-test slope -0.52 with interval $[-0.62, -0.40]$. These are empirical exponents for the implemented exact-Dupire design, not a dimension-free minimax theorem.

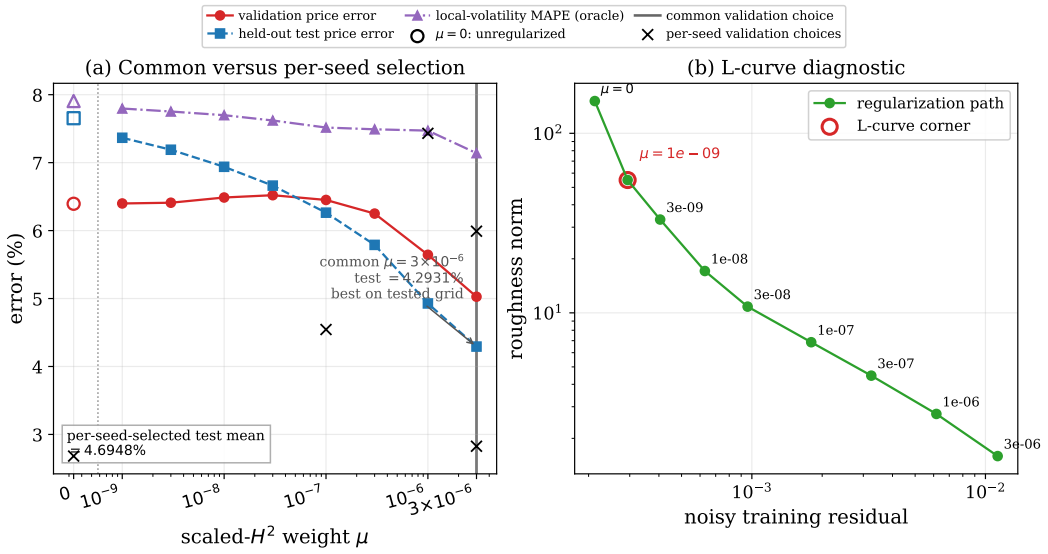

Figure 9: Validation-selected regularisation. Left: validation and held-out price errors together with oracle coefficient MAPE across scaled-H^2 weights, including the validation and oracle selections. Right: training residual against roughness along the same regularisation path, with the L-curve corner marked as a diagnostic.

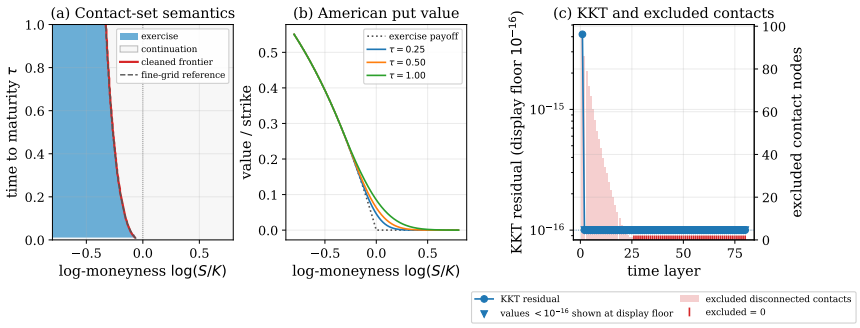

Figure 10: American obstacle extension. Left: economic contact set and exercise frontier against a fine-grid reference; centre: American put values and payoff at representative time layers; right: KKT residuals and excluded zero-payoff tail contacts across layers.

D ADJOINT-ENABLED UNCERTAINTY QUANTIFICATION

Let f denote a finite-dimensional pre-softplus field and suppose $y = \mathcal{Q}_D(\kappa(f)) + \varepsilon$, with $\varepsilon \sim \mathcal{N}(0, \Gamma)$ and prior $f \sim \mathcal{N}(m_0, C_0)$. At the MAP estimate $\hat{f}$, the Gauss–Newton Laplace approximation is

$$f \mid y \approx \mathcal{N}(\hat{f}, H^{-1}), \qquad H = J^\top \Gamma^{-1} J + C_0^{-1},$$

where J is assembled from tangent or adjoint solves.

Proposition 8 (Sandwich covariance)**.** *In the linearised regime, the sampling covariance of the penalised estimator is*

$$S = H^{-1}(J^\top \Gamma^{-1} J) H^{-1} = H^{-1} - H^{-1} C_0^{-1} H^{-1} \preceq H^{-1}.$$

Proof. Linearising the first-order condition gives $H(\hat{f} - f_\star) = J^\top \Gamma^{-1} \varepsilon$ plus a deterministic prior pull. Taking covariance yields the first identity; substituting $J^\top \Gamma^{-1} J = H - C_0^{-1}$ gives the second and the Loewner ordering. □

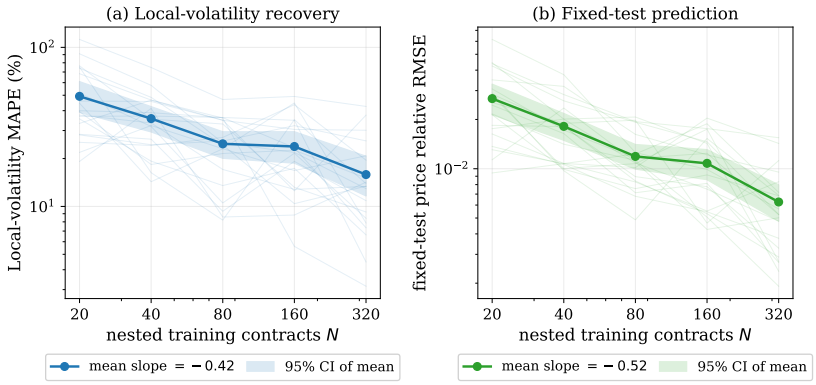

Figure 11: Nested-sample scaling over 20 repetitions. Left: coefficient RMSE versus the number of nested training contracts; right: relative price RMSE on a fixed independent test set. Thin curves show individual repetitions, solid curves show means, and shaded bands are bootstrap intervals for those means.

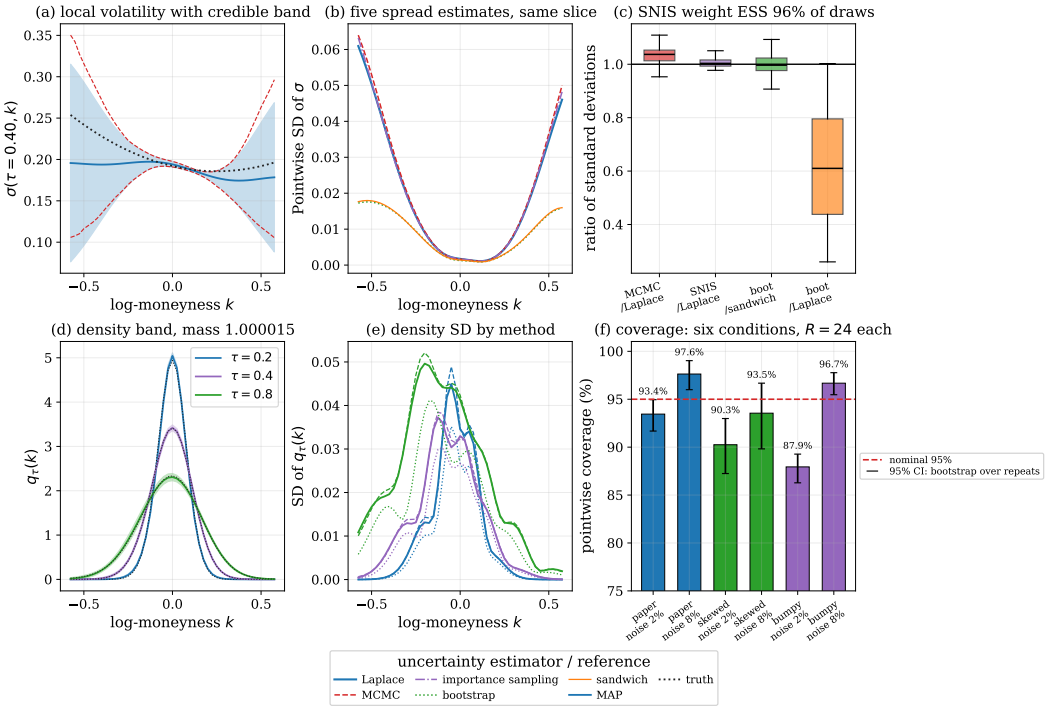

Figure 12: Adjoint-enabled uncertainty quantification with exact weighted Dupire. (a) MAP local-volatility slice with Laplace and MCMC bands; (b) pointwise spread estimates from Laplace, MCMC, importance sampling, bootstrap, and sandwich covariance; (c) standard-deviation ratio distributions; (d) dual-density bands; (e) density uncertainty by estimator and maturity; (f) repeated-noise coverage against the nominal 95% level.

The uncertainty-quantification study uses the exact weighted-Dupire map $\mathcal{Q}_D$ with a 36-dimensional tensor-Chebyshev field. Its bootstrap-to-Laplace standard-deviation ratio is 0.61, while the bootstrap-to-sandwich ratio is 1.00, as predicted. A posterior random-walk sampler gives ratio 1.04 and pointwise correlation 0.9995 with the Laplace spread. Across repeated-noise fits, the reported 95% bands cover 93.4% of evaluation nodes. These results validate a local approximation for the stated synthetic design and do not remove prior, model-bias, or boundary sensitivity.

E MARKET DATA AND REPRODUCIBILITY

Per-expiry forwards and discount factors are estimated from put–call parity, after which quotes are mapped to $k = \log(K/F_T)$ and out-of-the-money contracts are selected. Time to maturity uses ACT/365, and expiries with fewer than six clean quotes are discarded.

Weighted-Dupire operator $\mathcal{Q}_D$ assembles $\mathbf{K}(e^{-k})$ independently of the learned coefficient and places e^{-k}/κ in the weighted mass. It is used for synthetic coefficient recovery, density recovery, nested-sample scaling, out-of-sample S&P 500 calibration, the dual-density and score study, and adjoint-enabled uncertainty quantification. The symmetric benchmark $\mathcal{Q}_S$ is used for noise robustness, validation-selected regularisation, the American obstacle extension, hard constraints versus soft physics penalties, and neural-operator generalisation. In the dual-density and score study, convergence includes constant-κ analytic checks, whereas the Fokker–Planck, score, and sampling experiments use variable κ. This operator ledger is part of the claim boundary, not merely an implementation detail.