FINANCE-INFORMED OPERATOR LEARNING FOR OPTION PRICING WITH QUANTUM-COMPATIBLE REALIZATIONS

Anonymous authors
Paper under double-blind review

ABSTRACT

Pricing European options under local volatility requires solving a partial differential equation whose coefficients change with every recalibration, and practitioners repeatedly need not only prices but also their sensitivities to the underlying across spot–time surfaces for many strike and maturity configurations. Neural surrogates are an efficient alternative, approximating the solution operator directly so that model evaluation replaces repeated numerical solves. Near expiry, however, the solution loses regularity, and a network trained on prices alone can misstate the curvature needed for hedging or return prices outside the no-arbitrage bounds. We propose a finance-informed Deep Operator Network (FI-DeepONet) that decomposes the solution operator into a closed-form reference price and an additive learned correction. The framework is finance-informed in that financial structure enters the representation and the output map rather than the loss: the reference uses the strike-line integrated variance—the time integral of the squared local volatility evaluated at the strike, which captures the leading near-expiry curvature singularity, and a smooth monotone admissibility layer enforces pointwise no-arbitrage price bounds, at the cost of piecewise rather than global smoothness in the spot variable. A deterministic post-processing step then corrects the predicted grid after inference and is evaluated separately from the learned operator. We establish a short-maturity estimate and differentiated asymptotics for the exact correction, together with identities relating errors in the learned correction to pricing and sensitivity errors, as well as PDE-residual errors. On in-distribution (ID) tests, the model reduces the global relative price error by nearly an order of magnitude over vanilla and physics-informed DeepONet baselines across independent initializations. On out-of-distribution (OOD) tests, generalization is moderate when the input parameters lie outside the training range but within the same parametric class of local-volatility functions. We also test the trained model on real market data using index option quotes and find that it remains accurate without market-specific network retraining, though it does not improve on the analytic formula given the same volatility input. We further give a quantum-compatible realization of the trained operator, in which selected linear maps are either compiled exactly or approximated within a restricted diagonal–orthogonal family before supervised adaptation.

1 INTRODUCTION

Many scientific and engineering workflows solve the same family of partial differential equations (PDEs) many times, once for each new coefficient function or parameter setting. Neural operators pay off this cost by learning the map from PDE inputs to solutions, so that a single trained model replaces the numerical solver at inference time (Chen & Chen, 1995; Lu et al., 2021a; Li et al., 2021; Kovachki et al., 2023). Option pricing is a natural instance. A European call option gives its holder the right to buy an asset at a fixed price (the strike) on a fixed future date (the maturity). Under a local-volatility model, where the asset's volatility is a given function of asset price and time, the option price solves a linear parabolic PDE whose diffusion coefficient is set by that function (Black & Scholes, 1973; Merton, 1973; Dupire, 1994). Because the volatility function is recalibrated frequently and many contracts must be priced under each calibration, an operator that

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.

maps the volatility function and contract terms to the full price surface can replace a large number of repeated solves. Accurate prices alone are not enough in this setting. Risk management also relies on the sensitivities of the price to the asset price: its first derivative, Delta, and its second derivative, Gamma. These become especially challenging to learn near maturity. The payoff has a kink at the strike (it is zero below the strike and grows linearly above it), so as maturity approaches, Gamma grows without bound at the strike and concentrates in a shrinking neighborhood around it (Pooley et al., 2003; Giles & Carter, 2005). A model with small average price error can therefore still have large Gamma errors in the near-strike, short-maturity region. A generic network must learn this near-singular behavior from data, even though its form is known in closed form. This raises a question: where should known (prior) knowledge enter a neural operator? One common approach is physics-informed operator learning, which adds the PDE residual and boundary conditions to the training loss (Raissi et al., 2019; Wang et al., 2021b; Li et al., 2024). This encourages satisfaction of the equation but does not directly supply the singular structure of the solution, and its performance is sensitive to loss balancing, optimization, and the placement of collocation points (Wang et al., 2021a; Krishnapriyan et al., 2021; Chen et al., 2026b). An alternative, standard in classical numerical analysis but less common in operator learning, is to build the structure into the representation: subtract an analytic function that carries the singularity and let the network learn only the smoother remainder. Option pricing offers three such pieces of structure. (i) The payoff at maturity is known exactly. (ii) No-arbitrage arguments confine the price to a known interval: it cannot exceed the asset price, and it cannot fall below the larger of zero and the asset price minus the discounted strike. (iii) Near the strike at short maturities, the price is well approximated by the Black–Scholes formula parameterized by the strike-line integrated variance, that is, the time integral of the squared local volatility evaluated at the strike.

We propose the finance-informed Deep Operator Network (FI-DeepONet), which uses all three. Its core is a strike-matched carrier–residual ansatz (RA): the price is written as an analytic reference, which we call the carrier, plus a learned correction scaled by the remaining time to maturity. The carrier is the Black–Scholes price parameterized by this strike-line integrated variance. It therefore depends on the input volatility function, not only on the contract terms, and it reproduces the payoff kink exactly. Under standard well-posedness conditions, assuming that the local variance is bounded above and below by positive constants and is Lipschitz in log-moneyness, we prove that the exact correction vanishes at least linearly in the remaining time to maturity. This motivates the rescaling: dividing the correction by the remaining time gives a bounded learning target, and when the leading-order term is nonzero, linear scaling is the only power scaling that keeps the target both bounded and nonvanishing at maturity. A Financial Admissibility Layer (FAL), a smooth monotone map into the no-arbitrage interval, then enforces the price bounds at every output point. The model is trained on reconstructed prices with a standard regression loss. We retain a standard branch–trunk DeepONet architecture and encode the finance-informed structure through the representation and output map rather than the loss. The closest precedents are residual learning around analytic pricing approximations (Kim et al., 2025), exact-terminal and singularity-aware constructions (Zhang et al., 2026), and maturity-gated Black–Scholes representations (Sakuma, 2026). Our reference is distinguished by strike-line variance matching: it integrates the input local-variance field along the strike line to construct the carrier's strike-line integrated variance. We also analyze the resulting correction and its derivatives, rather than only reporting lower errors. Appendix F gives a broader comparison with related work.

Our contributions are as follows.

- **A finance-informed operator representation.** We propose a strike-matched carrier–residual ansatz in which an input-dependent Black–Scholes carrier, constructed from the strike-line integrated variance, captures the terminal singularity, while a DeepONet learns only a time-rescaled correction. A Financial Admissibility Layer maps every output into the no-arbitrage interval. The backbone and training objective are those of a standard DeepONet trained by price regression, so prior knowledge enters through the representation rather than through the loss.

- **Short-maturity analysis of the learning target.** We prove that the exact correction vanishes at least linearly in the remaining time to maturity and characterize its limiting shape, together with its first two spot derivatives, near the strike. We show that linear rescaling is the unique power scaling that keeps the target bounded and nonzero when the leading-order

term is nonzero, and we derive identities that relate errors in the learned correction to errors in price, Delta, Gamma, and the PDE residual.

- **Improved accuracy with controlled attribution.** On a local-volatility benchmark with finite-difference references, FI-DeepONet reduces global relative price error by about $6.2\times$ and $7.5\times$ relative to data-driven and physics-informed DeepONet baselines, respectively, and also reduces Delta and Gamma errors. Ablations attribute the gains to the change of learning target and the output layer, and quantify the trade-off between adding a PDE loss and price accuracy. The model generalizes to shifted contract and volatility parameters, while a selected final checkpoint does not accurately reproduce volatility perturbations outside the training family.
- **Quantum-compatible realization and market-data evaluation.** In contrast to Quantum DeepONet, which trains an orthogonally parameterized network from the start (Xiao et al., 2025), we begin from the pretrained dense model and realize selected linear maps with circuits that implement orthogonal transformations. We distinguish exact compilation of selected weight matrices from restricted circuit-compatible approximation followed by adaptation, and quantify the accuracy cost relative to classical structured approximations of equal computational cost. We further evaluate the model on index-option quotes, using implied volatilities as constant-volatility inputs and Black–Scholes prices on the same inputs as references, to assess transfer from synthetic to market-derived inputs. The learned operator remains accurate on market-derived inputs without market-specific retraining, although it does not outperform the same-input Black–Scholes reference. Neither study is intended to demonstrate a quantum speedup or a market calibration.

2 Pricing problem and finance-informed representation

We consider a European call option on an asset that pays no dividends. The holder may buy the asset at a fixed price $K > 0$ (the *strike*) on its expiry date $T > 0$ (the *maturity*). Let $V(S,t)$ be the option value when the asset's current (*spot*) price is S at time t. Its terminal payoff is $V(S,T) = \max(S - K, 0)$. The local volatility $\sigma(S,t)$ specifies the scale of random proportional price fluctuations at each asset price and time. With constant risk-free interest rate r, no-arbitrage pricing gives the backward PDE (Black & Scholes, 1973; Dupire, 1994)

$$\partial_t V + \tfrac{1}{2}\sigma^2(S,t)S^2\partial_{SS}V + rS\partial_S V - rV = 0, \qquad S > 0,\ t < T. \tag{2.1}$$

We write $a = \sigma^2$ for the local variance and $\tau = T - t$ for the remaining time. At the operator level, the learning task is $(a, K, T, r) \mapsto V(\cdot,\cdot)$, with (S,t) specifying where to evaluate the price surface. Delta, $\Delta = \partial_S V$, measures price sensitivity to the asset price; Gamma, $\Gamma = \partial_{SS}V$, measures its curvature. A financial glossary is provided in Appendix A.0. No-arbitrage arguments also give the pointwise price bounds

$$L(S,\tau) = \left(S - Ke^{-r\tau}\right)^+ \le V(S, T-\tau) \le U(S,\tau) = S. \tag{2.2}$$

Strike-matched carrier and learned correction. Instead of learning the full price directly, we separate an analytic reference (the carrier) from a learned correction. The reference fixes the spatial argument of the local variance at $S = K$ while retaining its time dependence. The resulting strike-line integrated variance is

$$A_0(\tau) = \int_0^\tau a(K, T-u)\, du. \tag{2.3}$$

The resulting Black–Scholes carrier is $C^{\mathrm{ref}} = S\Phi(d_+) - Ke^{-r\tau}\Phi(d_-)$, where $d_\pm = [\log(S/K) + r\tau \pm A_0/2]/\sqrt{A_0}$ and Φ is the standard normal distribution function. It exactly prices this spatially frozen model and equals the payoff at $\tau = 0$ (Appendix A.3). The benchmark directly evaluates $a(K,\cdot)$ from the coefficient generator, without sensor interpolation (Appendix B.3). Sensor-only inputs would require approximate strike-line evaluation; Appendix A.3 treats the resulting reference error. The strike-matched carrier–residual ansatz (RA) reconstructs the pre-FAL price as

$$z_\theta = C^{\mathrm{ref}} + \chi R_\theta, \qquad \chi = \tau/T. \tag{2.4}$$

Here χ is the remaining-life fraction and R_θ is the learned time-rescaled correction in price units. Under the value-estimate assumptions of Section 3, including positive boundedness and log-moneyness Lipschitz regularity, $W = V - C^{\mathrm{ref}} = O(\tau)$, so $R^* = W/\chi$ is bounded. Differentiated profiles require stronger local smoothness and allow spot derivatives to grow near expiry. Training supervises reconstructed prices, not correction labels.

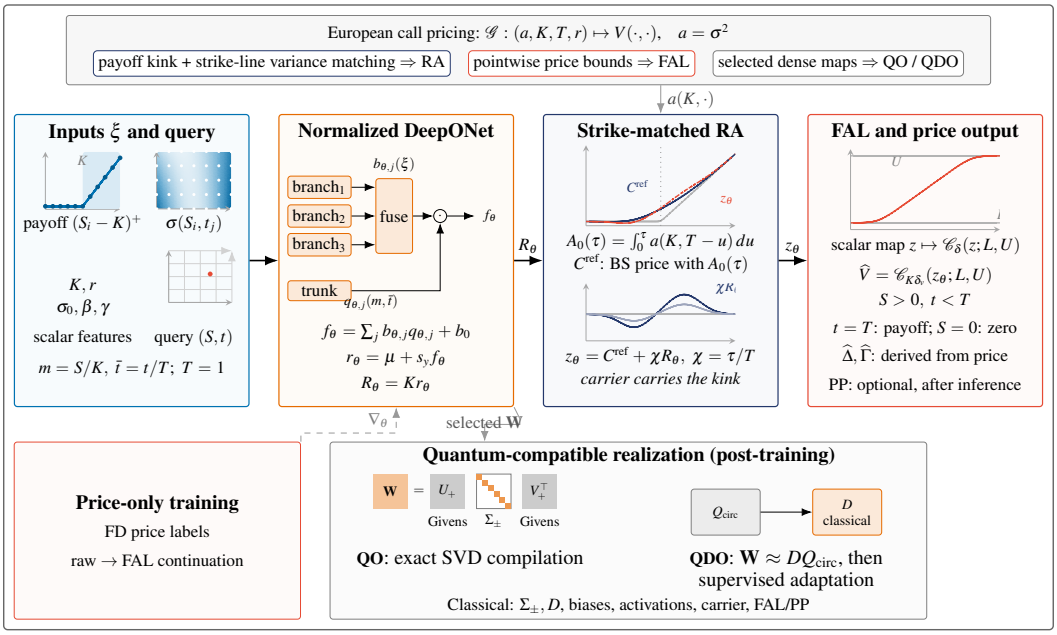


Figure 1: FI-DeepONet overview. The carrier uses strike-line integrated variance; DeepONet predicts R_θ, RA reconstructs z_θ, and FAL enforces pointwise price bounds. Training uses reconstructed-price regression on finite-difference (FD) labels. Post-training QO exactly compiles selected maps; QDO approximates and adapts them. Plots are schematic.

DeepONet architecture. The branch input ξ combines sampled payoff values, sampled volatilities $\sigma(S_i, t_j)$, and scalar features. Separate encoders are fused and combined with a trunk evaluated at $m = S/K$ (asset price relative to the strike) and $\bar{t} = t/T$ (normalized elapsed time). The branch–trunk readout gives $r_\theta = \mu + s_y f_\theta$ and $R_\theta = K r_\theta$, where μ, s_y are fixed training-normalization constants (Appendix B.4). The reported benchmark fixes $T = 1$ and uses scalar features $(K, r, \sigma_0, \beta, \gamma)$ for the contract and volatility family (Appendix B.2). Retaining absolute payoff samples and K with fixed standardizers makes the implementation scale-aware, not exactly scale-equivariant (Appendix A.1).

Financial Admissibility Layer and output stage. FAL maps the raw price z_θ into the interval in (2.2). Using the softplus $\mathrm{sp}_\delta(q) = \delta \log(1 + e^{q/\delta})$, a smooth approximation of $\max(q, 0)$, define

$$\mathcal{C}_\delta(z; L, U) = L + \mathrm{sp}_\delta(z - L) - \mathrm{sp}_\delta(z - U). \tag{2.5}$$

For fixed bounds, the map is smooth, increasing, and nonexpansive, but introduces smoothing bias (Appendix A.2). For $S > 0, t < T$, the prediction is $\widehat{V}_{\text{FI}} = \mathcal{C}_{K\delta_v}(z_\theta; L, U)$ with $\delta_v = 0.002$. Exact payoff and zero-spot assignments were added after training with unchanged weights and no held-out outcomes. They do not remove the fixed-width near-terminal limit bias. FAL enforces pointwise bounds, not spot monotonicity or convexity; the kinked lower bound makes the price only piecewise smooth, so Greek and PDE identities apply on smooth pieces.

FI denotes this output before optional grid post-processing (PP). PP repairs selected discrete price-bound, terminal, time-monotonicity, and spot-slope violations after inference, without changing weights; it is evaluated separately and provides no continuous-surface arbitrage certificate (Appendix B.5). The complete OOD study retains an earlier output convention (Appendix C.5).

3 Theoretical Analysis of Strike-Line Variance Matching

We write $x = \log(S/K)$ for log-moneyness and $W(S, t) = V(S, t) - C^{\text{ref}}(S, t)$ for the exact correction, using the calendar-time convention of Section 2. The value estimate assumes local variance bounded above and below by positive constants and uniformly Lipschitz in x. The differentiated profiles require the stronger local smoothness specified in Appendix A.3, which states the full assumptions.

Proposition 3.1 (Correction cancellation)**.** *Under the value-estimate assumptions in Appendix A.3, there is a constant C, independent of $\tau \in (0, T]$, such that*

$$\sup_{x\in\mathbb{R}} \frac{|W(Ke^x, T-\tau)|}{K} \le C\tau, \tag{3.1}$$

uniformly over the compact class of contracts stated there.

The correction has zero terminal data and source $\frac{1}{2}[a(S,T-\tau)-a(K,T-\tau)]S^2\partial_{SS}C^{\mathrm{ref}}$. In the $O(\sqrt{\tau})$ near-strike band, the Lipschitz coefficient difference is $O(\sqrt{\tau})$ and carrier curvature is $O(K/\sqrt{\tau})$; Gaussian decay controls the tails. The source is therefore bounded, giving (3.1) after time integration. A persistent strike-line mismatch can instead introduce an $O(\sqrt{\tau})$ contribution (Appendix A.3).

Proposition 3.2 (First-order skew profile and critical scaling)**.** *Under the differentiated-profile assumptions in Appendix A.3, let $a_* = a(K,T)$ be the local variance at the strike and maturity, $b_* = K\partial_S a(K,T)$ its log-moneyness slope, and g_{a_*} the centered Gaussian density with variance a_*. To resolve the shrinking $O(\sqrt{\tau})$ near-strike layer, we keep the scaled log-moneyness $y = x/\sqrt{\tau}$ fixed. For fixed y and $S_\tau = Ke^{y\sqrt{\tau}}$,*

$$\frac{W(S_\tau, T-\tau)}{K\tau} \longrightarrow \frac{b_*}{4}\, y\, g_{a_*}(y) \qquad (\tau \to 0). \tag{3.2}$$

Along these paths, if $b_* y \neq 0$, then W/χ^α with $\chi = \tau/T$ has a finite nonzero limit if and only if $\alpha = 1$. The corresponding profiles of the Delta and Gamma corrections, $\partial_S W$ and $\partial_{SS} W$, are given in (A.26)*. Thus the linear gate $\chi = \tau/T$ in (2.4) is the critical value-level rescaling near expiry.*

Boundedness of R^* does not imply bounded spot derivatives: along these paths, $\partial_S R^* = O(\tau^{-1/2})$ and $\partial_{SS} R^* = O(\tau^{-1})$. Manufactured-solution checks and the exponent ablation distinguish these asymptotics from finite-budget accuracy (Appendix C.1; Section 4.3).

Error transfer through the output stages. Let $\rho_\theta = R_\theta - R^*$ be the correction error. On the nonterminal domain Ω under consideration, define $N_\rho := \sup_\Omega\big(|\rho_\theta|/K + \sqrt{\chi}\,|\partial_S\rho_\theta| + \chi K|\partial_{SS}\rho_\theta|\big)$. Since $z_\theta - V = \chi\rho_\theta$, a finite N_ρ gives the pre-FAL bounds

$$\frac{|z_\theta - V|}{K} \le \chi N_\rho, \qquad |\partial_S z_\theta - \Delta| \le \sqrt{\chi}\, N_\rho, \qquad K|\partial_{SS} z_\theta - \Gamma| \le N_\rho, \tag{3.3}$$

where $\Delta = \partial_S V$ and $\Gamma = \partial_{SS} V$. These conditional bounds give $O(\chi)$ normalized price error and $O(\sqrt{\chi})$ Delta error; they bound the K-scaled Gamma error without requiring it to vanish. These bounds are conditional on finite N_ρ, which finite-grid checks do not establish. After FAL, price error includes smoothing bias, while Greek errors also involve derivatives of the bounds and output map (Section 2; Appendix A.4). Writing $P_a := \partial_\tau - \frac{1}{2}a(S,T-\tau)S^2\partial_{SS} - rS\partial_S + r$ for the time-to-maturity pricing operator, the exact pre-FAL identity is $P_a z_\theta = \chi P_a \rho_\theta + \rho_\theta/T$. This connects correction error to the governing-equation defect without requiring a PDE loss. The reported finite-grid residual RMS is insufficient to certify the required continuum regularity.

4 Numerical Experiments

Principal comparisons use pre-PP outputs (FI after FAL), FD Greeks and PDE diagnostics, and five-seed means; sample standard deviations are reported in Appendix C.2. Mechanism ablations use pre-FAL outputs; the PDE-loss factorial uses automatic-differentiation (AD) diagnostics.

4.1 Benchmark and metrics

Data and models. The fixed-$T = 1$ family varies volatility level σ_0, spot dependence β, time dependence γ, strike K, and rate r. Fixed splits contain 512/32/32 training/validation/ID surfaces and 320 parameter-support OOD surfaces. Sensors, inputs, and direct carrier access follow Section 2 and Appendix B.2. All principal models have width 128, 258,305 parameters, and validation-price-based checkpoint selection. DeepONet is the unnormalized direct-price baseline; FI uses RA/FAL; PI adds data, PDE, terminal, and boundary losses without RA/FAL. PI's baseline collocation omits the shortest maturities and has no near-strike enrichment (Appendix B.4).

Table 1: Principal pre-PP results. A: ID (32 surfaces); B: component ablation; C: parameter-shift OOD (320 surfaces). Five-seed means are shown unless deterministic; SDs are reported in Appendices C.2 and C.5. Bold: lowest mean.

A. In-distribution (32 surfaces)

Metric	DeepONet	PI	FI
Price ($\times 10^{-3}$)	5.99	7.24	**0.965**
Near p95	1.80	2.33	**0.158**
Delta p95	0.152	0.155	**0.0245**
Gamma p95	0.0332	0.0355	**0.00434**
PDE RMS	0.0461	0.0105	**0.00703**

B. Component ablation

Configuration	Price ($\times 10^{-3}$)	Near p95
Normalized	7.6	2.26
+ soft penalties	7.5	2.21
+ FAL	2.4	0.528
FI (RA+FAL)	**0.965**	**0.158**
Carrier only	6.89	0.393

C. Parameter-shift OOD (320 surfaces)

	DeepONet	FI
Price ($\times 10^{-3}$)	197	**1.52**

Metrics. Price RelL2 pools the test grid, including maturity; price p95 is the 95th percentile of absolute error near the strike. Near-strike means $|S/K - 1| \leq 0.05$; ATM-short further restricts $0 < \tau \leq 0.05$. FD Delta/Gamma p95 use nonterminal near-strike points on the 24-surface higher-resolution L3 subset. The common dimensionless FD PDE-residual RMS uses interior points, with terminal prices entering its last time stencil. Refining the reference preserves the FI–PI ordering, although shortest-maturity Gamma remains resolution-sensitive (Appendices B.6, C.1).

4.2 Main Results

FI-DeepONet achieves six- to eightfold lower mean global price error than the vanilla and PI baselines, with lower FD Greek p95 errors (Table 1A; Figure 2a). Despite price-only supervision, it also has the lowest common FD PDE-residual RMS. Lower residuals alone do not predict better prices: PI improves this diagnostic over DeepONet but worsens price error. The FI–PI gap persists under the predeclared collocation and boundary-target sensitivity studies (Appendices C.4, C.8). Figures 2c–e show the corresponding shortest-maturity error profiles and local error ratios, complementing these pooled metrics.

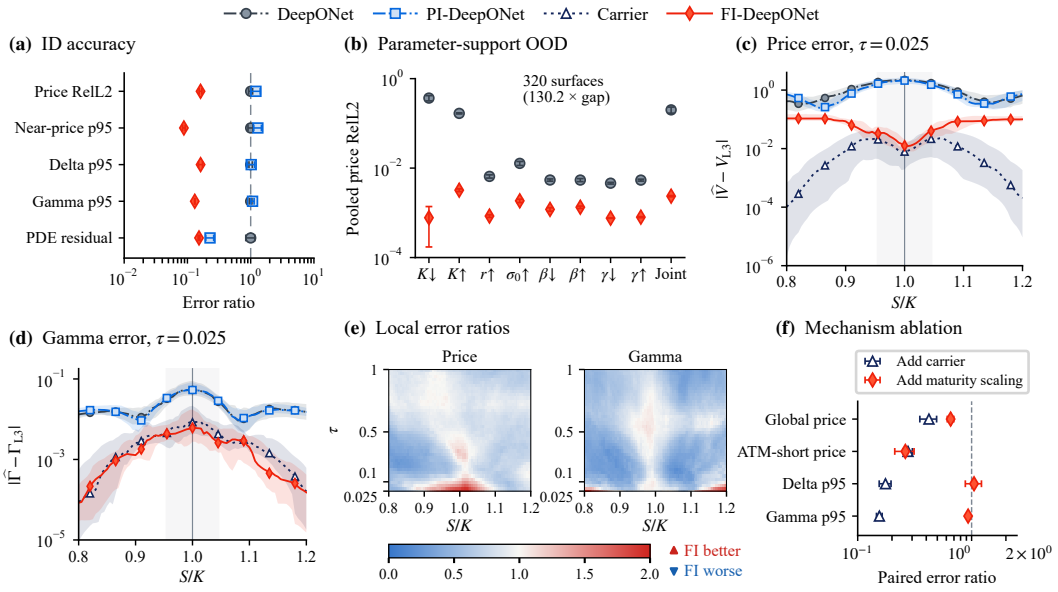


Figure 2: Classical evaluation before PP. (a) ID error ratios to DeepONet. (b) Complete parameter-support OOD set. (c,d) Price/Gamma errors at $\tau = 0.025$: medians and interquartile ranges over 24 L3 surfaces and five seeds. (e) Local error ratios; red favors FI. (f) Pre-FAL paired carrier (B0→A0) and rescaling (A0→A1) effects. Protocols: Appendix B.6.

4.3 Ablation study

Carrier and rescaling. FAL improves Normalized DeepONet, and full FI further improves both price metrics (Table 1B). A five-seed, 7,000-update matched study compares B0 (direct-price control), A0 (carrier without rescaling), and A1 (full RA). Before FAL, adding the carrier lowers global price and near-strike Delta/Gamma errors in every paired seed. Linear rescaling then improves global and ATM-short price accuracy, cutting the latter error by about three quarters, but slightly worsens near-strike Delta p95 (Figure 2f; Appendix C.2.1).

Exponent and representation controls. For the rescaling exponent α in the generalized reconstruction $z_{\theta,\alpha} = C^{\mathrm{ref}} + g_\alpha(\chi)R_\theta$, with $g_0(\chi) = 1$ and $g_\alpha(\chi) = \chi^\alpha$ for $\alpha > 0$, we test $\alpha \in \{0, \frac{1}{2}, 1, \frac{3}{2}\}$. The lowest mean post-FAL Gamma p95 occurs at $\alpha = 1$, while price error favors $\alpha = \frac{1}{2}$. Thus asymptotic criticality does not imply metric-wise training optimality. A separate three-seed study favors RA over the tested coordinate and sampling controls, but global low-rank reconstruction is worse for both correction targets than for full prices. The evidence supports local cancellation of the near-expiry singular structure; rank diagnostics show no reduction in global linear complexity (Appendices C.2.1, C.2).

4.4 Effect of PDE loss and limited data

The five-seed normalized-RA factorial separates the PDE-loss intervention from representation. Without FAL, adding the PDE loss lowers AD residual RMS and both Greek MAEs in every seed, but improves price error in only three. With FAL, it lowers residuals while worsening price error in all five seeds: optimizing the PDE loss and price accuracy need not improve together. Separate limited-data studies show benefits from PDE loss at 25% data and from FAL at every tested fraction; financial penalties add modest gains to FAL at the smallest fractions. These are development results, not changes to the final price-only FI objective (Appendix C.3).

4.5 Out-of-distribution generalization

Across all 320 parameter-shift surfaces, FI's pooled price error is about two orders of magnitude below DeepONet's by the ratio of five-seed means (Table 1C; Figure 2b). This tests the complete predictor within the declared functional family. The historical OOD output convention precludes a same-protocol ID/OOD degradation ratio (Appendix C.5).

To test beyond this coefficient family, we use carrier-null off-family perturbations that vanish at the strike, leaving the carrier unchanged. The selected FI checkpoint responds but does not accurately reproduce the reference directional price response or its spot derivatives, despite agreement between AD and FD of its own predictions. A separate three-seed multi-family study improves these responses with derivative supervision at a cost in price accuracy (Appendix C.6).

5 Quantum-Compatible Realizations and Market Evaluation

We study quantum compatibility and market-derived inputs as downstream evaluations of the trained FI operator, with the RA/FAL construction fixed.

5.1 Quantum-Compatible Realizations

Setup. Unlike Quantum DeepONet (Xiao et al., 2025), we start from a pretrained dense FI parent. QO exactly factors selected maps into orthogonal rotations and a classical signed core, without adaptation. QDO approximates eligible maps by $\mathbf{W} \approx DQ_{\mathrm{circ}}$, with diagonal D and restricted orthogonal $Q_{\mathrm{circ}} \in \mathcal{Q}_{\mathrm{circ}}$. Full QDO adapts all trainable parameters using pre-FAL price regression; a fixed composite score selects checkpoints and coverage. The 410/102 fitting/selection split comes entirely from the original synthetic training set, and all candidates share one parent. Cores, scalings, biases, nonlinearities, carrier, and FAL remain classical (Appendices D.1–D.5).

Fidelity and adaptation. QO reproduces parent predictions to numerical precision in idealized evaluation. At the selected C94 coverage, QDO roughly doubles global price error; near-strike price p95 changes little, Delta p95 worsens, and Gamma p95 improves (Table 2A,B). The selected C94 instance fails the fixed ID/OOD accuracy-preservation criteria (Appendix D.5). These synthetic evaluations use AD Greeks under a protocol separate from the principal FD study. In the three-seed matched C94 study, updating only substituted maps and their biases achieves almost all frozen-to-

Table 2: Selected C94 realization versus its fixed FI parent. A: MAC coverage and update scope; B: separate 32-surface AD synthetic evaluation; C: 2,586 held-out-IV quotes. Single selected instance, not seed averages; bold: lower error.

A. Realization scope		
Property	QO	QDO (C94)
Dense-map MAC coverage	49.6%	92.8%
Parameters updated	None	All trainable
B. Paired synthetic accuracy		
Metric	Parent FI-DeepONet	QDO (C94)
Global price RelL2 ($\times 10^{-3}$)	**0.837**	1.77
Near-strike price p95 error	**0.1315**	0.1324
Global AD Delta p95 error	**0.0127**	0.0148
Global AD Gamma p95 error	0.00244	**0.00224**
C. Market quotes with held-out implied volatility		
Metric	Parent FI-DeepONet	QDO (C94)
Market RMSE (quoted price units)	19.54	**14.92**
Operator RelL2 ($\times 10^{-3}$)	13.18	**5.48**

full error recovery (Figure 3a). Across paired C49–C99 configurations, mean price errors remain near 1.8×10^{-3} without a monotone coverage trend (Figure 3b; Appendix D.6).

Resource scope. QDO's lower listed errors relative to historical classical structured approximations of comparable computational cost are descriptive because their adaptation protocols are not matched (Appendix D.7); the comparison does not isolate the effect of the DQ parameterization. Computational cost here follows the arithmetic/storage accounting convention for such comparisons, rather than measured runtime. Whole-dense-map coverage ω counts reference dense-layer multiply–accumulate operations (MACs), while eligible-map coverage ρ uses a different denominator; both are static analytical coverage measures, and the saved partial implementation computes both dense and structured outputs. Figure 3c reports the resulting conditional cost scenarios analytically, without measuring hardware speedup. A complete end-to-end assessment would further require accounting for preparation, measurement, precision, and the retained classical workload (Guseynov et al., 2026) (Appendix D.8).

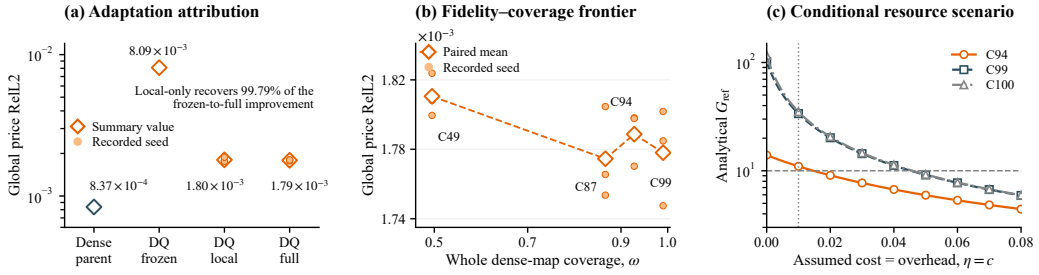


Figure 3: Quantum-compatible evaluation. (a) C94 adaptation attribution over three paired seeds. (b) Paired C49–C99 fidelity–coverage frontier; zoomed vertical scale. (c) Conditional $G_{\text{ref}} = [1 - \omega(1-\eta) + c]^{-1}$ with $\eta = c$. No hardware speedup is measured.

5.2 Market Evaluation with Constant-Volatility Inputs

Setup. We evaluate Nasdaq-100 (NDX) call quotes from 13 May to 7 August 2026. Implied volatility (IV) is a constant-volatility parameter inferred from a quote under a pricing model. Own-IV uses the target's supplied IV; held-out-IV reconstructs it from other strikes on the same date and expiry, excluding the target from the fit. Inputs set $\beta = \gamma = 0$ and use constant-volatility sensors. Under the effective-spot mapping in Appendix E, $\beta = \gamma = 0$ makes the supplied volatility field constant in spot and time, so the same-input exact solution is the Black–Scholes carrier: $V = C^{\text{ref}}$ and $W = R^* = 0$ for $\tau > 0$. The fixed-$T = 1$ network retains actual remaining maturity through the

effective-spot and contract-scaling map in Appendix E. Operator error uses the same-input Black–Scholes reference; market error uses the bid–offer midpoint. Weights and normalizers remain fixed, and all predictions precede PP.

Results. FI improves over both normalized neural baselines in both IV settings, but not the same-input Black–Scholes reference (Table 33). On held-out IV, its five-seed mean market relative RMSE is about thirteen times lower than Normalized DeepONet's and three times lower than the FAL-only baseline's. Separate diagnostics place FI's median market error at about one half-spread, with larger tail errors. Figure 4a visualizes market quotes against the frozen FI surface on one held-out date, while operator-error checks remain stable across the tested IV reconstructions and late dates (Figure 4b).

Despite the synthetic fidelity loss above, across C49–C99, all twelve paired coverage–adaptation–seed comparisons reduce market RMSE against the same fixed FI parent. Full-sample geometric means indicate roughly 20–25% reductions; all four coverages also improve on the sealed late-date subset. The reported date/chain bootstrap intervals lie below one conditional on the trained realizations (Figure 4c; Appendix E.4). The error decomposition supports sample-dependent price-error alignment as the explanation for these reductions, without evidence of a quantum-specific regularization effect. The study evaluates constant-volatility inputs; local-volatility calibration is outside its scope, and incomplete upstream rate/dividend and contract records limit source-to-model reproduction (Appendix E.5).

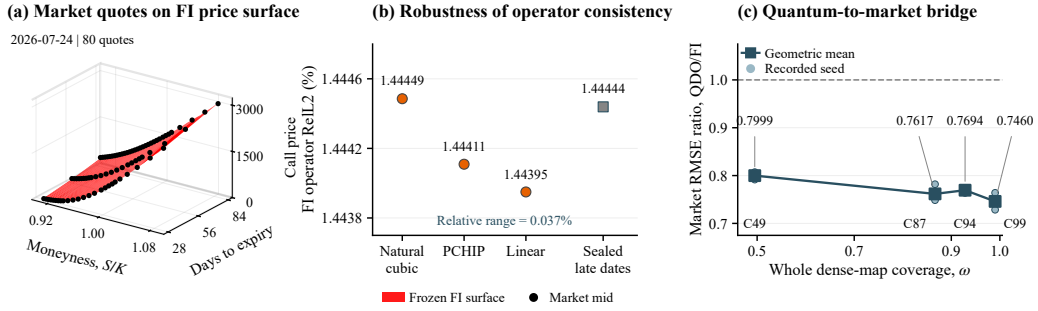


Figure 4: Frozen-network market evaluation before PP. (a) FI price surface on 2026-07-24 with 80 market mid-quotes. (b) Same-input Black–Scholes operator error under IV reconstruction and late-date checks; zoomed vertical scale. (c) Paired C49–C99 QDO/FI market-RMSE ratios; squares show three-seed geometric means and circles recorded seeds. Protocols: Appendix E.

6 Limitations and Conclusion

FI-DeepONet shows that known structure of a parabolic pricing problem can enter an operator network through its representation and output map, rather than through additional loss terms. A carrier matched to the strike-line integrated variance captures the leading near-expiry singular structure. The network learns a time-rescaled correction whose exact target is bounded under the stated assumptions; FAL enforces pointwise price bounds. Trained on prices alone, the model lowers global price error six- to eightfold relative to vanilla and physics-informed DeepONet, and it also reduces errors in Delta and Gamma, the first and second spot derivatives of the price, and in the PDE residual. Ablations attribute these gains to the carrier and the rescaling, with trade-offs that depend on the metric. Selected layers can be compiled exactly into orthogonal circuits, or approximated within a restricted family and then adapted. The frozen model also stays consistent on market-derived constant-volatility inputs. These conclusions are limited to one low-dimensional coefficient family, fixed sensors, and a fixed horizon; off-family tests leave general functional extrapolation unresolved. Theory and matched ablations support local cancellation of the leading near-expiry singular structure, while rank diagnostics show no reduction in global linear complexity. Fixed-width FAL retains near-terminal bias and a spot kink; PP repairs only a discrete grid. Neither step certifies an arbitrage-free continuous surface. Quantum results are classical or idealized, so hardware speedup remains unmeasured. Market tests concern constant-volatility inputs; same-input Black–Scholes remains more accurate, and source-to-model provenance is incomplete. Future work includes terminal-scaled FAL, broader coefficient families, and end-to-end resource measurements.

AI USE STATEMENT

Generative AI tools were used to assist literature and source discovery; research ideation and experimental-methodology refinement; implementation, debugging, and execution of numerical analysis and evaluation workflows; mathematical derivation, proof drafting, and proof checking; analysis and interpretation of results; manuscript organization, section drafting, language editing, and consistency checks; and development of the LaTeX code for scientific figures. This includes assistance with mathematical claims and their proofs, rather than language polishing alone. The synthetic numerical datasets and reference labels used in the experiments were generated by analytical formulas or finite-difference solvers under specified parameterized protocols; generative AI was not used to synthesize observations, numerical labels, or replacement benchmark data. AI tools also assisted preparation and auditing of the submission. AI-assisted calculations and consistency checks are not independent validation of a proof or an empirical claim. The authors retain responsibility for the final text, mathematical arguments, experimental design, numerical results, citations, code, and data-use permissions, including all AI-assisted content.

REPRODUCIBILITY STATEMENT

Appendix A states the assumptions, proofs, notation, and output-stage qualifications. Appendix B specifies the fixed data, inputs, architecture, optimization, checkpoint selection, masks, and metric units. Appendix C reports additional numerical results, reference checks, ablations, and sensitivity studies. Appendix D distinguishes exact selected-map compilation from adapted diagonal–orthogonal approximation and conditional resource accounting. Appendix E defines the market-input mapping, target-excluded IV reconstruction, evaluation populations, and remaining provenance limitations. Appendix F summarizes the replay scope across the historical synthetic records, the separate five-seed E0 refit, and the complete-population OOD replay. Reproducibility claims concern fixed realized data, documented selected states, and explicitly identified refits and replays; a fresh retraining of every historical principal run has not been performed. Some split-generation, circuit-topology, and market-feature provenance remains incomplete, and licensed quote-level market records cannot be redistributed.

ETHICS STATEMENT

The models are research approximations and have not been validated for deployment in trading or risk management. Pointwise price bounds, discrete grid repairs, and finite numerical checks do not establish reliable hedging or continuous-surface absence of arbitrage. The market study evaluates constant-volatility inputs and does not establish local-volatility calibration or superiority to the same-input Black–Scholes reference. The quantum-compatible studies do not demonstrate measured hardware speedup. The market data are licensed; any artifact release must respect applicable access and redistribution permissions. The manuscript and figure-source package does not confer rights to redistribute the underlying market records.

REFERENCES

M. Abid, A. Sojitra, B. Tiwari, and O. San. Quantum SEDONet: Spectrally-embedded quantum deep operator networks for partial differential equations. arXiv:2608.27626, 2026. URL https://arxiv.org/abs/2608.27626.

A. M. Aboussalah, X. Li, C. Chi, and R. Patel. Finance-informed neural network: Learning the geometry of option pricing. arXiv:2412.12213v2, 2026. URL https://arxiv.org/abs/2412.12213v2. Revised 2026.

Y. Achdou and O. Pironneau. *Computational Methods for Option Pricing*, volume 30 of *Frontiers in Applied Mathematics*. SIAM, Philadelphia, 2005.

D. Ackerer, N. Tagasovska, and T. Vatter. Deep smoothing of the implied volatility surface. In *Advances in Neural Information Processing Systems*, volume 33, 2020.

J. Berner, P. Grohs, and A. Jentzen. Analysis of the generalization error: Empirical risk minimization over deep artificial neural networks overcomes the curse of dimensionality in the numerical approximation of Black–Scholes partial differential equations. *SIAM Journal on Mathematics of Data Science*, 2(3):631–657, 2020. doi: 10.1137/19M125649X.

F. Black and M. Scholes. The pricing of options and corporate liabilities. *Journal of Political Economy*, 81(3):637–654, 1973.

D. T. Breeden and R. H. Litzenberger. Prices of state-contingent claims implicit in option prices. *Journal of Business*, 51(4):621–651, 1978.

H. Buehler, B. Horvath, and A. Kratsios. DYSANOS: Generative dynamic smooth arbitrage-free non-parametric option surfaces. arXiv:2608.12587, 2026a. URL https://arxiv.org/abs/2608.12587.

H. Buehler, B. Horvath, A. Kratsios, Y. Limmer, and R. Saqur. SANOS: Smooth strictly arbitrage-free non-parametric option surfaces. arXiv:2601.11209, 2026b. URL https://arxiv.org/abs/2601.11209.

P. Carr and D. B. Madan. A note on sufficient conditions for no arbitrage. *Finance Research Letters*, 2(3):125–130, 2005. doi: 10.1016/j.frl.2005.04.005.

S. Chakrabarti, R. Krishnakumar, G. Mazzola, N. Stamatopoulos, S. Woerner, and W. J. Zeng. A threshold for quantum advantage in derivative pricing. *Quantum*, 5:463, 2021. doi: 10.22331/q-2021-06-01-463.

J. Chen, Y. Li, and A. Neufeld. Quantum Monte Carlo algorithm for option pricing and its complexity analysis. *Mathematical Finance*, 2026a. doi: 10.1111/mafi.70036. Advance online publication.

N. Chen, C. Cui, A. Chen, S. Wang, and R. Ma. On the training of physics-informed neural operators for solving parametric partial differential equations. arXiv:2606.06164, 2026b. URL https://arxiv.org/abs/2606.06164.

T. Chen and H. Chen. Universal approximation to nonlinear operators by neural networks with arbitrary activation functions and its application to dynamical systems. *IEEE Transactions on Neural Networks*, 6(4):911–917, 1995. doi: 10.1109/72.392253.

Q. Cheng, M. H. Sahadath, H. Yang, S. Pan, and W. Ji. MD-PNOP: Equation-recast neural operators for minimal-data extrapolation and PDE solver acceleration. *Journal of Computational Physics*, 557:114844, 2026. doi: 10.1016/j.jcp.2026.114844.

W. Cheng, N. Costanzino, J. Liechty, A. Mazzucato, and V. Nistor. Closed-form asymptotics and numerical approximations of 1D parabolic equations with applications to option pricing. *SIAM Journal on Financial Mathematics*, 2(1):901–934, 2011. doi: 10.1137/100796832.

F. Corielli, P. Foschi, and A. Pascucci. Parametrix approximation of diffusion transition densities. *SIAM Journal on Financial Mathematics*, 1(1):833–867, 2010. doi: 10.1137/080742336.

F. de Feo. Universal approximation of nonlinear operators and their derivatives. arXiv:2605.15285v2, 2026. URL https://arxiv.org/abs/2605.15285v2.

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

C. Dugas, Y. Bengio, F. Bélisle, C. Nadeau, and R. Garcia. Incorporating second-order functional knowledge for better option pricing. In *Advances in Neural Information Processing Systems*, volume 13, 2000.

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

H. Eivazi, J.-A. Tröger, S. Wittek, S. Hartmann, and A. Rausch. EquiNO: A physics-informed neural operator for multiscale simulations. *Journal of Computational Physics*, 554:114745, 2026. doi: 10.1016/j.jcp.2026.114745.

A. Friedman. *Partial Differential Equations of Parabolic Type*. Prentice-Hall, Englewood Cliffs, NJ, 1964.

W. Gao, R. Xu, Y. Deng, and Y. Liu. Discretization-invariance? on the discretization mismatch errors in neural operators. In *International Conference on Learning Representations*, 2025.

J. Gatheral, E. P. Hsu, P. Laurence, C. Ouyang, and T.-H. Wang. Asymptotics of implied volatility in local volatility models. *Mathematical Finance*, 22(4):591–620, 2012.

M. B. Giles and R. Carter. Convergence analysis of Crank–Nicolson and Rannacher time-marching. *Journal of Computational Finance*, 9(4):89–112, 2005.

N. Guseynov, N. Liu, C. S. Pun, and T. Vaidya. End-to-end PDE-based quantum algorithms for multi-asset option pricing under local and stochastic volatility. arXiv:2605.26610, 2026. URL https://arxiv.org/abs/2605.26610.

J. Han, A. Jentzen, and W. E. Solving high-dimensional partial differential equations using deep learning. *Proceedings of the National Academy of Sciences*, 115(34):8505–8510, 2018. doi: 10.1073/pnas.1718942115.

D. Herman, Y. Sun, J.-P. Liu, M. Pistoia, C. Che, R. Otter, S. Chakrabarti, and A. Harrow. Quantum speedups for derivative pricing beyond Black–Scholes. arXiv:2602.03725, 2026. URL https://arxiv.org/abs/2602.03725.

B. Horvath, A. Muguruza, and M. Tomas. Deep learning volatility: A deep neural network perspective on pricing and calibration in (rough) volatility models. *Quantitative Finance*, 21(1):11–27, 2021. doi: 10.1080/14697688.2020.1817974.

J. M. Hutchinson, A. W. Lo, and T. Poggio. A nonparametric approach to pricing and hedging derivative securities via learning networks. *Journal of Finance*, 49(3):851–889, 1994.

A. Itkin. Deep learning calibration of option pricing models: Some pitfalls and solutions. arXiv:1906.03507, 2019. URL https://arxiv.org/abs/1906.03507.

P. Jin, S. Meng, and L. Lu. MIONet: Learning multiple-input operators via tensor product. *SIAM Journal on Scientific Computing*, 44(6):A3490–A3514, 2022. doi: 10.1137/22M1477751.

I. Kerenidis, J. Landman, and N. Mathur. Classical and quantum algorithms for orthogonal neural networks. arXiv:2106.07198, 2021. URL https://arxiv.org/abs/2106.07198.

Donghyun Kim, Jeonggyu Huh, and Ji-Hun Yoon. Improved accuracy of an analytical approximation for option pricing under stochastic volatility models using deep learning techniques. *Computers and Mathematics with Applications*, 187:150–165, 2025. doi: 10.1016/j.camwa.2025.03.029.

N. Kovachki, Z. Li, B. Liu, K. Azizzadenesheli, K. Bhattacharya, A. Stuart, and A. Anandkumar. Neural operator: Learning maps between function spaces with applications to PDEs. *Journal of Machine Learning Research*, 24(89):1–97, 2023.

A. S. Krishnapriyan, A. Gholami, S. Zhe, R. M. Kirby, and M. W. Mahoney. Characterizing possible failure modes in physics-informed neural networks. In *Advances in Neural Information Processing Systems*, volume 34, 2021.

I. E. Lagaris, A. Likas, and D. I. Fotiadis. Artificial neural networks for solving ordinary and partial differential equations. *IEEE Transactions on Neural Networks*, 9(5):987–1000, 1998. doi: 10.1109/72.712178.

J. Landman, N. Mathur, Y. Y. Li, M. Strahm, S. Kazdaghli, A. Prakash, and I. Kerenidis. Quantum methods for neural networks and application to medical image classification. *Quantum*, 6:881, 2022. doi: 10.22331/q-2022-12-22-881.

S. Lanthaler, S. Mishra, and G. E. Karniadakis. Error estimates for DeepONets: A deep learning framework in infinite dimensions. *Transactions of Mathematics and Its Applications*, 6(1): tnac001, 2022. doi: 10.1093/imatrm/tnac001.

Z. Li, N. Kovachki, K. Azizzadenesheli, B. Liu, K. Bhattacharya, A. Stuart, and A. Anandkumar. Fourier neural operator for parametric partial differential equations. In *International Conference on Learning Representations*, 2021.

Z. Li, H. Zheng, N. Kovachki, D. Jin, H. Chen, B. Liu, K. Azizzadenesheli, and A. Anandkumar. Physics-informed neural operator for learning partial differential equations. *ACM/IMS Journal of Data Science*, 1(3):1–27, 2024. doi: 10.1145/3648506.

G. M. Lieberman. *Second Order Parabolic Differential Equations*. World Scientific, Singapore, 1996.

L. Lu, P. Jin, G. Pang, Z. Zhang, and G. E. Karniadakis. Learning nonlinear operators via DeepONet based on the universal approximation theorem of operators. *Nature Machine Intelligence*, 3:218–229, 2021a. doi: 10.1038/s42256-021-00302-5.

L. Lu, R. Pestourie, W. Yao, Z. Wang, F. Verdugo, and S. G. Johnson. Physics-informed neural networks with hard constraints for inverse design. *SIAM Journal on Scientific Computing*, 43(6): B1105–B1132, 2021b. doi: 10.1137/21M1397908.

D. Luo, T. O'Leary-Roseberry, P. Chen, and O. Ghattas. Dimension reduction for derivative-informed operator learning: An analysis of approximation errors. *Journal of Machine Learning Research*, 27(182):1–94, 2026.

L. D. McClenny and U. M. Braga-Neto. Self-adaptive physics-informed neural networks. *Journal of Computational Physics*, 474:111722, 2023. doi: 10.1016/j.jcp.2022.111722.

R. C. Merton. Theory of rational option pricing. *Bell Journal of Economics and Management Science*, 4(1):141–183, 1973.

C. Mou, Y. Zhang, X. Zhu, and Q. Zhuang. PAS-Net: Physics-informed adaptive scale deep operator network. arXiv:2511.14925, 2025. URL https://arxiv.org/abs/2511.14925.

M. Noguer i Alonso. Derivative-informed operator learning for finance: On-the-fly Greeks, surfaces, hedging, and control. arXiv:2606.05900, 2026. URL https://arxiv.org/abs/2606.05900.

T. O'Leary-Roseberry, P. Chen, U. Villa, and O. Ghattas. Derivative-informed neural operator: An efficient framework for high-dimensional parametric derivative learning. *Journal of Computational Physics*, 496:112555, 2024. doi: 10.1016/j.jcp.2023.112555.

D. Pirjol and L. Zhu. Short-maturity asymptotics for option prices with interest rate effects. *International Journal of Theoretical and Applied Finance*, 26(6–7):2350023, 2023. doi: 10.1142/S0219024923500231.

D. M. Pooley, K. R. Vetzal, and P. A. Forsyth. Convergence remedies for non-smooth payoffs in option pricing. *Journal of Computational Finance*, 6(4):25–40, 2003.

Y. Qiu, N. Bridges, and P. Chen. Derivative-enhanced deep operator network. In *Advances in Neural Information Processing Systems*, volume 37, pp. 20945–20981, 2024.

M. Raissi, P. Perdikaris, and G. 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. doi: 10.1016/j.jcp.2018.10.045.

R. Rannacher. Finite element solution of diffusion problems with irregular data. *Numerische Mathematik*, 43:309–327, 1984. doi: 10.1007/BF01390130.

P. Rebentrost, B. Gupt, and T. R. Bromley. Quantum computational finance: Monte Carlo pricing of financial derivatives. *Physical Review A*, 98:022321, 2018. doi: 10.1103/PhysRevA.98.022321.

M. Sakarvadia, K. Hegazy, A. Totounferoush, K. Chard, Y. Yang, I. Foster, and M. W. Mahoney. The false promise of zero-shot super-resolution in machine-learned operators. In *International Conference on Learning Representations*, 2026.

T. Sakuma. Differential machine learning for 0DTE options with stochastic volatility and jumps. arXiv:2603.07600, 2026. URL https://arxiv.org/abs/2603.07600.

P. H. Schönemann. A generalized solution of the orthogonal Procrustes problem. *Psychometrika*, 31:1–10, 1966.

J. Sirignano and K. Spiliopoulos. DGM: A deep learning algorithm for solving partial differential equations. *Journal of Computational Physics*, 375:1339–1364, 2018. doi: 10.1016/j.jcp.2018.08.029.

N. Stamatopoulos, D. J. Egger, Y. Sun, C. Zoufal, R. Iten, N. Shen, and S. Woerner. Option pricing using quantum computers. *Quantum*, 4:291, 2020. doi: 10.22331/q-2020-07-06-291.

N. Sukumar and A. Srivastava. Exact imposition of boundary conditions with distance functions in physics-informed deep neural networks. *Computer Methods in Applied Mechanics and Engineering*, 389:114333, 2022. doi: 10.1016/j.cma.2021.114333.

E. van Tegelen, T. Kapoor, G. A. K. van Voorn, P. van Heijster, and I. N. Athanasiadis. Late fusion neural operators for extrapolation across parameter space in partial differential equations. arXiv:2604.16721, 2026. URL https://arxiv.org/abs/2604.16721.

R. Wang, X. Zhong, Z. Xia, and J. Yan. Quantum neural operators with implicit quadratic frame and expressivity advantages. *Nature Machine Intelligence*, 2026. doi: 10.1038/s42256-026-01289-7.

S. Wang, Y. Teng, and P. Perdikaris. Understanding and mitigating gradient flow pathologies in physics-informed neural networks. *SIAM Journal on Scientific Computing*, 43(5):A3055–A3081, 2021a. doi: 10.1137/20M1318043.

S. Wang, H. Wang, and P. Perdikaris. Learning the solution operator of parametric partial differential equations with physics-informed DeepONets. *Science Advances*, 7(40):eabi8605, 2021b. doi: 10.1126/sciadv.abi8605.

S. Wang, X. Yu, and P. Perdikaris. When and why PINNs fail to train: A neural tangent kernel perspective. *Journal of Computational Physics*, 449:110768, 2022. doi: 10.1016/j.jcp.2021.110768.

R. Wiedemann, A. Jacquier, and L. Gonon. Operator deep smoothing for implied volatility. In *International Conference on Learning Representations*, 2025. URL https://arxiv.org/abs/2406.11520. arXiv:2406.11520.

P. Xiao, M. Zheng, A. Jiao, X. Yang, and L. Lu. Quantum DeepONet: Neural operators accelerated by quantum computing. *Quantum*, 9:1761, 2025. doi: 10.22331/q-2025-06-04-1761.

W. Zhang, Y. Guo, and B. Lu. Exact terminal condition neural network for American option pricing based on the Black–Scholes–Merton equations. *Journal of Computational and Applied Mathematics*, 480:117253, 2026. doi: 10.1016/j.cam.2025.117253.

A ANALYTICAL DETAILS AND PROOFS

The analysis separates properties of the pricing operator from properties of the implemented predictor. Appendices A.1 and A.2 establish scale symmetry of the exact pricing operator and pointwise admissibility of the scalar output map; Appendices A.3 and A.4 analyze the correction and its transfer through the output stages. Appendix A.5 gives a terminal-scaled smoothing construction whose empirical performance is not evaluated here.

A.0 FINANCIAL TERMINOLOGY

The terms below follow the pricing conventions used in this paper (Black & Scholes, 1973; Merton, 1973; Dupire, 1994). They distinguish the financial contract, its PDE inputs, and the market-data quantities.

Term	Meaning in this paper
Underlying / spot S	The asset on which the option is written; S is its current price.
European call	A contract giving the holder the right, but not the obligation, to buy the asset at price K only at expiry T.
Strike K	The purchase price fixed by the option contract.
Maturity T / remaining time τ	T is the expiry time; $\tau = T - t$ is the time left from valuation time t.
Payoff	The cash value at expiry: $(S-K)^+ = \max(S-K, 0)$ for the call considered here.
Risk-free rate r	The constant continuously compounded rate used for discounting; the discounted strike is $Ke^{-r\tau}$.
Local volatility σ / variance a	$\sigma(S,t)$ sets the instantaneous scale of proportional price fluctuations; $a(S,t) = \sigma^2(S,t)$.
Strike-line integrated variance A_0	The time integral of the input local variance evaluated at $S = K$, as in (2.3); not the integral of σ.
Delta Δ / Gamma Γ	The first and second price derivatives in S: $\partial_S V$ and $\partial_{SS} V$. They describe sensitivity and curvature.
Hedging	Taking offsetting positions to reduce financial risk; accurate price sensitivities are relevant to this task.
Moneyness / at the money (ATM)	The synthetic benchmark uses $m = S/K$, with ATM near $m = 1$. Market tests use their stated forward-based convention.
No-arbitrage bounds	Necessary pointwise price restrictions (2.2); satisfying them alone does not ensure cross-spot or cross-contract consistency.
Implied volatility (IV)	The constant volatility that matches an observed option price in a specified pricing formula; it is not a local-volatility field.
Calibration	Inferring pricing-model parameters or a volatility function from market quotes; distinct from evaluating a trained pricing surrogate.
Dividend yield q / forward F	The market-input convention uses $F = Se^{(r-q)\tau}$ and effective spot $Se^{-q\tau}$; see Appendix E.
Bid, offer, mid, spread	Bid is the price quoted by a buyer; offer is the price quoted by a seller. Mid is their average, and spread is offer minus bid.

Notation shared with the main text. We reserve $a = \sigma^2$ for local variance, $x = \log(S/K)$ for log-moneyness, $m = S/K$ for moneyness, $\bar{t} = t/T$ for normalized elapsed time, and $\chi = \tau/T = 1 - \bar{t}$ for remaining-life fraction. $W = V - C^{\mathrm{ref}}$ is the exact price correction; $R^* = W/\chi$ and $R_\theta = Kr_\theta$ have price units, while $r_\theta = \mu + s_y f_\theta$ is dimensionless. z_θ denotes the reconstructed pre-FAL price and $\widehat{V}_{\mathrm{FI}}$ the terminal-corrected FI price before optional PP. The Greek errors refer to physical-spot derivatives. Bold $\mathbf{W}$ in Appendix D denotes a weight matrix, not the scalar correction W. The exponent α always denotes maturity rescaling; coverage uses ρ and ω. The main-text P_a and appendix $\mathcal{P}_a$ denote the same pricing operator in (A.2).

A.1 PRICING, NORMALIZATION, AND EXACT SCALE SYMMETRY

For a zero-dividend asset, the risk-neutral dynamics are $dS_u = rS_u\,du + S_u\sigma(S_u,u)\,dW_u$, with $a = \sigma^2$. Risk-neutral valuation is

$$V_a^K(S,t) = e^{-r(T-t)}\mathbb{E}^{\mathbb{Q}}\big[(S_T - K)^+ \mid S_t = S\big]. \tag{A.1}$$

The backward equation is $V_t + \frac{1}{2}a(S,t)S^2V_{SS} + rSV_S - rV = 0$. With $v(S,\tau) = V(S, T-\tau)$, define

$$\mathcal{P}_a = \partial_\tau - \mathcal{A}_{a,\tau}, \qquad \mathcal{A}_{a,\tau} = \tfrac{1}{2}a(S,T-\tau)S^2\partial_{SS} + rS\partial_S - r. \tag{A.2}$$

Then $\mathcal{P}_a v = 0$ and $v(S,0) = (S-K)^+$. Whenever $\mathcal{P}_a$ acts on V, C^{ref}, or their corrections, these functions are understood in time-reversed coordinates. Delta and Gamma are $\Delta = V_S$ and $\Gamma = V_{SS}$. Classical Gamma at the terminal strike is not an evaluation target. Cross-strike and calendar-spread restrictions require consistent contract families; they are distinct from the pointwise interval in (2.2). The operator-learning target is $(a, K, T, r) \mapsto V_a^K(\cdot,\cdot)$, with function-valued a. The finite-dimensional sampled family in Appendix B.2 is narrower than this mathematical input class. These pricing formulations follow the classical risk-neutral and local-volatility constructions (Black & Scholes, 1973; Merton, 1973; Dupire, 1994; Derman & Kani, 1994).

For $m = S/K$, $\bar{t} = t/T$, $c(m,\bar{t}) = V(Km, T\bar{t})/K$, the equation and sensitivities are

$$c_{\bar{t}} + T\left[\tfrac{1}{2}a(Km,T\bar{t})m^2c_{mm} + rmc_m - rc\right] = 0, \qquad \Delta = c_m, \quad \Gamma = K^{-1}c_{mm}. \tag{A.3}$$

The variable $\chi = \tau/T = 1 - \bar{t}$ is remaining, not elapsed, normalized time. All reported synthetic training uses $T = 1$; normalization alone is not an unrestricted maturity-invariance result.

Proposition A.1 (Joint spot–strike–field symmetry)**.** *For $\lambda > 0$, set $a_\lambda(S,u) = a(S/\lambda, u)$. Then*

$$V_{a_\lambda}^{\lambda K}(\lambda S,t) = \lambda V_a^K(S,t), \quad \Delta_{a_\lambda}^{\lambda K}(\lambda S,t) = \Delta_a^K(S,t), \quad \Gamma_{a_\lambda}^{\lambda K}(\lambda S,t) = \lambda^{-1}\Gamma_a^K(S,t), \tag{A.4}$$

where the derivatives exist.

Proof. If S_u solves the original SDE, then $\widetilde{S}_u = \lambda S_u$ solves the transformed model because $a_\lambda(\widetilde{S}_u, u) = a(S_u, u)$. Its payoff is $(\widetilde{S}_T - \lambda K)^+ = \lambda(S_T - K)^+$. Discounted expectation and uniqueness give the price identity. Differentiating with respect to physical spot gives the Delta and Gamma identities. □

A trained network inherits this identity only if its inputs and normalization statistics transform compatibly. The implementation retains absolute payoff samples and the unnormalized K, with fixed training standardization. If $\mathrm{Std}(y) = (y - \mu_{\mathrm{tr}})/s_{\mathrm{tr}}$, then

$$\mathrm{Std}(\lambda y) - \mathrm{Std}(y) = (\lambda - 1)y/s_{\mathrm{tr}}. \tag{A.5}$$

Thus the implemented model is scale-aware, not exactly equivariant. Equation (A.4) is an operator property, not a learned OOD theorem.

A.2 Admissibility, stability, and units

For $\delta > 0$ and $\ell < u$, define the softplus map and the Financial Admissibility Layer using stable softplus evaluation:

$$\mathrm{sp}_\delta(z) = \delta\log(1 + e^{z/\delta}), \qquad \mathcal{C}_\delta(z;\ell,u) = \ell + \mathrm{sp}_\delta(z-\ell) - \mathrm{sp}_\delta(z-u). \tag{A.6}$$

We use dimensionless smoothing $\delta_v = 0.002$, corresponding to $K\delta_v$ in price units. The complete FI output is

$$\widehat{V}_{\mathrm{FI}}(S,t) = \begin{cases} 0, & S = 0, \\ (S-K)^+, & t = T, \\ \mathcal{C}_{K\delta_v}(z_\theta; L, U), & S > 0,\ t < T. \end{cases} \tag{A.7}$$

This is the post-FAL, terminal-corrected output before PP. Historical continuation training applied ordinary FAL even at terminal nodes; (A.7) was evaluated later as a parameter-free analytic assignment using unchanged selected weights, no held-out outcomes, and no retraining (Appendix B.4). Appendix A.5 studies a time-scaled smoothing alternative, but it was not trained here.

Proposition A.2 (Scalar FAL properties)**.** *For $\ell < u$, finite z, and $\delta > 0$, the map in* (A.6) *satisfies*

$$\ell < \mathcal{C}_\delta(z;\ell,u) < u, \qquad 0 < \partial_z\mathcal{C}_\delta < 1, \tag{A.8}$$

$$|\mathcal{C}_\delta(z_1;\ell,u) - \mathcal{C}_\delta(z_2;\ell,u)| \le |z_1 - z_2|, \tag{A.9}$$

$$|\mathcal{C}_\delta(z;\ell,u) - P_{[\ell,u]}(z)| \le \delta\log 2, \qquad \mathcal{C}_{\lambda\delta}(\lambda z;\lambda\ell,\lambda u) = \lambda\mathcal{C}_\delta(z;\ell,u). \tag{A.10}$$

For $v \in [\ell,u]$, $|\mathcal{C}_\delta(z;\ell,u) - v| \le |z - v| + \delta\log 2$.

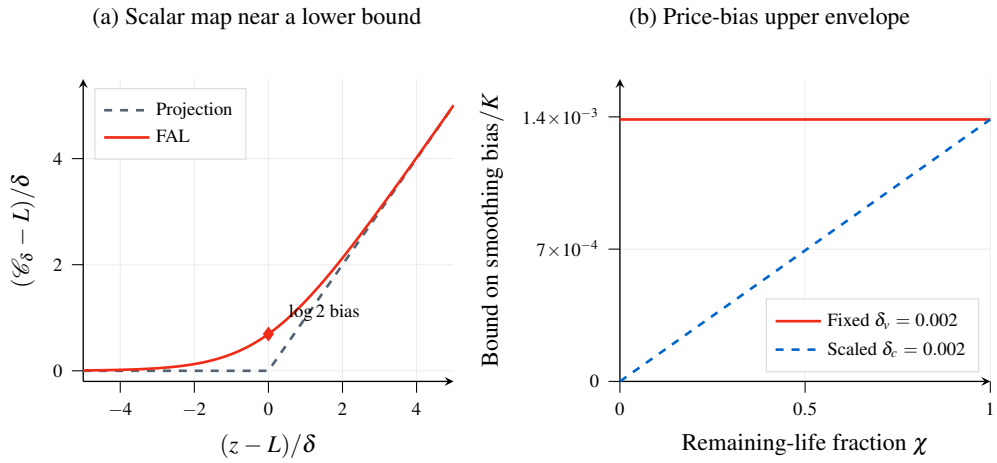


Figure 5: Analytic FAL diagnostics. (a) The scalar map and hard projection on an illustrative interval with $(U - L)/\delta = 20$; the lower-bound bias approaches $\delta \log 2$. This local coordinate plot is not a price-versus-spot surface. (b) The bounds $\delta_v \log 2$ and $\chi\delta_c \log 2$ on normalized price smoothing error, with equal illustrative coefficients $\delta_c = \delta_v = 0.002$. The fixed-width curve concerns the nonterminal output: an exact assignment at $\chi = 0$ does not remove its one-sided limit bias. Terminal-scaled FAL is the untested theoretical refinement of Appendix A.5, not a new trained model.

Proof. Put $A = \mathrm{sigm}((z - \ell)/\delta)$, $B = \mathrm{sigm}((z - u)/\delta)$, where $\mathrm{sigm}(s) = (1 + e^{-s})^{-1}$. Since $0 < B < A < 1$,

$$\partial_z \mathcal{C} = A - B, \qquad \partial_\ell \mathcal{C} = 1 - A, \qquad \partial_u \mathcal{C} = B. \tag{A.11}$$

These derivatives are positive and sum to one. The limits as $z \to -\infty, +\infty$ are ℓ, u, respectively, so strict monotonicity proves the interval result. The mean-value theorem proves scalar nonexpansiveness; integrating the joint gradient also gives one-Lipschitz continuity in the maximum norm along admissible triples. Since $P_{[\ell,u]}(z) = \ell + (z - \ell)^+ - (z - u)^+$ and $0 \le \mathrm{sp}_\delta(q) - q^+ \le \delta \log 2$, their difference is the difference of two numbers in $[0, \delta \log 2]$, not the sum of their absolute upper bounds. This proves the single-$\delta \log 2$ constant. Projection nonexpansiveness gives the error bound relative to v. Finally, $\mathrm{sp}_{\lambda\delta}(\lambda q) = \lambda\, \mathrm{sp}_\delta(q)$ proves homogeneity. □

As a function of the independent scalar arguments (z, ℓ, u), FAL is a smooth approximation to projection, not a retraction that fixes every interior price. This scalar smoothness does not imply global smoothness of the composed price as a function of spot when $\ell = L(S, \tau)$ is kinked. The dimensionless parameter δ_v acting on V/K corresponds to $K\delta_v$ in price units. At maturity, the payoff equals L and

$$\mathcal{C}_\delta(L; L, U) - L = \delta \log 2 - \delta \log\left(1 + e^{-(U-L)/\delta}\right) > 0 \quad (L < U). \tag{A.12}$$

The explicit terminal assignment in (A.7) does not remove this nonterminal limit bias for fixed δ. At $S = 0$, $L = U = 0$, and the output is assigned zero directly rather than invoking the strict-interval proposition.

On a margin region $z - \ell \ge m_0 > 0$, $u - z \ge m_0$,

$$|\mathcal{C}_\delta(z) - z| \le 2\delta e^{-m_0/\delta}, \quad 1 - \partial_z \mathcal{C} \le 2e^{-m_0/\delta}, \quad \partial_\ell \mathcal{C}, \partial_u \mathcal{C} \le e^{-m_0/\delta}. \tag{A.13}$$

For $q = (z, \ell, u)$, its Hessian is

$$\nabla_q^2 \mathcal{C} = h_\ell v_\ell v_\ell^\top - h_u v_u v_u^\top, \quad v_\ell = (1, -1, 0)^\top, \quad v_u = (1, 0, -1)^\top, \tag{A.14}$$

where $h_\ell = A(1 - A)/\delta$, $h_u = B(1 - B)/\delta$, each bounded by $e^{-m_0/\delta}/\delta$ on that region. Small scalar distortion therefore need not imply small derivative distortion near an active bound.

Figure 5 isolates the scalar smoothing effect and its terminal implication; neither panel is a trained-model experiment.

A.3 Correction equation, cancellation, and differentiated profiles

With A_0 from (2.3), the strike-matched carrier for $S > 0$ and $A_0(\tau) > 0$ is

$$C^{\mathrm{ref}}(S,t) = S\Phi(d_+) - Ke^{-r\tau}\Phi(d_-),$$
$$d_+ = \frac{\log(S/K) + r\tau + A_0(\tau)/2}{\sqrt{A_0(\tau)}}, \qquad d_- = d_+ - \sqrt{A_0(\tau)}. \tag{A.15}$$

Here Φ is the standard normal CDF. The carrier exactly prices the spatially frozen, time-dependent variance $a(K,t)$. Its limiting assignments are $C^{\mathrm{ref}}(S,T) = (S-K)^+$ and $C^{\mathrm{ref}}(0,t) = 0$.

Assumptions and correction equation. Set $x = \log(S/K)$ and $\bar{a}(x,\tau) = a(Ke^x, T-\tau)$. Contract parameters range over a compact class with K, T bounded away from zero. For value estimates, assume a full-line extension with $0 < a_- \le \bar{a} \le a_+ < \infty$, uniform spatial Lipschitz continuity, and unique pricing/correction solutions with comparison in an at-most-exponential-growth class. For differentiated profiles, additionally assume a smooth extension with $\bar{a} \in C_b^{3+\zeta,1+\zeta/2}$, $0 < \zeta < 1$. A smooth neighborhood of the strike suffices by localization; clipping elsewhere need not be differentiable, as detailed below. Constants below are uniform over the stated compact contract class. The stronger differentiated assumptions concern the smooth extension, not every clipped production coefficient globally.

The exact correction $W = V - C^{\mathrm{ref}}$ satisfies, wherever classical derivatives exist,

$$\mathcal{P}_a W = f, \qquad W(S,T) = 0,$$
$$f = \tfrac{1}{2}[a(S,T-\tau) - a(K,T-\tau)]S^2 C^{\mathrm{ref}}_{SS}. \tag{A.16}$$

Proof of Proposition 3.1. The carrier solves the equation with $a(S,t)$ replaced by $a(K,t)$. Subtracting its forward equation from the full price equation gives (A.16). Set $c^{\mathrm{ref}}(x,\tau) = C^{\mathrm{ref}}(Ke^x, T-\tau)/K$ and $w = W(Ke^x, T-\tau)/K$. Then

$$(\partial_{xx} - \partial_x)c^{\mathrm{ref}} = \frac{e^x\phi(d_+)}{\sqrt{A_0(\tau)}}, \tag{A.17}$$

$$\partial_\tau w = \tfrac{1}{2}\bar{a}\, w_{xx} + (r - \bar{a}/2)w_x - rw + \widetilde{f},$$
$$\widetilde{f} = \tfrac{1}{2}[\bar{a}(x,\tau) - \bar{a}(0,\tau)](\partial_{xx} - \partial_x)c^{\mathrm{ref}}. \tag{A.18}$$

Here $\phi = \Phi'$. Uniform ellipticity gives $a_-\tau \le A_0(\tau) \le a_+\tau$. Completing the square in $e^x\phi(d_+)$, with $r\tau + A_0/2 = O(\tau)$, yields

$$0 \le (\partial_{xx} - \partial_x)c^{\mathrm{ref}} \le C\tau^{-1/2}e^{-c_0x^2/\tau}. \tag{A.19}$$

Lipschitz continuity implies $|\widetilde{f}| \le C|x|\tau^{-1/2}e^{-c_0x^2/\tau} \le C_f$. Duhamel/comparison with zero initial correction gives

$$\|w(\cdot,\tau)\|_\infty \le \int_0^\tau e^{|r|(\tau-s)}C_f\, ds \le C\tau. \tag{A.20}$$

The full-line estimate concerns the declared extension. A bounded-domain analogue must also include differences in lateral boundary data. This proves the stated value bound. □

Alternative coupling estimate. For initial spots in a fixed compact interval containing K, assume bounded spot-Lipschitz σ, a globally Lipschitz extension of $s \mapsto s\sigma(s,u)$, and finite required fourth moments, uniformly over the contract class. Drive the true process X and strike-frozen process Y from common initial S with the same Brownian motion. The latter has discounted call value C^{ref}. Its geometric representation gives, for $q = 2, 4$,

$$\mathbb{E}\sup_{t\le v\le u}|Y_v - K|^q \le C_q\big(|S-K|^q + (u-t)^{q/2}\big). \tag{A.21}$$

For $D = X - Y$, split the diffusion discrepancy as

$$X_u\sigma(X_u,u) - Y_u\sigma(K,u) = [X_u\sigma(X_u,u) - Y_u\sigma(Y_u,u)]$$
$$+ Y_u[\sigma(Y_u,u) - \sigma(K,u)]. \tag{A.22}$$

Stochastic integral estimates and Gronwall reduce the mean-square difference to the integral of $\mathbb{E}[Y_u^2|Y_u - K|^2]$. Hölder and the fourth-moment bounds give $\mathbb{E}\sup_{t\le u\le T}|D_u|^2 \le C(\tau|S-K|^2 + \tau^2)$. The payoff is one-Lipschitz, hence

$$|V(S,t) - C^{\mathrm{ref}}(S,t)| \le C\big(\sqrt{\tau}|S-K| + \tau\big). \tag{A.23}$$

This is an $O(\tau)$ value estimate on a parabolic strike neighborhood; it does not permit derivative estimates by differentiation of its right-hand side.

Approximate strike-line integration. Let $\widetilde{A}_0$ be absolutely continuous with $a_-\tau \le \widetilde{A}_0(\tau) \le a_+\tau$ and $\widetilde{a}_K = \partial_\tau \widetilde{A}_0$ almost everywhere. Its carrier $\widetilde{C}^{\mathrm{ref}}$ satisfies

$$\mathcal{P}_a(V - \widetilde{C}^{\mathrm{ref}}) = \tfrac{1}{2}[a(S,T-\tau) - \widetilde{a}_K(\tau)]S^2\widetilde{C}^{\mathrm{ref}}_{SS}. \tag{A.24}$$

Separate spatial mismatch from line error $\varepsilon_K(q) = |\widetilde{a}_K(q) - a(K,T-q)|$. The preceding Gaussian bound yields

$$\sup_x \frac{|V - \widetilde{C}^{\mathrm{ref}}|}{K} \le C\tau + C\int_0^\tau \varepsilon_K(q)q^{-1/2}\,dq \le C(\tau + \bar{\varepsilon}_K\sqrt{\tau}) \tag{A.25}$$

when $\varepsilon_K \le \bar{\varepsilon}_K$. A fixed nonzero effective line error can contaminate the remainder at $O(\sqrt{\tau})$. Generator-based strike evaluation avoids sensor interpolation error but is not a guarantee for arbitrary approximate quadrature.

Full statement of Proposition 3.2. Under the differentiated assumptions, put $a_* = \bar{a}(0,0)$, $b_* = \partial_x\bar{a}(0,0)$, and let g_{a_*} be the centered Gaussian density of variance a_*. In spot coordinates, $b_* = K\partial_S a(K,T)$, so its nonvanishing is the nonzero terminal variance-slope condition used in the main text. Along $S_\tau = Ke^{y\sqrt{\tau}}$, for fixed y,

$$\begin{aligned}
\frac{W(S_\tau, T-\tau)}{K\tau} &\longrightarrow G_0(y) = \tfrac{b_*}{4}yg_{a_*}(y),\\
\frac{W_S(S_\tau, T-\tau)}{\sqrt{\tau}} &\longrightarrow G_1(y) = \tfrac{b_*}{4}(1 - y^2/a_*)g_{a_*}(y),\\
KW_{SS}(S_\tau, T-\tau) &\longrightarrow G_2(y) = \tfrac{b_*}{4}(y^3/a_*^2 - 3y/a_*)g_{a_*}(y).
\end{aligned} \tag{A.26}$$

Along the same rays, $\sqrt{\tau}KC^{\mathrm{ref}}_{SS} \to g_{a_*}$. For $b_*y \neq 0$, $W/(\tau/T)^\alpha$ has a bounded nonzero limit precisely at $\alpha = 1$. The criticality concerns the value target, not global unweighted C^2 regularity. Generically $R^* = O(1)$, $R^*_S = O(\tau^{-1/2})$, and $KR^*_{SS} = O(\tau^{-1})$ on these rays. If $b_* = 0$, the uniqueness claim for this exponent does not apply. Manufactured reference calculations support the scaled profiles, not a theorem about training (Appendix C.1).

Proof of Proposition 3.2. Put $W_\varepsilon(y,s) = \varepsilon^{-2}w(\varepsilon y, \varepsilon^2 s)$. Equation (A.18) becomes

$$\begin{aligned}
\partial_s W_\varepsilon = \tfrac{1}{2}\bar{a}(\varepsilon y, \varepsilon^2 s)\partial_{yy}W_\varepsilon + \varepsilon\big(r - \bar{a}(\varepsilon y, \varepsilon^2 s)/2\big)\partial_y W_\varepsilon\\
- \varepsilon^2 r W_\varepsilon + \widetilde{f}(\varepsilon y, \varepsilon^2 s).
\end{aligned} \tag{A.27}$$

Taylor expansion gives $\bar{a}(\varepsilon y, \varepsilon^2 s) - \bar{a}(0, \varepsilon^2 s) = \varepsilon b_* y + o(\varepsilon)$ and $A_0(\varepsilon^2 s)/\varepsilon^2 \to a_* s$. Consequently $\widetilde{f}(\varepsilon y, \varepsilon^2 s) \to (b_*/2)yg_{a_* s}(y)$ in a local parabolic Hölder norm away from $s = 0$. The value bound gives $|W_\varepsilon(y,s)| \le Cs$. Interior Schauder estimates and compactness produce locally $C^{2,1}$-convergent subsequences on $|y| \le M$, $s_0 \le s \le 1$. Each limit has zero initial trace and solves

$$\partial_s W_0 = \tfrac{1}{2}a_*\partial_{yy}W_0 + \tfrac{1}{2}b_* y g_{a_* s}(y), \qquad W_0(y,0) = 0. \tag{A.28}$$

Uniqueness in the bounded mild-solution class identifies the whole limit. The Gaussian convolution identity

$$\int_{\mathbb{R}} g_{a_*(s-q)}(y-z)\, z g_{a_* q}(z)\, dz = \frac{q}{s}\, y g_{a_* s}(y) \tag{A.29}$$

evaluates Duhamel's integral as $W_0(y,s) = (b_*/4)syg_{a_* s}(y)$. Taking $s = 1$, $\varepsilon = \sqrt{\tau}$, and using

$$W_S = e^{-x}w_x, \qquad KW_{SS} = e^{-2x}(w_{xx} - w_x) \tag{A.30}$$

proves all three profiles in (A.26). The carrier limit follows directly from $A_0(\tau)/\tau \to a_*$. Finally,

$$\frac{W(S_\tau, T-\tau)}{(\tau/T)^\alpha} = KT^\alpha \tau^{1-\alpha}\big(G_0(y) + o(1)\big). \tag{A.31}$$

For $b_* y \neq 0$, $G_0(y) \neq 0$, which gives a vanishing, finite nonzero, or divergent limit according as $\alpha < 1$, $\alpha = 1$, or $\alpha > 1$. □

Figure 6 visualizes the universal dimensionless shapes obtained from (A.26). It separates bounded correction values from the derivative scales that remain singular after rescaling.

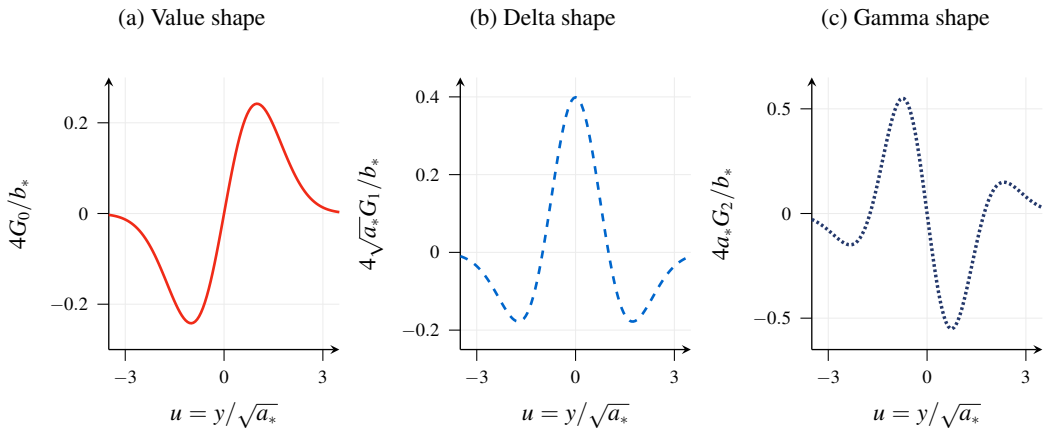


Figure 6: Dimensionless first-order skew profiles from (A.26), for $b_* \neq 0$. With $u = y/\sqrt{a_*}$ and standard normal density ϕ, the three curves are $u\phi(u)$, $(1-u^2)\phi(u)$, and $(u^3-3u)\phi(u)$. They correspond to the limits of $W/(K\tau)$, $W_S/\sqrt{\tau}$, and KW_{SS} after the displayed amplitude normalizations. These are exact analytic shapes, not fitted curves or empirical convergence measurements. Nonzero-target criticality requires $b_* y \neq 0$; the value profile vanishes at $y = 0$.

Localization near an unclipped strike. Suppose the value assumptions hold globally and the coefficient is smooth on $|x| < d$ for small remaining times. Choose a smooth, bounded, uniformly elliptic extension agreeing there. Its strike-line carrier is identical. The difference of the two normalized prices solves the same homogeneous parabolic equation inside the neighborhood, with zero initial data and bounded lateral data. A bounded-drift, bounded-volatility log diffusion started in $|x| \le d/2$ exits $|x| < d$ before τ with probability at most $Ce^{-c/\tau}$. The stopped representation bounds the value difference by the same scale. Interior derivative estimates add only powers of τ^{-1}, so the first two derivative differences vanish faster than the scales in (A.26) on $x = y\sqrt{\tau}$. Thus the local profiles survive clipping away from the strike. This is not a discretization or finite-domain error estimate.

A.4 Error transfer, full FAL derivatives, and PDE residuals

Here PDE residual refers to the governing-equation defect $\mathcal{P}_a u$; it is distinct from the carrier-relative correction learned by the network. For correction error $\rho_\theta = R_\theta - R^*$, define the weighted norm on the nonterminal domain Ω under consideration:

$$N_\rho = \sup_{\Omega}\left(\frac{|\rho_\theta|}{K} + \sqrt{\chi}|\rho_{\theta,S}| + \chi K|\rho_{\theta,SS}|\right), \quad \chi > 0. \tag{A.32}$$

Equation (3.3) follows from $z_\theta - V = \chi\rho_\theta$ and the independence of χ from S. Each term in (A.32) bounds the corresponding correction error, giving the three inequalities separately when $N_\rho < \infty$. Finite checks of these identities cannot establish a uniform bound on the learned norm. In particular, the maturity gate does not force Gamma error to vanish: for a fixed smooth decoder, $\chi K R_{\theta,SS}$ can vanish while KW_{SS} has a nonzero limit. At nonterminal points, FAL gives the global price estimate $|\widehat{V}_{\mathrm{FI}} - V| \le \chi|\rho_\theta| + K\delta_v \log 2$. Its Greek derivatives contain bound-derivative and Hessian terms, so pre-FAL Greek guarantees do not transfer globally.

Before FAL, the exact PDE-residual identity is

$$\mathcal{P}_a z_\theta = -f + \chi \mathcal{P}_a R_\theta + R_\theta/T = \chi \mathcal{P}_a \rho_\theta + \rho_\theta/T. \tag{A.33}$$

This identity connects the correction target to physics information without a PDE loss; it does not guarantee successful optimization. Post-FAL differentiation adds the curvature and smoothing terms derived below.

For fixed price-unit smoothing δ, define $F = \mathcal{C}_\delta(z_\theta; L, U)$, $A = \mathrm{sigm}((z_\theta - L)/\delta)$, $B = \mathrm{sigm}((z_\theta - U)/\delta)$, $p = A - B$, $q = 1 - A$, $s = B$, $h_L = A(1-A)/\delta$, and $h_U = B(1-B)/\delta$. On a smooth region,

$$F_S = p z_{\theta,S} + q L_S + s U_S, \tag{A.34}$$

$$F_{SS} = p z_{\theta,SS} + q L_{SS} + s U_{SS} + h_L (z_{\theta,S} - L_S)^2 - h_U (z_{\theta,S} - U_S)^2. \tag{A.35}$$

These follow from (A.11)–(A.14). The composition need not be globally smooth in spot. For fixed $\tau > 0$, let $S_b = Ke^{-r\tau}$ be the kink of the lower bound. If $z_\theta(\cdot, t)$ is smooth and finite at S_b, then L_S jumps from 0 to 1 while p, q, s and the other smooth factors remain continuous, so

$$F_S(S_b^+, t) - F_S(S_b^-, t) = 1 - \mathrm{sigm}\left(\frac{z_\theta(S_b, t)}{\delta}\right) > 0. \tag{A.36}$$

Thus the post-FAL predictor is generally only piecewise C^1 in spot; this does not invalidate grid-based error metrics, but it prevents interpreting scalar softplus smoothness as a global Greek-smoothness guarantee. On a margin region $z_\theta - L \geq m_0$, $U - z_\theta \geq m_0$,

$$|F_S - V_S| \leq \chi |\rho_{\theta,S}| + e^{-m_0/\delta}(|L_S - V_S| + |U_S - V_S|), \tag{A.37}$$

$$\begin{aligned} |F_{SS} - V_{SS}| \leq{}& \chi |\rho_{\theta,SS}| + e^{-m_0/\delta}(|L_{SS} - V_{SS}| + |U_{SS} - V_{SS}|) \\ &+ \frac{e^{-m_0/\delta}}{\delta}\big[(z_{\theta,S} - L_S)^2 + (z_{\theta,S} - U_S)^2\big]. \end{aligned} \tag{A.38}$$

Subtract $V_S = (p + q + s)V_S$ and its second-derivative analogue, then use $p \leq 1$, $q, s \leq e^{-m_0/\delta}$. These are local, not global, Greek bounds: L is not C^2 at $S = Ke^{-r\tau}$ and no positive margin is uniform at maturity.

For time-dependent $\delta_e(\tau) > 0$ and $b = \frac{1}{2}aS^2$, the full post-FAL PDE residual on smooth nonterminal regions is

$$\begin{aligned} \mathcal{P}_a F = p \mathcal{P}_a z_\theta - b\big[h_L (z_{\theta,S} - L_S)^2 - h_U (z_{\theta,S} - U_S)^2\big] \\ + \big(\delta_e'(\tau) + r\delta_e\big)\partial_\delta \mathcal{C}_{\delta_e}(z_\theta; L, U). \end{aligned} \tag{A.39}$$

Indeed, for $q_i \in \{z_\theta, L, U\}$ and corresponding weights p_i, differentiation gives $F_\tau = \sum_i p_i q_{i,\tau} + \delta_e' \partial_\delta \mathcal{C}$ and $F_{SS} = \sum_i p_i q_{i,SS} + q_S^\top \nabla_q^2 \mathcal{C} q_S$. Homogeneity yields $F - \sum_i p_i q_i = \delta_e \partial_\delta \mathcal{C}$. On either smooth branch of L, $\mathcal{P}_a L = \mathcal{P}_a U = 0$, proving (A.39). For the implemented constant $\delta_e = K\delta_v$, the $r\delta_e$ term remains. Product differentiation of $C^{\mathrm{ref}} + \chi R_\theta$ gives (A.33) because $\mathcal{P}_a C^{\mathrm{ref}} = -f$ and $\chi' = 1/T$. Discrete PP is outside these differentiation identities.

Interior recovery requires a continuum norm. Let $u(x, \tau) = V(Ke^x, T - \tau)/K$ and $\widehat{u}$ be a prediction at one specified output stage. On nested parabolic cylinders $Q' \Subset Q$, separated from maturity, boundaries, and output kinks, assume uniform ellipticity, $C^{\zeta,\zeta/2}$ coefficients, bounded drift/reaction, and $u, \widehat{u} \in C^{2+\zeta,1+\zeta/2}(Q)$. The interior estimate (Friedman, 1964; Lieberman, 1996) is

$$\|\widehat{u} - u\|_{C^{2+\zeta,1+\zeta/2}(Q')} \leq C\left(\|\widehat{u} - u\|_{C^0(Q)} + \|\mathcal{L}_a \widehat{u}\|_{C^{\zeta,\zeta/2}(Q)}\right), \tag{A.40}$$

where $\mathcal{L}_a = \partial_\tau - \frac{1}{2}\bar{a}\partial_{xx} - (r - \bar{a}/2)\partial_x + r$. This follows by applying the parabolic estimate to $e = \widehat{u} - u$, which satisfies $\mathcal{L}_a e = \mathcal{L}_a \widehat{u}$. Since $\Delta = e^{-x} u_x$ and $K\Gamma = e^{-2x}(u_{xx} - u_x)$, bounded log cylinders also give

$$\sup_{Q'}\left(\frac{|\widehat{V} - V|}{K} + |\widehat{\Delta} - \Delta| + K|\widehat{\Gamma} - \Gamma|\right) \leq C\left(\|e\|_{C^0(Q)} + \|\mathcal{L}_a \widehat{u}\|_{C^{\zeta,\zeta/2}(Q)}\right). \tag{A.41}$$

A scalar FD RMS on one grid controls neither the continuum supremum nor the Hölder seminorm, and it also includes truncation error. It cannot certify this premise.

A.5 TERMINAL-SCALED FAL: A THEORETICAL REFINEMENT, NOT A TESTED MODEL

For a dimensionless $\delta_c > 0$ and $\chi > 0$, define

$$L_R = (L - C^{\mathrm{ref}})/\chi, \quad U_R = (U - C^{\mathrm{ref}})/\chi, \qquad \widetilde{V}_\theta = C^{\mathrm{ref}} + \chi \mathcal{C}_{K\delta_c}(R_\theta; L_R, U_R). \tag{A.42}$$

Translation and homogeneity show $\widetilde{V}_\theta = \mathcal{C}_{K\chi\delta_c}(z_\theta; L, U)$. Assign the payoff at $\chi = 0$ and zero at $S = 0$.

Proposition A.3 (Conditional value-level closure)**.** *For $\chi > 0$ and $L < U$, $L < \widetilde{V}_\theta < U$, and*

$$\frac{|\widetilde{V}_\theta - V|}{K} \le \chi \left(\frac{|\rho_\theta|}{K} + \delta_c \log 2 \right). \tag{A.43}$$

If R_θ/K is locally bounded near maturity, $\widetilde{V}_\theta$ converges to the payoff in value. If in addition

$$\begin{aligned} R_\theta^{K,T}(S,t) &= K\mathcal{R}_\theta\big(\mathfrak{a}_{K,T}, rT, \textit{scale-invariant inputs}; S/K, \tau/T\big), \\ \mathfrak{a}_{K,T}(m,s) &= Ta(Km, T(1-s)). \end{aligned} \tag{A.44}$$

with dimensionless sensors and compatible statistics, the composed predictor inherits (A.4)*.*

Proof. Move χ through softplus to obtain the equivalent smoothing $K\chi\delta_c$, then apply (A.8)–(A.10). Boundedness of the decoder, $\chi \to 0$, and carrier convergence give the terminal limit. On the joint scaling orbit, dimensionless fields, contract combinations, and sensor data do not change; the decoder, carrier, bounds, and smoothing all scale by λ. Homogeneity proves price scaling, and differentiation gives the Greek identities where defined. □

This is not a theorem for the implemented dimensional branch or fixed $0.002K$ smoothing. It does not transfer pre-FAL weighted Greek convergence through the refined map: its Hessian grows like $(K\chi\delta_c)^{-1}$ near active bounds. Empirical claims would require a separate matched experiment.

B IMPLEMENTATION, DATA, AND EVALUATION PROTOCOLS

B.1 MODEL VOCABULARY AND ANALYTICAL SCOPE

Figure 1 summarizes the FI construction. Table 3 defines the model names used throughout the paper, and Table 4 separates mathematical assumptions from properties established for the implemented predictor.

Table 3: Model names and components. RA, FAL, and PP have distinct roles.

Name	Meaning
DeepONet	Deep Operator Network
Normalized DeepONet	DeepONet with contract normalization
FAL	Financial Admissibility Layer; smooth-in-raw-price pointwise call-price-bound map; spot composition is piecewise smooth
RA	Strike-matched carrier–residual ansatz
PP	Deterministic post-processing of a predicted grid
PI-DeepONet	Physics-informed DeepONet; canonical comparator has no RA/FAL/PP
FI-DeepONet	Normalization + RA + FAL; principal ID output uses terminal correction
FI-DeepONet-PP	The same selected FI checkpoint followed by PP
QO-FI-DeepONet	Exact compilation of selected frozen maps with orthogonal factors and a classical signed core
QDO-FI-DeepONet	Restricted diagonal–orthogonal map approximation followed by supervised adaptation

Table 4: Analytical statements and applicability. Finite verification does not establish a uniform theorem premise.

Result	Required regime	Implemented-model interpretation
Scale symmetry	Joint spot–strike–field transformation	Exact for the operator; learned model only scale-aware
FAL interval	Finite values, common units, $L < U$	Pointwise only; scalar map smooth, but spot composition inherits lower-bound kink
RA cancellation	Value-level assumptions or stated coupling hypotheses	Continuous canonical-field value bound; reference errors separate
First-order skew profile	Strong local smoothness; $b_* y \neq 0$ for criticality	Local near the strike; clipping away from it is handled by localization
Weighted transfer	Finite N_ρ in (A.32); pre-FAL stage	Exact identities; uniform learned norm unverified
Interior recovery	Same-stage PDE residual in Hölder norm	Not certified by scalar grid RMS
Terminal-scaled FAL	Smoothing $K\chi\delta_c$	Theoretical refinement, not a principal experiment
Post-FAL PDE residual	Smooth region and full chain rule	Excludes kinks, terminal line, and PP

B.2 FIXED COEFFICIENT FAMILY, SENSORS, AND SPLIT SUPPORT

We use European calls with $T = 1$ and zero dividends. The local-volatility family is

$$\sigma_{\mathrm{loc}}(S,t) = \mathrm{clip}_{[0.05,1]}\left\{\sigma_0\left[1+\beta\tanh\left(2\frac{S-K}{K}\right)\right]\left[1+\gamma\left(\frac{T-t}{T}-\frac{1}{2}\right)\right]\right\}. \tag{B.1}$$

The raw coefficient in (B.1) is clipped to $[0.05, 1.00]$ before squaring. Implementation safeguards replace K, T in denominators by $\max(K, 10^{-8})$, $\max(T, 10^{-8})$; these are inactive on the reported support. The 101 payoff sensors are equally spaced on $[0, 220]$. The 231 volatility values use 21 spot positions on $[0, 220]$ and 11 calendar-time positions on $[0, 1]$, stored time-major. The scalars are $(K, r, \sigma_0, \beta, \gamma)$. The reported generalization experiments evaluate held-out and parameter-support performance within this declared coefficient family; off-family functional response is assessed separately.

Table 5 lists the fixed support and split sizes; Table 6 decomposes the 320-surface OOD population into eight single-axis shifts and the joint-shift group.

Table 5: Fixed synthetic data support. All intervals are half-open. Validation and held-out ID share the training support.

Item	Specification
Surface counts	512 training; 32 validation; 32 held-out ID; 320 strict parameter-support OOD
Training/ID support	$K \in [55, 135)$, $r \in [0, 0.13)$, $\sigma_0 \in [0.10, 0.45)$, $\beta \in [-0.90, 0.60)$, $\gamma \in [-0.45, 0.50)$
Strike shifts	Low $K \in [40, 52)$; high $K \in [138, 160)$
Rate / volatility	High $r \in [0.14, 0.18)$; high $\sigma_0 \in [0.48, 0.60)$
Skew shifts	$\beta \in [-1.15, -0.95)$ or $[0.65, 0.80)$
Term shifts	$\gamma \in [-0.65, -0.50)$ or $[0.55, 0.70)$
Joint shifts	Two or three shifted coordinates under the recorded design
Unshifted coordinates	$K \in [70, 120)$, $r \in [0.03, 0.08)$, $\sigma_0 \in [0.20, 0.30)$, $\beta \in [-0.30, 0.10)$, $\gamma \in [-0.10, 0.15)$

Table 6: Disjoint parameter-support OOD design. Excursions are farthest proposal distances beyond the training boundary, normalized by its support width.

Regime	Count	Shifted support	Max. excursion
Low K	32	$[40, 52)$	18.75%
High K	32	$[138, 160)$	31.25%
High r	32	$[0.14, 0.18)$	38.46%
High σ_0	32	$[0.48, 0.60)$	42.86%
Negative β	32	$[-1.15, -0.95)$	16.67%
Positive β	32	$[0.65, 0.80)$	13.33%
Negative γ	32	$[-0.65, -0.50)$	21.05%
Positive γ	32	$[0.55, 0.70)$	21.05%
Joint	64	Two or three coordinates	Multi-axis
Total	320	Eight single-axis + joint	—

Parameter rows and their ordering are recorded in the project, but the original split-generation seed and complete sampling distribution are not established. Comparisons therefore use fixed realized data without asserting a uniform sampling law. The single-axis design shifts K, β, γ in both directions but r, σ_0 only upward. A separate 64-surface high-strike diagnostic lies *inside* training support and is not part of the parameter-support OOD evidence. Realized excursions need not reach proposal endpoints (Table 21).

The data-flow schematic in Figure 7 distinguishes the sampled network input from the carrier's direct coefficient access. The 410/102 realization split is nested within the original 512 training surfaces; it does not use the 32-surface development-validation or ID sets.

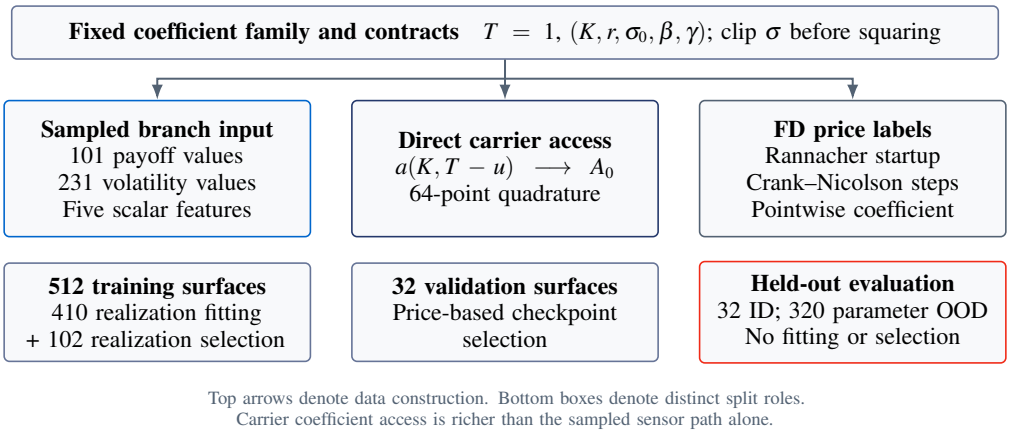


Figure 7: Data access and split roles. The carrier evaluates the generating coefficient on the strike line; it does not infer that line from the 231 volatility sensors. The original 512 training surfaces also supply the disjoint 410/102 fitting/selection split for quantum-compatible realization, whereas the 32-surface validation set selects the principal neural checkpoints. The bottom boxes specify roles, not a random resampling algorithm: the realized split rows are fixed and the original generation stream is incomplete. The OOD solver and historical output adapter are separately specified in Appendix C.5.

B.3 Finite-difference reference and analytic carrier evaluation

For the principal training and ID data, the reference advances $v_\tau = \mathcal{A}_{a,\tau} v$ directly in spot on $[0, 605]$, with $N_S = 880$ spatial intervals, $\Delta S = 0.6875$, $N_t = 320$ time intervals, and $\Delta\tau = 0.003125$, in double precision. Coefficients are evaluated at time-step midpoints. Four backward-Euler half steps replace the first two full time intervals (Rannacher startup), followed by Crank–Nicolson steps with tridiagonal Thomas solves (Rannacher, 1984; Pooley et al., 2003; Giles & Carter, 2005; Achdou & Pironneau, 2005). The conditions are

$$v(0,\tau) = 0, \quad v(605,\tau) = \max(605 - Ke^{-r\tau}, 0), \quad v(S,0) = (S-K)^+. \tag{B.2}$$

Reference Delta and Gamma use centered second-order spot differences. The strike is not shifted onto a grid node or midpoint. The output is restricted to 41 calendar times and 81 spots on $[0,1] \times [0,220]$ with coordinate tolerance 2×10^{-13}; arrays are ordered (surface,time,spot). The smallest positive remaining maturity is 0.025, then 0.05. The query edge 220 is not the computational boundary. The complete OOD evaluation uses a separate archived solver configuration described in Appendix C.5.

The boundary in (B.2) is an asymptotic finite-domain approximation. Refining a fixed domain does not test truncation, and Rannacher startup alone does not guarantee accurate shortest-maturity Gamma for every strike (Pooley et al., 2003). The distinct checks in Appendix C.1 address these issues empirically.

The carrier integral A_0 is evaluated from the coefficient generator, not interpolated sensors, with 64-point Gauss–Legendre quadrature. When strike-line clipping is inactive,

$$A_0(\tau) = \sigma_0^2 \left[(1-\gamma/2)^2 \tau + (1-\gamma/2)\gamma \frac{\tau^2}{T} + \frac{\gamma^2 \tau^3}{3T^2} \right]. \tag{B.3}$$

This is regular at $\gamma = 0$. The quadrature integrates the polynomial exactly in exact arithmetic, subject to floating-point rounding in numerical evaluation. Equation (A.25) treats a general effective line-integration error.

B.4 Architecture, training, and checkpoint selection

DeepONet architecture and normalized coordinates. For the normalized FI parameterization, let ξ collect the sampled payoff, sampled volatility, and contract scalars. The fused branch features

$b_{\theta,j}(\xi)$ and query-trunk features $q_{\theta,j}$ produce the scalar logit

$$f_\theta(\xi; m, \bar{t}) = \sum_{j=1}^{p} b_{\theta,j}(\xi) q_{\theta,j}(m, \bar{t}) + b_0, \quad m = S/K, \quad \bar{t} = t/T, \quad c = V/K. \tag{B.4}$$

Here p is the latent width and b_0 is a learned scalar bias. The relation between f_θ and the correction is specified in the FI training protocol below; the canonical E0 baseline uses physical query coordinates as described separately. The principal models have hidden/latent width 128, tanh activations, and 258,305 trainable parameters. The network combines separate payoff, volatility, and scalar branches with a fusion map and a query trunk, as in (B.4). Each of the payoff, volatility, scalar, fusion, and trunk MLPs has three biased linear layers, with tanh after the first two. Their input dimensions are respectively $101, 231, 5, 384, 2$; hidden and output dimensions are 128. The branch features are concatenated before fusion; the fusion–trunk dot product has a separate learned scalar bias.

DeepONet: canonical unnormalized baseline (E0). E0 uses physical (S, t) queries and price V, with frozen statistical standardization of branch inputs, query coordinates, and output. The E0 output normalizer is fit on the final Rannacher training labels only, using population standard deviation plus 10^{-8}. The resulting mean and scale are 40.613327773255371 and 45.26865379343748, respectively. Validation, test, and OOD labels are excluded from the fit. Training uses 5,000 Adam updates at 10^{-3}, reloads the best raw-phase validation checkpoint, and uses a new Adam optimizer for 2,000 updates at 10^{-4}. Adam has betas $(0.9, 0.999)$, epsilon 10^{-8}, zero weight decay, and no AMSGrad. For each principal seed, `default_rng(seed)` first draws all $7{,}000 \times 32$ surface indices and then all $7{,}000 \times 512$ query indices with replacement; each update shares its query vector across surfaces. Training is float32 on one CPU thread with deterministic Torch algorithms and seeded Python, NumPy, and Torch initialization. Validation pools standardized-price MSE over all 32 validation surfaces at each phase start, its first update, and every 100 updates: 74 evaluations. Strict improvement selects the checkpoint, ties retain the earlier candidate, and E0 retains the raw-phase best score during continuation. There is no early stopping, gradient clipping, scheduler, warmup, auxiliary loss, RA, FAL, or PP in this baseline.

Canonical PI-DeepONet. The PDE residual of a destandardized dimensionless price prediction $\widehat{c}$ is

$$R_{\mathrm{el}}[\widehat{c}] = \widehat{c}_{\bar{t}} + \frac{T}{2} a(Km, T\bar{t}) m^2 \widehat{c}_{mm} + Trm\widehat{c}_m - Tr\widehat{c}, \quad \mathcal{P}_a(K\widehat{c}) = -(K/T) R_{\mathrm{el}}[\widehat{c}]. \tag{B.5}$$

The PDE loss $\mathcal{L}_{\mathrm{PDE}}$ is the mean squared value of $R_{\mathrm{el}}[\widehat{c}]$ at the sampled collocation points. The objective is $\mathcal{L}_{\mathrm{data}} + \mathcal{L}_{\mathrm{PDE}} + \mathcal{L}_{\mathrm{terminal}} + \mathcal{L}_{\mathrm{boundary}}$ with four unit weights. Each of the five principal seeds runs 5,000 Adam updates at 10^{-3}, reloads the lowest-validation-error raw-phase checkpoint, creates a *new* Adam optimizer, and continues for 2,000 updates at 10^{-4}. Adam has betas $(0.9, 0.999)$, epsilon 10^{-8}, and zero weight decay.

Each update samples 32 training surfaces and a shared vector of 512 supervised grid-query indices, with replacement. Independently for each surface, it samples 512 PDE points $m = 0.01 + U(220/K - 0.01)$, $\bar{t} = 0.975U$; 512 terminal points $m = (220/K)U$, $\bar{t} = 1$; and 512 times $U \in [0, 1)$ shared by the boundaries $m = 0$ and $m = 220/K$. Uniform draws are independent apart from the stated sharing. There is no near-strike enrichment. Supervised draws use `default_rng(seed)`; float32 collocation coordinates use `default_rng(seed+5051005)`. Each loss has its own mean-square reduction; the boundary loss averages concatenated lower and upper samples. Four sequential backward calls accumulate gradients before one optimizer update.

Derivatives are with respect to raw $m = S/K$ and elapsed $\bar{t}$, with input standardization and output destandardization inside the differentiation graph. Volatility is bilinearly interpolated from reconstructed sensors, clamping coordinates to the sensor rectangle, and *then squared*. Validation is standardized price MSE on complete 32-surface development-validation grids at each phase start and every 100 updates (72 evaluations in total). Continuation resets the best-validation state; strict improvement selects, ties retain the earlier checkpoint. ID/OOD results do not enter selection. This comparator has no RA, FAL, or PP.

Historical FI price-loss algebra and later inference rule. Let f_θ be the standardized scalar network logit, with frozen dimensionless-price normalization

$$\mu = 0.5017557622056907, \qquad s_y = 0.6236731671283599. \tag{B.6}$$

The historical E5 training forward is

$$
\begin{aligned}
r_\theta &= \mu + s_y f_\theta, \quad z_v = C^{\mathrm{ref}}/K + \chi r_\theta, \quad y_n = (V_{\mathrm{FD}}/K - \mu)/s_y, \\
\mathcal{L}_{\mathrm{raw}} &= \mathrm{mean}\left[\left((z_v - \mu)/s_y - y_n\right)^2\right], \\
\mathcal{L}_{\mathrm{cont}} &= \mathrm{mean}\left[\left((\mathcal{C}_{\delta_v}(z_v; L/K, U/K) - \mu)/s_y - y_n\right)^2\right],
\end{aligned}
\tag{B.7}
$$

with $\delta_v = 0.002$. For $\chi > 0$, define the numerical-label correction $R^*_{\mathrm{FD}} = (V_{\mathrm{FD}} - C^{\mathrm{ref}})/\chi$. The pointwise raw-stage loss in (B.7) then equals $\chi^2(R_\theta - R^*_{\mathrm{FD}})^2/(K^2 s_y^2)$. Thus reconstructed-price supervision implicitly weights squared numerical-correction error by χ^2, without an additional user-specified maturity weight. The numerical target differs from the exact correction by $R^*_{\mathrm{FD}} - R^* = (V_{\mathrm{FD}} - V)/\chi$; comparison with the continuous target therefore also involves reference error. This identity applies before FAL and does not carry unchanged through the nonlinear continuation loss. E5 uses 5,000 Adam updates at 10^{-3} and 2,000 at 10^{-4}, with 32-surface and 512-query batches and validation-price selection. There is no direct correction-label MSE, extra user-specified χ point weight, near-strike loss weight, Greek loss, PDE loss, or financial penalty in final E5. Terminal nodes remain eligible; $\chi = 0$ removes their learned correction. Ordinary FAL was applied during continuation *including terminal nodes*. Consequently, at terminal nodes the continuation-stage FAL output is parameter-independent, so these samples contribute no gradient to the learned network despite remaining in the batch. The terminal-preserving output in (A.7) is a later evaluation change with frozen selected weights, not a retrained architecture. FAL-only E3 interprets the destandardized network logit directly as normalized price and also applies FAL in continuation.

The principal seeds are $\{2026, 3407, 5201, 7713, 9109\}$. The separate kink-treatment seeds are $\{20260719, 20260720, 20260721\}$. Training specifics of development studies or historical baselines are not assumed to be identical merely because an update count is shared. Alternative kink treatments include gradient-balanced soft financial penalties, diffusion-distance coordinate enlargement based on $\log(S/K)/\sqrt{A_0}$ with near-terminal regularization, and near-strike oversampling. The coordinate intervention is motivated by adaptive-scale inputs (Mou et al., 2025); it is not a full reproduction of every such method.

Matched carrier–rescaling mechanism ablation. To separate the analytic carrier from maturity rescaling, we use the generalized reconstruction

$$
z_{\theta,\alpha} = C^{\mathrm{ref}} + g_\alpha(\chi) R_\theta, \qquad g_\alpha(\chi) = \begin{cases} 1, & \alpha = 0, \\ \chi^\alpha, & \alpha > 0, \end{cases}
\tag{B.8}
$$

with

$$
\alpha \in \{0, \tfrac{1}{2}, 1, \tfrac{3}{2}\}.
$$

The learned arms are denoted A0, A05, A1, and A15, respectively. B0 is the normalized direct-price FAL control. All arms use the same frozen H1 data, width-128 architecture, input/output statistics, five principal seeds, surface/query sampling schedules, $5{,}000{+}2{,}000$ Adam-update budget, validation-price checkpoint rule, and FAL continuation protocol. No PDE loss, Greek loss, near-strike weighting, oversampling, scheduler, or PP is introduced.

A0, A05, and A15 were trained as fifteen new paired runs. Historical A1/E5 and B0/E3 checkpoints were reused only after exact generalized-path replay, initialization, data, checkpoint, and schedule audits passed. No held-out metric was used to alter the exponent set or training protocol. The primary mechanism analysis uses pre-FAL nonterminal outputs; a second evaluation applies the same terminal-preserving post-FAL convention to all arms. This distinction prevents FAL or terminal assignment from being attributed to the residual parameterization itself.

Selected checkpoints and replay scope. In the principal seed order above, the selected global updates are $(6900, 7000, 6900, 6900, 6900)$ for refitted E0, $(7000, 6900, 7000, 7000, 6600)$ for canonical PI, and $(6800, 7000, 6800, 6600, 6600)$ for FI. The E0 results in Tables 1 and 8 use the output normalizer fitted to the final Rannacher training labels. Their configurations, training logs, checkpoint hashes, and selected-state replay are recorded separately from the historical synthetic-core archive. That archive retains the earlier E0 states and output normalizer, so its replay does not reproduce the current E0 column. Appendix F distinguishes these versions and the separate OOD replay extension.

B.5 DETERMINISTIC GRID POST-PROCESSING

For each float64 array with shape $N \times 41 \times 81$, ordered (sample,time,spot), PP performs exactly two cycles. Each cycle applies: call-price lower/upper clipping and $S = 0$ overwrite; terminal payoff overwrite; reverse-calendar-time cumulative maximum; unit-weight pooled-adjacent-violators (PAVA) repair of forward spot slopes; slope clipping to $[0, 1]$; left-anchored reconstruction; and bounds/zero-spot enforcement. After two cycles, apply a final terminal overwrite, time-monotonicity pass, and bounds/boundary enforcement. There is no convergence test or repair tolerance; 10^{-8} is the post hoc violation threshold.

Only the slope-PAVA step is an isotonic projection. The whole noncommuting sequence is order-dependent, not a simultaneous metric projection onto the intersection of all constraints. It acts after inference, outside backpropagation, and changes no parameters, optimizer state, checkpoint selection, or ranking. The recorded five-model re-evaluation reproduces the reported PP outputs. It does not give continuous-surface no-arbitrage certification or independently differentiated AD Greeks. In a time-inhomogeneous pricing problem, a backward surface indexed by valuation time must also be distinguished from a family of contract maturities at a fixed valuation date; the time-repair operation is not used as a general calendar-arbitrage theorem.

Inference forms normalized queries, evaluates A_0 and C^{ref}, predicts R_θ, reconstructs z_θ, applies unit-consistent FAL, and assigns terminal/zero-spot values. PP is an optional final grid transformation. Eligible linear-map replacements leave this pricing pipeline unchanged.

B.6 METRIC DEFINITIONS, MASKS, AND COEFFICIENT CONVENTIONS

For a specified reference vector $\mathbf{V}$ and prediction $\widehat{\mathbf{V}}$,

$$\mathrm{RelL2} = \frac{\left\|\widehat{\mathbf{V}} - \mathbf{V}\right\|_2}{\|\mathbf{V}\|_2}, \quad \mathrm{RMSE} = \sqrt{N^{-1}\sum_i |\widehat{V}_i - V_i|^2}, \quad \mathrm{MAE} = N^{-1}\sum_i |\widehat{V}_i - V_i|. \quad \text{(B.9)}$$

Price p95 is a pooled absolute-error quantile, not relative price error. Principal price $\mathrm{RelL2}$ pools $32 \times 41 \times 81 = 106,272$ points before computing one norm ratio per seed. Near-strike price uses $|S/K - 1| \le 0.05 + 8\epsilon_{64}$, retains terminal nodes, and has no explicit spatial-edge exclusion. Seed means and sample SDs are computed afterward.

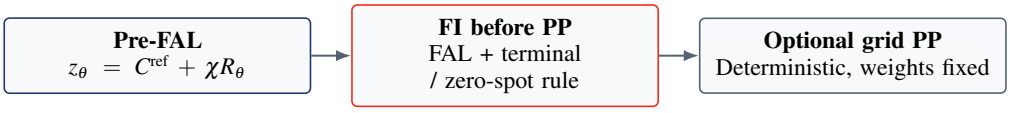
The principal Greek predictions are physical-spot `np.gradient(edge_order=2)` of saved prices, applied twice for Gamma. References come from the fixed 24-surface L3 ID/test subset. The historical near-strike mask intersected with $\tau > 0$ retains 3,160 finite prediction/reference pairs per Greek and seed, excluding 79 terminal pairs. The historical including-terminal mask has 3,239 pairs and is confined to Table 25 and its attribution discussion. The principal PDE residual RMS pools dimensionless residuals computed from the saved pre-PP output with the common interior-grid FD evaluator (98,592 points per seed). The terminal row itself is not scored, but the recorded centered time stencil at the last interior level $\bar{t} = 0.975$ uses the $\bar{t} = 1$ terminal price; consequently the near-maturity PDE residual RMS depends on the terminal-value convention. This principal FD evaluation is distinct from the PDE loss used in PI training and from the AD evaluation used in the controlled PDE-loss factorial. Price RelL2 is dimensionless. Absolute price errors, including RMSE, MAE, and absolute-error p95, have price units; Delta errors are dimensionless, and Gamma errors have inverse-price units.

The archived $\tau = 0.025$ profile analysis aggregates all 24 L3 surfaces and five seeds: absolute native errors are interpolated to the fixed common moneyness grid, then equally weighted medians/IQRs are formed over 120 surface–seed values per coordinate. Both price and Gamma there use the L3 population, unlike the principal 32-surface price table. IQRs are dispersion, not confidence intervals; no surface is selected by its model error.

Reference refinement in Table 7 excludes spatial edge columns, uses $|S/K - 1| \le 0.05$, and defines ATM-short by $0 < \tau \le 0.05$. Carrier-only regional price comparison uses the near-strike mask and $0 < \tau \le 0.05$ for pooled regional $\mathrm{RelL2}$ per seed. The shortest-time check uses only $\tau = 0.025$. The PI coefficient-sensitivity mask instead has $0 < \tau \le 0.10$ within the legal sensor rectangle. Market ATM-short uses $|\log(K/F)| \le 0.03$ and $7 \le \mathrm{DTE} \le 30$. These masks and populations are not interchangeable.

For a constraint $g_j \geq 0$, violation frequency at threshold ε_j counts $g_j < -\varepsilon_j$; mean severity averages $(-g_j)^+$. The principal financial magnitude sums seven native-array mean severities: lower and upper price-bound violations, negative price, negative forward spot slope, positive elapsed-time price increment, negative FD Gamma, and absolute terminal discrepancy. Time increments are not divided by step size. The factorial averages nine rule means, adding lower/upper FD-Delta violations. The threshold 10^{-8} defines frequencies, not magnitudes. These mixed-unit scores are protocol-specific, not a calibrated arbitrage scale.

Output-stage and population registry. Figure 8 provides a reading guide to the principal protocols. "Pre-PP" alone is not a complete output specification: FI is post-FAL, mechanism comparisons may be pre-FAL, and the complete OOD study retains its historical FI adapter. The canonical PI baseline, the normalized-RA PDE-loss factorial, and the selected-parent realization study are different experiments. Neither AD versus FD nor nonterminal versus including-terminal Greek statistics are interchangeable. Unless a caption explicitly states a paired estimator, a quoted improvement computed from a summary table is a ratio of reported means, not a mean within-seed improvement.



Study	Population / stage	Estimator / scope
Principal ID	32 price surfaces; 24 L3 for Greeks FI after FAL; before PP	Five-seed mean ± sample SD FD Greeks, $\tau > 0$
Mechanism ablation	B0 / A0 / A05 / A1 / A15 Pre-FAL and post-FAL reported separately	Five paired seeds Stage-dependent rankings
Complete OOD	320 canonical-clipped surfaces Historical FI output adapter	Pooled price norm per seed No ID/OOD degradation ratio
PDE-loss factorial	Normalized RA; paired objectives FAL is an output-map factor	AD PDE residual and Greek MAE Distinct from principal PI
Realization / market	One fixed dense parent; selected QDO Market uses constant-volatility inputs	Separate synthetic AD Greeks Market mid $\neq$ operator reference

Figure 8: Output-stage and protocol map. Arrows show the FI inference order, not a guarantee that errors decrease at every stage. DeepONet and canonical PI are direct-price comparators and do not follow the FI carrier–FAL path. The table identifies the principal distinctions needed to read the appendix: FD and AD, nonterminal and including-terminal masks, per-seed summaries and selected-parent comparisons, and market versus same-input operator references. See Appendix B.6 for exact masks and aggregation.

PI coefficient-path sensitivity. PI interpolates clipped 21×11 volatility sensors before squaring, whereas the reference evaluates and clips the generator pointwise. For fixed selected weights,

$$q_{\text{coef}} = \frac{\text{RMS}(R_{\text{sensors}} - R_{\text{generator}})}{\text{RMS}(R_{\text{sensors}})} \tag{B.10}$$

lies in $[0.0103, 0.0148]$ on the admissible validation region and $[0.00163, 0.00210]$ on this check's ATM-short subset. The direct PDE-residual perturbation is small on these fixed predictions, but it does not bound the price effect of retraining with a different coefficient path. FI's generator-based carrier access is also distinct from PI's sensor-interpolated PDE residual.

Finite-precision admissibility. An initial normalized float32 evaluation reported apparent lower-bound violations at about 2.0% of FI and 11.0% of FAL-only predictions. Double/higher-precision re-evaluation found no genuine interval violation at those points. This verifies those outputs, not every numerical protocol or cross-point shape condition.

C Additional Numerical Experiments

The additional experiments examine reference accuracy, representation choices, supervision, and generalization. Each study retains its own model population, output stage, and error estimator; the formal comparisons and development studies are identified below.

C.1 Reference accuracy and analytical checks

The canonical clipped variance obeys $0.05^2 \le a \le 1$. In log-moneyness, the derivative of the unclipped spatial factor contains $2e^x \operatorname{sech}^2(2(e^x-1))$, which is bounded on $\mathbb{R}$. Clipping is Lipschitz; squaring preserves this property on the bounded nonnegative range. Thus the coefficient part of the value assumptions holds. At the strike, when clipping is inactive,

$$\sigma(K, T-u) = \sigma_0(1-\gamma/2+\gamma u/T). \tag{C.1}$$

On the closure of training support this lies in $[0.075, 0.5625]$ for $0 \le u \le T$, strictly inside $[0.05, 1]$. A smooth neighborhood of the strike therefore exists. The carrier is independent of β, while

$$a_* = \sigma_0^2(1-\gamma/2)^2, \qquad b_* = 4\beta a_*. \tag{C.2}$$

Local profiles depend on skew even though the carrier does not. These facts concern the canonical continuous field, not every historical coefficient convention.

Table 7 reports the primary check, which refines $(N_S, N_t) = (880, 320)$ to $(1760, 640)$ with the 32 ID surfaces and selected models fixed. It excludes the 82-surface parameter-support OOD archive with partly model-error-based selection. FI–PI ordering is preserved in all five aggregate metrics and all 25 seed–metric comparisons. Joint refinement does not independently identify space/time order; main results retain the base reference.

Table 7: Reference-solver sensitivity on the fixed 32 ID surfaces: p95 absolute base-to-doubled-resolution discrepancies in the corresponding physical units. ATM-short uses $0 < \tau \le 0.05$. This is a separate 32-surface sensitivity population, not the 24-surface Greek population. Discrepancies are not certified error bounds.

Slice	Price	Delta	Gamma
All admissible points	4.776101×10^{-4}	5.962378×10^{-5}	6.385583×10^{-6}
Near strike	2.145785×10^{-3}	3.096243×10^{-4}	1.103966×10^{-4}
ATM-short	4.730651×10^{-3}	1.654274×10^{-3}	1.508516×10^{-3}

A separate domain check moves the upper boundary from 605 to 1210 while preserving $\Delta S = 0.6875$ and the time step. Near-strike price/Delta/Gamma p95 discrepancies are 5.3701×10^{-12}, 9.0501×10^{-13}, and 1.3680×10^{-13}; the largest checked absolute discrepancy is 4.5276×10^{-9}. At $\tau = 0.025$, independent spatial refinement gives price/Delta/Gamma p95 differences 0.00337994, 0.00187478, and 0.00159277; independent temporal refinement gives 0.00286804, 0.000573062, and 0.000454783. The combined Gamma discrepancy is 0.00204970, with no reversal of the nine checked five-seed p95 rankings. The matched refined-reference FI–PI gap for this slice is 0.06301548116729162, so the unrounded discrepancy 0.002049703724169065 is approximately 3.25% of that gap. This is not the principal table's different Greek statistic or a continuum error bound.

A smooth uniformly elliptic manufactured first-order-skew family uses a 3×3 product of independent spatial/temporal refinements. Observed spatial orders for price, Delta, and Gamma are 1.9976, 1.9989, and 2.0039; temporal orders are 2.0007, 2.0002, and 2.0018. All 18 tested nonzero self-similar rays per quantity support the predicted scales. Median relative discrepancies of scaled profiles from their limits are 0.00256248, 0.01000658, and 0.0152407. These manufactured convergence orders are not transferred to every shortest-maturity production-grid Gamma.

The pre-FAL weighted identities pass on 575 reliable points per verification path. Direct post-FAL differentiation agrees with the complete chain-rule reconstruction at 360 points, with maximum absolute discrepancy 1.7833×10^{-15}. Such finite identity checks do not supply the continuum norm in (A.40). The archived manufactured-field formula, all refinement grids, and finite-check point lists are needed for independent replay; they are not reconstructed from the aggregate observations.

Figure 9 compares the reported refinement discrepancies by region and the separately measured manufactured-solution orders. Different panels retain their own units and numerical populations.

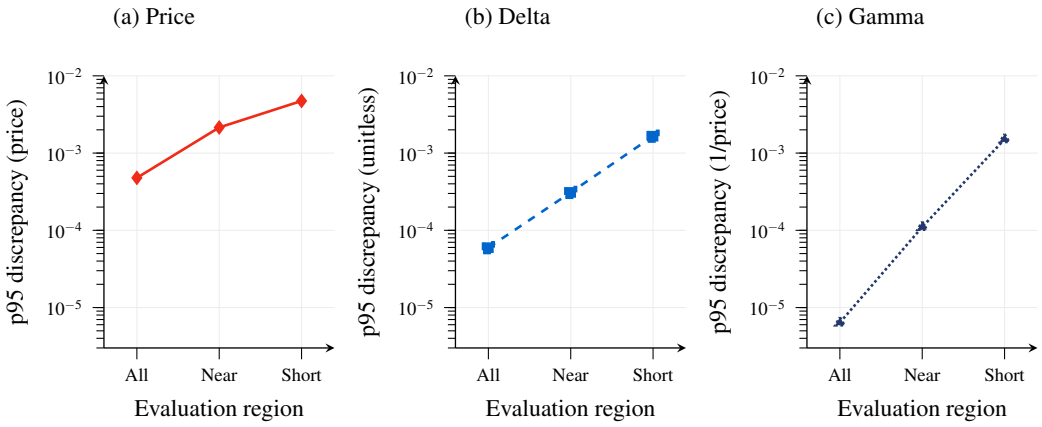


Figure 9: Reference-solver refinement on the fixed 32 ID surfaces, from $(N_S, N_t) = (880, 320)$ to $(1760, 640)$, using Table 7. "All" denotes admissible points, "Near" the near-strike region, and "Short" the ATM-short subset $0 < \tau \leq 0.05$. Each panel uses its own physical units and a logarithmic vertical axis. Points are pooled p95 base-to-refined discrepancies, not sample means; no confidence intervals are available from these summaries. Lines connect categories only. The annotation reports a separate manufactured-solution experiment in price/Delta/Gamma order. Neither check certifies continuum error.

C.2 Principal comparisons, component ablations, and representation controls

Table 1 reports means only for compact presentation. Tables 8 and 9 retain the corresponding five-seed sample standard deviations and full reported precision. The repeated FI row uses the principal-comparison values; carrier-only price results are stated below. The E0 column uses the refitted baseline; the version-specific replay scope is described in Appendices B.4 and F.

Table 8: Principal five-seed comparison, mean ± sample SD. Price: 32 held-out ID surfaces. Greeks: nonterminal near-strike FD errors on the fixed 24-surface L3 subset. PDE residual RMS uses the common interior FD evaluator (98,592 points/seed). PP is excluded. Lower is better; bold marks the lowest mean in each row.

Metric	DeepONet	PI-DeepONet	FI-DeepONet
Price RelL2	0.0059854 ± 0.00043	0.0072437 ± 0.0012	$\mathbf{0.0009645} \pm 0.000049$
Near price p95	1.7967 ± 0.169	2.3338 ± 0.33	$\mathbf{0.15794} \pm 0.0073$
Delta p95	0.15183 ± 0.0092	0.15474 ± 0.021	$\mathbf{0.024518} \pm 0.0018$
Gamma p95	0.033225 ± 0.0015	0.035549 ± 0.0021	$\mathbf{0.0043378} \pm 0.000035$
PDE Residual RMS	0.046102 ± 0.0075	0.010494 ± 0.0015	$\mathbf{0.0070251} \pm 0.00022$

The kink intervention in Table 10 compares price representation with alternative query representation and sampling. Its results concern the tested controls and do not establish superiority over adaptive-coordinate or adaptive-sampling architectures in general.

On the principal 32-surface ID price population the deterministic carrier alone has global RelL2 $=$ 0.00688952 and near-strike price p95 0.392665. The full predictor improves these by 86.00% and 59.78%. Carrier-only is a diagnostic rather than a trained baseline; the matched learned-model ablation below separately identifies the carrier and maturity-rescaling contributions. On the experiment-specific ATM-short price mask, the analytic carrier is already highly accurate and the learned correction raises RelL2 by approximately 40.46%; the improvement is therefore not uniform over every maturity slice. Together with the global-rank result below, these observations prevent interpreting

Table 9: Principal component ablation (five seeds, mean ± SD). "Normalized" denotes contract normalization; RA is the carrier-residual representation. The complete FI predictor has normalization, RA, and FAL.

Configuration	Global RelL2	Near-strike price p95
DeepONet	$0.0059853982 \pm 0.000433$	1.7967389 ± 0.1692
Normalized DeepONet	0.0076 ± 0.0013	2.2553 ± 0.3486
Normalized + soft financial penalties	0.0075 ± 0.0013	2.2125 ± 0.3474
Normalized + FAL	0.0024 ± 0.000488	0.5277 ± 0.0581
Normalized + FAL + PP	0.0023 ± 0.000498	0.4581 ± 0.0688
FI-DeepONet	0.0009645 ± 0.0000488	0.15794 ± 0.0073
FI-DeepONet-PP	0.001043 ± 0.0000455	0.1579 ± 0.0073

Table 10: Payoff-kink treatments on a common three-seed population before PP, distinct from the principal five-seed population. Entries are means. Lower is better; bold marks the lowest mean in each column.

Treatment	Global RelL2	Near price p95	Near Delta p95	Near Gamma p95
Coordinate enlargement	0.001795	0.348724	0.065799	0.025100
Near-strike oversampling	0.001811	0.364922	0.085414	0.036080
RA	**0.0009050**	**0.145524**	**0.020482**	**0.003367**

the asymptotic construction as proof that the learned correction is globally simpler or that $\alpha = 1$ is training-optimal. Historical including-terminal Greek attribution is retained below rather than mixed into nonterminal results.

C.2.1 Matched carrier–rescaling mechanism ablation

Within the carrier arms, the pre-registered five-seed ablation varies the reconstruction exponent in

$$z_{\theta,\alpha} = C^{\mathrm{ref}} + g_\alpha(\chi) R_\theta,$$

with $g_0(\chi) = 1$ and $g_\alpha(\chi) = \chi^\alpha$ for $\alpha > 0$, as in (B.8). B0 is a separate direct-price control without the carrier. Table 11 reports the pre-FAL nonterminal mechanism metrics.

Table 11: Matched carrier–rescaling ablation before FAL; five-seed mean ± sample SD. B0 is the normalized direct-price control, and A0/A05/A1/A15 use the same carrier with $\alpha = 0, \frac{1}{2}, 1, \frac{3}{2}$. Lower is better.

Arm	Global price RelL2	ATM-short price RelL2	Near Delta p95	Near Gamma p95
B0	0.04680 ± 0.00996	0.1752 ± 0.0228	0.1362 ± 0.0194	0.03690 ± 0.00274
A0	0.01975 ± 0.00117	0.04846 ± 0.01045	0.02388 ± 0.00335	0.005722 ± 0.000121
A05	0.01345 ± 0.00094	0.01974 ± 0.00350	$\mathbf{0.02052} \pm 0.00233$	0.005416 ± 0.0000547
A1	0.01292 ± 0.000520	0.01270 ± 0.000747	0.02508 ± 0.00174	0.005331 ± 0.0000529
A15	$\mathbf{0.01248} \pm 0.000744$	$\mathbf{0.01028} \pm 0.000216$	0.03025 ± 0.00130	$\mathbf{0.005232} \pm 0.00000563$

Adding the strike-matched carrier without maturity rescaling (B0→A0) lowers the mean pre-FAL global price, Delta, and Gamma errors by approximately 57.8%, 82.5%, and 84.5%, respectively; the corresponding paired directions improve in all five seeds. Thus the carrier has an empirical contribution beyond normalization and FAL even before the maturity gate is introduced.

Adding the linear gate (A0→A1) lowers global price RelL2 by approximately 34.6% and ATM-short price RelL2 by 73.8%, while near-strike Gamma p95 decreases by about 6.8%. Near-strike Delta p95 instead increases by about 5.0%. Hence maturity rescaling has a clear but metric-dependent finite-budget effect.

For $\alpha > 0$, the raw correction vanishes at maturity by construction. The pre-FAL terminal maximum error is 5.03 ± 0.99 for A0 but approximately 2.26×10^{-5} for A05, A1, and A15. This terminal behavior is reported separately from the nonterminal mechanism metrics.

Table 12 evaluates the same checkpoints after the common FAL and terminal-preserving output rule.

Table 12: Matched carrier–rescaling ablation after FAL and the common terminal-preserving inference rule; five-seed mean ± sample SD. These values use a common evaluation stage for all arms and are distinct from the historical component table above. Lower is better.

Arm	Price RelL2	Near price p95	Delta p95	Gamma p95	FD PDE RMS
B0	0.002336 ± 0.000493	0.4581 ± 0.0688	0.08459 ± 0.01166	0.02535 ± 0.00265	0.02577 ± 0.00291
A0	0.001179 ± 0.000234	0.1616 ± 0.0168	0.02289 ± 0.00394	0.004527 ± 0.000125	0.008158 ± 0.00131
A05	**0.0008339** ± 0.0000657	**0.1424** ± 0.0162	**0.01959** ± 0.00249	0.004354 ± 0.0000506	**0.006339** ± 0.000232
A1	0.0009645 ± 0.0000488	0.1579 ± 0.00730	0.02452 ± 0.00178	**0.004338** ± 0.0000353	0.007025 ± 0.000224
A15	0.001192 ± 0.0000705	0.1677 ± 0.00768	0.02982 ± 0.00132	0.004413 ± 0.0000859	0.007912 ± 0.000220

No single exponent minimizes every finite-budget metric. A15 has the lowest pre-FAL global and ATM-short price errors and the lowest pre-FAL Gamma p95, whereas A05 has the lowest pre-FAL Delta p95 and the lowest post-FAL price, near-price, Delta, and PDE errors; A1 has the lowest post-FAL Gamma p95. These observations do not contradict Proposition 3.2: $\alpha = 1$ is critical for a bounded, nondegenerate value target in the stated asymptotic regime, not a theorem of metric-wise training optimality.

Several arm rankings change between the pre- and post-FAL stages. FAL therefore interacts nontrivially with the residual parameterization, and post-FAL differences are not attributed to the exponent alone.

Figure 10 shows the post-FAL metric trade-offs with the reported sample standard deviations. These error bars describe variation across five seeds; they are not confidence intervals and do not establish significance of the small differences between exponents.

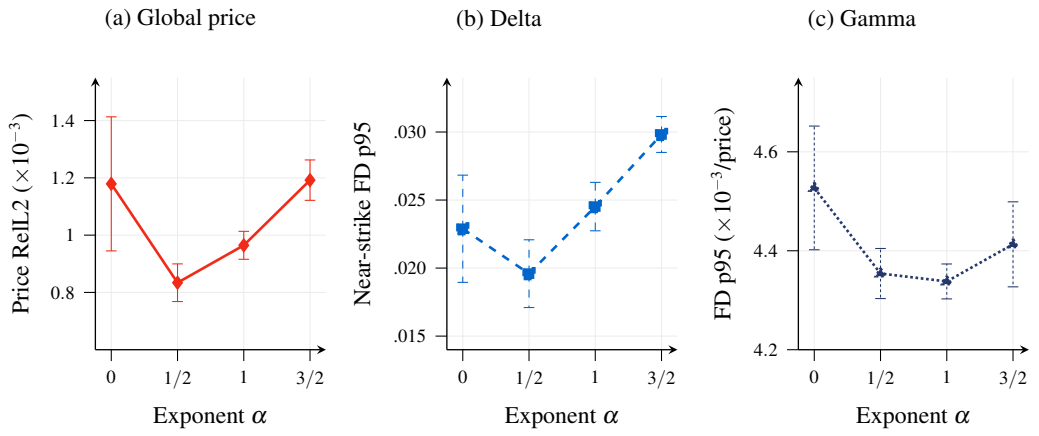


Figure 10: Post-FAL exponent ablation with the common terminal-preserving output rule, from Table 12. Points and error bars are five-seed means and sample standard deviations for A0, A05, A1, and A15. The lowest mean price and Delta errors occur at $\alpha = 1/2$, while the lowest mean Gamma error occurs at $\alpha = 1$. Vertical axes are zoomed; Greek units differ. Connecting lines are visual guides between the four tested exponents, not an interpolated optimization landscape. The critical asymptotic value scaling at $\alpha = 1$ is not a claim of finite-budget optimality for every metric.

The global linear-reconstruction experiment fits the mean and basis using only 512 training surfaces, with a common 3,240-point nonterminal output grid and uniform evaluator weights. It compares $v_{\text{full}} = V/K$, $w_{\text{corr}} = (V - C^{\text{ref}})/K$, and $r_{\text{corr}} = w_{\text{corr}}/\chi$. It does not replace this common-coordinate problem with irregular contract-dependent ATM subsets.

For a square-integrable centered Hilbert-space random output R, with covariance eigenvalues $\lambda_1 \geq \lambda_2 \geq \cdots$, the optimal rank-$p$ mean-square reconstruction error is

$$\inf_{\dim E = p} \mathbb{E}\,\|R - P_E R\|^2 = \sum_{j>p} \lambda_j. \tag{C.3}$$

Indeed, the error is $\operatorname{tr}(M_R) - \operatorname{tr}(P_E M_R)$, and the latter trace is maximized by leading eigenvectors. This standard identity (Lanthaler et al., 2022) does not order the spectra of full price and correc-

Table 13: Global linear reconstruction: energy ranks and held-out error ratios relative to normalized full price at the same rank. Ratios below one would mean easier reconstruction.

Representation	99% rank	99.9% rank	$p = 8$	$p = 32$	$p = 128$
Full normalized price v_{full}	1	3	1	1	1
Pricing correction w_{corr}	8	16	31.666	37.484	60.457
Rescaled correction r_{corr}	13	29	62.013	133.118	291.398

tion. Table 13 shows that both correction representations require more modes to reach the reported variance thresholds and have higher held-out reconstruction error at the tested ranks. This does not support reduced global linear complexity as the explanation for the observed gains. It does not contradict the exact correction equation or local short-time cancellation; the oracle linear reconstruction error also does not determine the error of a trained branch–trunk network.

C.3 PDE LOSS AND LIMITED DATA

The five-seed factorial uses a common normalized carrier-residual parameterization. The price-only arm and the arm with an added PDE loss do not use FAL, whereas the corresponding FAL arms activate the output map during continuation. The financial factor therefore changes the output map without adding a soft financial-penalty loss. The data-only control is not the principal unnormalized DeepONet, and none of these arms redefines the principal FI model. Separate development studies vary PDE-loss weights, training-gradient calibration, late learning-rate instability, and registered-mixture, uniform, and PDE-residual-adaptive collocation.

For the formal factorial, all five seeds use paired initialization and supervised schedules, with validation normalized-price MSE checkpoint selection. The PDE-loss arms share 1,024 collocation pairs per update, uniformly sampled over $m \in [0.01, 2.75)$ and $\bar{t} \in [0, 0.975)$. Their final PDE-loss coefficient is 0.865302726925: the first 10% of updates have zero weight, the next 20% ramp linearly, and the remaining 70% use the fixed coefficient. All arms reduce the learning rate tenfold at update 3,800, retaining that multiplier in continuation. These settings differ from the canonical comparator.

The separate stabilized collocation study compares a registered random mixture, uniform draws, and a PDE-residual-adaptive bank under the carrier-residual parameterization using development data. The registered mixture allocates probabilities $0.50/0.25/0.25$ to global, near-strike ($m \in [0.95, 1.05)$), and late-time ($\bar{t} \in [0.80, 0.975)$) draws. All three schemes retain the cutoff $\bar{t} < 0.975$; none resolves the omitted shortest-maturity PDE region of the canonical comparator. The full scheme comparison uses one seed; the selected uniform scheme has a separate second-seed development check. These results are not a matched collocation-only rerun of canonical PI.

Table 14: Paired effects in the normalized carrier-residual factorial. FAL is an output-map change; it adds no soft financial loss. Entries are mean *within-seed* percentage changes, with improved seeds out of five in parentheses. Negative is better. PDE residual RMS and Greeks are evaluated by AD, whereas Table 8 uses the principal FD evaluation for the PDE metric and FD p95 for Greeks.

Comparison	Price RelL2	PDE residual RMS	Delta MAE	Gamma MAE
FAL vs price control	-33.95 (5)	-8.38 (5)	-37.69 (5)	-4.80 (5)
PDE loss vs price control	-2.34 (3)	-34.18 (5)	-7.87 (5)	-5.85 (5)
PDE loss + FAL vs FAL	$+34.38$ (0)	-30.67 (5)	-3.83 (4)	-3.33 (4)

Table 15 reports the corresponding arm means, which are distinct from the paired percentage-change estimator above.

The ratio of means and the mean of paired within-seed changes are different estimators. For FP versus F, the former gives a 34.07% price increase and a 30.68% PDE-residual RMS decrease; the latter gives 34.38% and 30.67% in Table 14. The qualitative trade-off agrees.

In the separate development study with 25% data, the control has price $\text{RelL2} = 0.0013530935$, versus 0.0011967161 with the PDE loss; PDE residual RMS is 0.0059653614 versus 0.0037237662. The reported improvements are 11.557% for price, 37.577% for PDE residual RMS, 11.039% for Delta MAE, and 12.052% for Gamma MAE. Improved-seed counts for price, PDE residual, Delta,

Table 15: Five-seed means for the normalized carrier-residual factorial: C uses price supervision, F adds FAL, P adds the PDE loss, and FP combines them. No arm adds a soft financial loss. Lower is better; bold marks the lowest mean in each column. These estimators differ from the principal comparison. PDE residual RMS is evaluated by AD in this factorial.

Objective	Price RelL2	PDE residual RMS	Delta MAE	Gamma MAE	Financial
C	0.001576342	0.006682306	0.005316740	0.000777047	0.002907677
F	**0.001036124**	0.006115663	0.003305758	0.000739836	**0.000087879**
P	0.001525491	0.004392578	0.004876109	0.000731263	0.003298899
FP	0.001389139	**0.004239599**	**0.003172085**	**0.000714711**	0.000093171

and Gamma are 4/5, 5/5, 4/5, and 5/5. These observations were used during development and are not independent confirmation. Another adaptation with a reinitialized optimizer lowered PDE residual RMS while increasing price and financial errors and did not change the selected final model.

Table 16 reports global price errors across six data fractions; Table 17 records the corresponding near-strike price and Gamma tails.

Table 16: Six-level limited-data study: mean global RelL2 over five seeds, before PP. Additions refer to Normalized DeepONet; this study does not redefine the final RA model. Bold marks the lowest mean within each data fraction.

Data retained	Normalized	+ balanced penalties	+ FAL	+ FAL + penalties
100%	0.007066	0.008253	**0.002648**	0.002684
50%	0.007299	0.008733	**0.002752**	0.002883
25%	0.007526	0.008868	**0.003306**	0.003421
12.5%	0.009351	0.010642	0.004504	**0.004489**
5%	0.016960	0.017343	0.008249	**0.007894**
1%	0.061610	0.061183	0.030917	**0.030302**

Twenty deterministic training batches calibrate the aggregate penalty/data gradient ratio to $[0.8, 1.2]$; weights are then fixed. This differs from arbitrary hand weighting or a continuously adaptive scheme. FAL benefits all six fractions. Standalone balanced penalties are unfavorable at most fractions; adding them to FAL gives modest global gains at the smallest fractions and improves near-strike tails.

Table 17: Near-strike tails under limited data. Each pair compares FAL with FAL plus balanced penalties on Normalized DeepONet. Bold marks the lower value within each metric pair and data fraction.

	Near price p95		Near Gamma p95	
Data retained	FAL	FAL + penalties	FAL	FAL + penalties
100%	0.601981	**0.566507**	0.027989	**0.026772**
50%	0.616515	**0.561270**	0.028161	**0.025659**
25%	0.836214	**0.796390**	0.028038	**0.026413**
12.5%	1.182752	**1.104384**	0.031705	**0.028233**
5%	2.246798	**2.138333**	0.035645	**0.031917**
1%	6.899662	**6.523885**	0.071334	**0.069443**

At both 5% and 1% data, FAL-plus-penalty global error improves in four of five seeds, not all. This supports a benefit from the added penalties under limited data, not a recommendation that all FI models should contain soft penalties.

Table 18 retains the separate financial aggregates and their paired changes.

The threshold for violation frequency is not used as a mean-severity threshold. Small aggregate values after PP are not continuous-surface certificates. The QDO selection score in Appendix D.5 is a third, distinct composite and is not substituted for these financial magnitudes.

Figure 11 keeps the paired PDE-loss effects and limited-data study in separate panels. Its displayed points summarize existing experiments; the connecting lines in the limited-data panel only guide the eye between tested fractions.

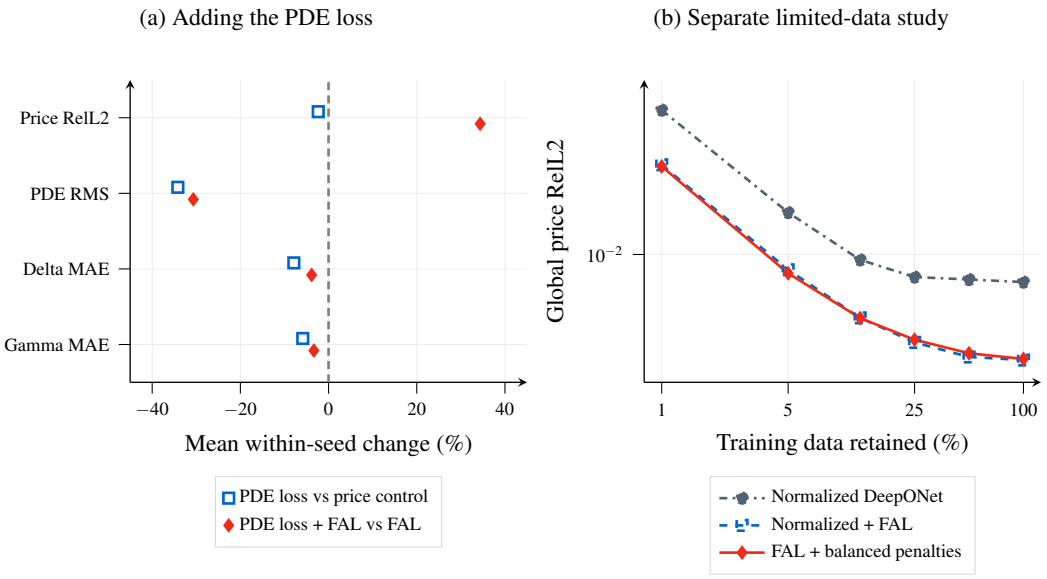


Figure 11: Distinct PDE-loss and limited-data studies. (a) Mean paired within-seed percentage changes from Table 14; negative is better. The normalized-RA factorial uses AD PDE residuals and Greek MAEs. Adding the PDE loss to FAL lowers the residual but raises price error. (b) Five-seed mean price errors from Table 16; all three curves start from Normalized DeepONet, not the final RA model. Both axes in (b) are logarithmic. The source tables do not provide uncertainty for these plotted summaries, so no error bars are inferred. Connecting lines in (b) only join tested fractions. The final FI objective remains price-only.

Table 18: Protocol-specific financial aggregates retained for traceability. The principal seven-rule sum and factorial nine-rule mean have different definitions and mixed units.

Configuration / comparison	Value
Principal PI	0.7917 ± 0.1426
Principal FI	$(7.236 \pm 0.237) \times 10^{-4}$
Principal FI-DeepONet-PP	$(8.088 \pm 2.314) \times 10^{-10}$
Factorial data control	0.002907677
Factorial FAL	0.000087879
Factorial PDE loss	0.003298899
Factorial combined	0.000093171
FAL vs control	-96.94%; 5/5 improve
PDE loss vs control	$+14.52\%$; 1/5 improve
Combined vs FAL	$+6.03\%$; 0/5 improve

C.4 Matched PI collocation sensitivity

We predeclared and hashed a single collocation-only sensitivity before training. It uses the five canonical PI seeds, 512 PDE points per surface per update, and a fixed equal mixture of global samples with full nonterminal time support and enriched near-strike/short-maturity samples satisfying $|S/K-1| \le 0.05$ and $0 < \tau/T \le 0.10$, with float32 endpoint guards fixed in advance. Architecture, normalization, supervised/terminal/boundary draws, the four unit-weight losses, Adam settings, the $5{,}000 + 2{,}000$ update budget, validation population, and phase-wise validation checkpoint selection are unchanged. No held-out ID/OOD outcome is used to tune the sampler or select checkpoints. All five runs completed the locked 7,000-update protocol before held-out evaluation.

Table 19 reports the same five metrics and populations as Table 8. Relative to canonical PI, the sensitivity lowers mean price $\mathrm{RelL2}$ by 1.59%, near-strike price p95 by 5.67%, Delta p95 by 7.61%, and Gamma p95 by 0.40%, while increasing PDE residual RMS by 18.26%. The corresponding per-seed improvement counts are $4/5$, $3/5$, $4/5$, $1/5$, and $1/5$, respectively, so the small mean Gamma change in particular is not a seed-stable improvement. The sensitivity-to-FI mean ratios are 7.39, 13.94, 5.83, 8.16, and 1.77 for those five metrics. No importance weights map the enriched collocation samples back to the canonical uniform collocation measure; therefore the intervention changes the empirical measure underlying the PDE loss and its spatial/temporal coverage, even though the four explicit loss coefficients remain at unit weight. The reported PDE residual RMS, however, uses the same principal FD evaluator as Table 8. The FI–PI gap persists under this fixed predeclared near-strike/short-maturity collocation intervention. A separate upper-boundary target sensitivity produces the same qualitative conclusion. Because this is one sampler and reuses the held-out populations, it is robustness evidence rather than independent confirmation, a universal PI-fairness result, or a characterization of all possible physics-informed collocation objectives.

Table 19: Predeclared five-seed PI collocation sensitivity. Entries are mean $\pm$ sample SD. The historical canonical PI and FI columns are the same models/populations as Table 8. Price p95 includes terminal points; Delta/Gamma p95 exclude the terminal line. Lower is better; bold marks the lowest mean in each row.

Metric	Canonical PI	PI collocation sensitivity	FI-DeepONet
Price $\mathrm{RelL2}$	0.0072437 ± 0.00123	0.0071289 ± 0.00138	$\mathbf{0.00096450} \pm 0.0000488$
Near price p95	2.3338 ± 0.327	2.2014 ± 0.298	$\mathbf{0.15794} \pm 0.00730$
Delta p95 ($\tau > 0$)	0.15474 ± 0.0205	0.14296 ± 0.0162	$\mathbf{0.024518} \pm 0.00178$
Gamma p95 ($\tau > 0$)	0.035549 ± 0.00205	0.035406 ± 0.00275	$\mathbf{0.0043378} \pm 0.0000353$
PDE residual RMS	0.010494 ± 0.00151	0.012410 ± 0.00223	$\mathbf{0.0070251} \pm 0.000224$

Table 20 exposes the five selected-state results behind the collocation-sensitivity summary.

Table 20: Per-seed PI collocation-sensitivity metrics under the locked intervention. These are the selected checkpoints after the unchanged phase-wise validation rule; lower is better.

Seed	Price $\mathrm{RelL2}$	Near price p95	Delta p95	Gamma p95	PDE residual RMS
2026	0.00584792	1.90532	0.126435	0.0330828	0.0108313
3407	0.00700464	2.25867	0.135171	0.0357557	0.0109860
5201	0.00808553	2.42720	0.164153	0.0375136	0.0149560
7713	0.00893305	2.53432	0.156070	0.0385162	0.0147449
9109	0.00577322	1.88157	0.132980	0.0321622	0.0105338

C.5 Complete parameter-support OOD population and secondary selected subset

Table 21: Proposed and realized normalized parameter-support OOD distances beyond the nearest training boundary, divided by the corresponding training width. The joint row is multi-axis.

Regime	Proposed near/mid/far	Realized min/median/max	Count
Low K	.037/.113/.188	.042/.109/.179	32
High K	.037/.175/.312	.043/.174/.308	32
High r	.077/.231/.385	.079/.247/.383	32
High σ_0	.086/.257/.429	.098/.247/.428	32
Negative β	.033/.100/.167	.035/.051/.065	32
Positive β	.033/.083/.133	.036/.079/.129	32
Negative γ	.053/.132/.211	.054/.126/.193	32
Positive γ	.053/.132/.211	.061/.112/.210	32
Joint	—	Multi-axis	64

The primary parameter-support OOD evaluation compares refitted E0 and FI on the complete canonical-clipped 320-surface population. Across five seeds, mean pooled $\mathrm{RelL2}$ $\pm$ sample SD

is 0.1974974 ± 0.0269031 for E0 and $0.00151740 \pm 0.0000751062$ for FI. Table 1C reports these means at reduced display precision; the sample standard deviations are retained here.

The archived OOD reference uses $S_{\max} = 825$, $N_S = 1200$, $N_t = 320$, and four Rannacher half steps, rather than the principal ID solver domain. Each seed pools all $320 \times 41 \times 81$ price points, including terminal and boundary points. FI uses the same selected weights as in the principal study but retains the historical OOD output adapter; the corrected terminal-preserving ID adapter is a different evaluation convention. These results therefore support a matched E0–FI comparison on the OOD population, not a matched OOD/ID degradation ratio. This complete-population comparison does not isolate the OOD contributions of normalization, FAL, or RA; intermediate-architecture OOD values below come from the secondary selected archive and are not substituted for a complete-320 ablation. The full initial OOD implementation used the same *unclipped* coefficient in sensors and reference labels, making inputs/labels internally consistent but different from canonical training/ID clipping. Table 22 restores canonical clipping on identical parameter rows with fixed models. Clipping affects all 32 negative-skew, 15 joint, and one positive-skew surface. FI's negative-skew result changes about 11.369%, more than the 1.4553% aggregate change. Whole-population and affected-surface statistics are reported separately.

Table 22: Historical complete 320-row clipping sensitivity, fixed pre-refit trained models and identical parameter rows.

Model	Unclipped	Canonical clipped	Relative change
DeepONet (historical E0)	0.197656594	0.197654514	-0.0011%
FI-DeepONet	0.0014956348	0.0015174005	$+1.4553\%$

The following 82-surface L3 archive is secondary. Selection depended partly on model errors; its DeepONet aggregate 0.2636 must not be used as the complete-320 error. It does not establish a representative parameter-support OOD comparison for all intermediate architectures. Table 23 preserves the regime-level values from this selected archive for traceability.

Table 23: Secondary OOD archive: price $\mathrm{RelL2}$ on 82 selected higher-resolution surfaces. Columns with FAL are normalized; this is not the complete-population primary evidence.

Regime	DeepONet	FAL	FAL + PP	FI	FI + PP
Aggregate	0.2636	0.0042	0.0041	0.0015	0.0014
High strike	0.2006	0.0081	0.0077	0.0032	0.0038
High rate	0.0074	0.0035	0.0035	0.000748	0.000883
High volatility	0.0138	0.0091	0.0089	0.0017	0.0017
Joint shift	0.1934	0.0108	0.0106	0.0038	0.0039
Low strike	0.3822	0.000590	0.000564	0.000833	0.000381
Negative skew	0.0062	0.0025	0.0024	0.0010	0.0011
Negative term	0.0050	0.0022	0.0020	0.000789	0.000939
Positive skew	0.0060	0.0028	0.0027	0.0014	0.0015
Positive term	0.0054	0.0017	0.0017	0.000645	0.000773

The proposed shifts are moderate; realized negative-skew excursions are especially smaller than their proposal extreme. Substituting $S = Km$ into the family removes explicit K dependence from its volatility formula, so strike shifts partly test scale handling. Skew/term shifts change spatial/time shape. These shifts provide the primary OOD evidence within the stated coefficient family; off-family directional responses are examined separately in Appendix C.6. Principal ID, complete OOD, and selected-82 values have distinct model/population/aggregation identities; no cross-study OOD/ID ratio is used.

C.6 Beyond-family functional-response stress tests and development studies

At fixed (K, r, T), the generating family is a low-dimensional manifold $\mathcal{M} = \{\sigma_{\sigma_0,\beta,\gamma}\}$. A direction normal to its sampled tangent span is not produced by an infinitesimal change of its generating parameters. A carrier-null perturbation also satisfies $h(K, t) = 0$, so its first-order analytic-reference response is zero.

To distinguish volatility from variance inputs, set $\mathcal{G}_\sigma[\sigma] = \mathcal{G}[\sigma^2]$. For a smooth admissible perturbation $\sigma + \varepsilon h$, $Z = D\mathcal{G}_\sigma[\sigma]h$ satisfies

$$\partial_\tau Z = \mathcal{A}_{\sigma^2,\tau} Z + \sigma h S^2 v_{SS}, \qquad Z(S,0) = 0. \tag{C.4}$$

Both σ, h are evaluated at $(S, T-\tau)$. The factor σh is the derivative of $\frac{1}{2}\sigma^2$; this is not a variance perturbation. Z_S, Z_{SS} are mixed input–spot sensitivities, distinct from ordinary Delta/Gamma.

For a numerical step $M_n v_{n+1} = N_n v_n + b_n$, differentiation gives

$$M_n Z_{n+1} = N_n Z_n + \dot{N}_n v_n + \dot{b}_n - \dot{M}_n v_{n+1}. \tag{C.5}$$

Initial/boundary values, restriction, and Greek stencils are differentiated consistently. Model AD versus its own FD checks implementation; comparison with this reference tangent checks the *learned response*. Derivative-informed operator learning concerns the latter problem (O'Leary-Roseberry et al., 2024; Qiu et al., 2024; Luo et al., 2026; Noguer i Alonso, 2026; de Feo, 2026).

The completed final-FI study probes a stricter off-family regime than the parameter-support OOD test. It uses seed 7713, selected update 6600, and 137 identifiable perturbation pairs from 23 bases. Carrier-null median directional $\mathrm{RelL2}$ is 1.10659 for price, 1.01399 for Delta response, and 1.02228 for Gamma response, with slightly negative near-zero cosines. The model responds measurably to the perturbations but does not accurately reproduce the off-family reference tangent. This is a single selected-model functional stress test rather than a five-seed comparison.

An earlier normalized FAL-only input ablation compares five information variants over three seeds. Removing all volatility information increases price error by factors 2.27 before FAL and 8.32 after FAL. This establishes within-family value of volatility information; off-family directional response is evaluated in the dedicated stress tests. That earlier model's one-shot off-family response has $\mathrm{RelL2} = 9.47$ and cosine 0.066. Five subsequent input/fusion architectures on the original family retain small price errors but do not repair the tested function-direction response. Removing generating parameters or delaying fusion is not sufficient in those tests; it is not a reproduction of every late-fusion design (van Tegelen et al., 2026).

Controlled multi-family development study. A separate set includes benchmark-family, B-spline, Fourier, Matérn/Karhunen–Loève, and localized-bump fields, with 512 training, 40 validation, and 40 Wavelet development surfaces. Price/directional labels use (C.5). An initial derivative-informed study was confounded by a floor on a loss coefficient, output precision, terminal handling, and output-stage ambiguity; those results do not prove intrinsic target incompatibility.

The controlled follow-up uses one fixed-sensor late-fusion architecture and three configurations: value supervision plus preservation; that baseline plus calibrated JVP/mixed-derivative supervision; and a control retaining the coefficient floor. All retain the preservation term, corrected output implementation, three seeds, and 7,000 updates. A 2,000-resample bootstrap clusters the 40 base surfaces, retaining both directions and all three fixed trained seeds in each sampled cluster. Its intervals condition on these models and reused development surfaces, not new seeds or prospective unseen families. Table 24 reports the resulting conditional directional and value-error ratios.

Table 24: Multi-family development paired error ratios, geometric means. Conditional 95% intervals cluster surfaces, retaining three fixed seeds. Lower than one is improvement.

Quantity	Error ratio	Conditional 95% interval
Composite directional error	0.584911	[0.554260, 0.618680]
Price error	2.413665	[2.344490, 2.484836]
Composite value error	1.430692	[1.416221, 1.444485]

The three within-seed directional ratios are approximately 0.5834, 0.5017, and 0.6837. Nevertheless, benchmark-family, Wavelet, and worst-family price-preservation criteria are not met. Removing the coefficient floor improves pooled results versus that control, with heterogeneous seed/direction effects. No tested configuration meets both price-preservation and directional-response criteria. This development study targets a distinct beyond-family response objective from the held-out ID and parameter-support evaluations. A prospective new-family test has not been run and would address a distinct extrapolation setting.

A separate variable-sensor study has worst 96-versus-231-sensor price-error ratio 2.2235. Its population is not pooled with the 40-surface JVP experiment. The fixed-sensor follow-up does not retest sensor transfer. Resolution acceptance and numerical accuracy after a discretization change are different properties (Gao et al., 2025; Sakarvadia et al., 2026). Principal pricing, spot-Greek, and quantum-compatible results continue to refer to the benchmark-family FI model, not these development models.

C.7 Historical terminal sensitivity, PP, and recorded timing

Table 25: Historical including-terminal Greek sensitivity and PP price/PDE diagnostics, mean $\pm$ sample SD over five seeds. Price uses 32 ID surfaces; Greeks use the 24-surface L3 subset with 3,239 pairs including 79 terminal pairs. These are not Table 8's 3,160-pair nonterminal Greeks. A dash indicates no independently differentiated continuous PP output.

Metric	PI	FI	FI–PP
Global $\mathrm{RelL2}$	0.007244 ± 0.001228	0.0009645 ± 0.0000488	0.001043 ± 0.0000455
Global price p95	0.9294 ± 0.1979	0.1331 ± 0.0046	0.1330 ± 0.0037
Near price p95	2.3338 ± 0.3268	0.1579 ± 0.0073	0.1579 ± 0.0073
Delta p95	0.1726 ± 0.0190	0.02535 ± 0.00188	—
Gamma p95	0.03760 ± 0.00075	0.006469 ± 0.000038	—
PDE residual RMS	0.01049 ± 0.00151	0.007025 ± 0.000224	0.006931 ± 0.000220

These historical and primary masks view the same frozen predictions, not different trained models. PP slightly worsens global price error. Identical displayed near-strike p95 before/after PP does not imply pointwise equality. Although the PDE residual RMS omits the terminal evaluation row, its last centered interior time stencil uses terminal prices, so the reported near-maturity residual is not independent of the terminal-value convention. Under the including-terminal mask, carrier-only Delta/Gamma p95 is $0.0945468/0.00874230$, versus FI $0.0253491/0.00646887$, reductions of approximately $73.19\%/26.00\%$. This comparison includes the learned correction and output map jointly; it is not mixed into the primary nonterminal attribution.

Table 26 records the historical timers; their unequal scopes are explained below.

Table 26: Recorded CPU timing, not an equal-scope efficiency comparison. Training includes validation but PI/FI timer boundaries differ.

Quantity	PI	FI	FI–PP
Training time (s)	4779.8 ± 196.8	437.6 ± 42.9	Same checkpoint
Inference time (s)	0.1490 ± 0.0031	0.1266 ± 0.0021	0.2055 ± 0.0055

PI excludes initial dataset/model preparation and later held-out inference, but includes collocation, differentiation, checkpointing, and periodic log rewrites. FI also includes initial data/reference preparation and final selected-checkpoint held-out inference/prediction writing. PI has 7,000 updates and 72 validation calls; differentiation time is not separately recorded. Historical mean training totals for DeepONet, Normalized DeepONet, normalized soft penalties, normalized FAL, and FI are 41.78, 475.89, 536.49, 499.24, and 437.61 seconds, respectively, with 258,305 parameters each. Neither equal timer scope nor architecture-only efficiency follows from these values. No amortization crossover is inferred without label generation and training costs.

C.8 Upper-boundary target sensitivity

The canonical PI comparator uses an asymptotic call-price target at the query edge $S = 220$, whereas the ID reference solver extends to $S = 605$. We test this difference by replacing only the upper-boundary target with the interpolated reference-solver trace at $S = 220$. The architecture, optimizer schedule, collocation protocol, loss weights, validation checkpoint rule, and five seeds remain fixed. The sensitivity was predeclared before training; all five runs completed the 7,000-update protocol. Per-seed configurations, training logs, and selected-checkpoint hashes are retained.

Table 27 shows small changes in PI errors and a persistent gap to FI. The tested boundary-target difference therefore does not account for the principal FI–PI separation.

Table 27: Upper-boundary target sensitivity: five-seed means under the principal evaluation protocol. Trace-matched PI changes only the upper-boundary target. Lower is better; bold marks the lowest mean in each row.

Metric	Canonical PI	Trace-matched PI	FI
Price RelL2	0.00724	0.00723	**0.000965**
Near price p95	2.3338	2.3051	**0.1579**
Delta p95	0.1547	0.1528	**0.0245**
Gamma p95	0.0355	0.0358	**0.00434**
PDE residual RMS	0.01049	0.01045	**0.00703**

D QUANTUM-COMPATIBLE REALIZATIONS AND RESOURCE ACCOUNTING

This appendix studies two quantum-compatible realizations of a pretrained dense FI model. QO exactly compiles selected frozen linear maps through signed-core orthogonal factorizations, whereas QDO replaces eligible dense maps by restricted diagonal–orthogonal (DQ) parameterizations followed by supervised adaptation. These constructions differ from Quantum DeepONet, which trains an orthogonally parameterized classical analogue and transfers the learned angles for quantum-circuit evaluation (Xiao et al., 2025), and from recent quantum neural operators based on alternative structured parameterizations or spectral embeddings (Wang et al., 2026; Abid et al., 2026). All reported results here are classical or idealized algebraic evaluations: no physical routed depth, finite-shot quantum execution, or measured end-to-end hardware speedup is claimed.

D.1 EXACT FROZEN-MODEL COMPILATION

For $\mathbf{W} \in \mathbb{R}^{m\times n}$, take a full SVD $\mathbf{W} = U\Sigma V^\top$. Select diagonal sign matrices D_U, D_V with $\det D_U = \det U$, $\det D_V = \det V$. Then

$$U_+ = UD_U \in SO(m), \quad V_+ = VD_V \in SO(n), \quad \Sigma_\pm = D_U\Sigma D_V, \qquad \mathbf{W} = U_+\Sigma_\pm V_+^\top. \tag{D.1}$$

Because $D_U^2 = D_V^2 = I$, multiplication gives $(UD_U)(D_U\Sigma D_V)(VD_V)^\top = U\Sigma V^\top$. Determinants place the factors in the special-orthogonal groups; plane-rotation elimination gives a finite Givens decomposition. Reflection signs must remain in the signed core. Unary amplitude encoding and Givens/RBS circuits implement these orthogonal actions in the unary subspace (Xiao et al., 2025; Kerenidis et al., 2021; Landman et al., 2022). A single replacement $\mathbf{W} \mapsto Q$ would not preserve a general matrix.

QO's eligible set consists of all 15 dense maps. The realized compiler selects `trunk.2`, a 128×128 map evaluated at 3,321 queries, and `fusion.0`, a 128×384 map evaluated once:

$$128^2 \cdot 3321 + 128 \cdot 384 = 54,460,416 \tag{D.2}$$

semantic MACs, or 49.5563% of the whole dense-map census $109,896,064$. In this all-dense-map census, eligible and whole-map coverage coincide. Reconstructed blocks and ideal orthogonal actions pass the recorded checks, with compilation-induced relative price discrepancy 9.06×10^{-17}. This measures an idealized selected computation, not finite-shot accuracy. The signed singular core, biases, nonlinearities, carrier, FAL, and other retained operations remain classical; hardware additionally needs preparation, readout, and precision control.

D.2 STRUCTURED REPRESENTATION, LIMITATIONS, AND INITIALIZATION

For a square block, classical Procrustes gives (Schönemann, 1966)

$$\min_{Q^\top Q=I} \|\mathbf{W} - Q\|_F^2 = \sum_i (\varsigma_i - 1)^2, \qquad Q_* = UV^\top. \tag{D.3}$$

Here ς_i are the singular values of $\mathbf{W}$. The restricted diagonal–orthogonal family used by QDO is

$$\mathbf{W} \approx DQ, \qquad D = \mathrm{diag}(d_1, \ldots, d_n), \quad Q = Q_{\mathrm{circ}} \in \mathcal{Q}_{\mathrm{circ}} \subseteq O(n). \tag{D.4}$$

Positive scales are parameterized by $d_i = e^{s_i}$; determinant/reflection signs require separate treatment. For a fixed Q, rowwise least squares gives $d_i = \langle w_i, q_i \rangle$ without sign constraints, or $\max(\langle w_i, q_i \rangle, 0)$ for nonnegative scales. A zero optimum is approached but not attained by a strictly positive exponential parameterization.

Proposition D.1 (Representability and approximation obstruction)**.** *For square* $\mathbf{W}$*, unrestricted* $Q \in O(n)$ *and nonnegative diagonal* D *represent* $\mathbf{W} = DQ$ *exactly iff its nonzero rows are mutually orthogonal, equivalently* $\mathbf{W}\mathbf{W}^\top$ *is diagonal. Put* $\gamma_{\mathrm{row}}(\mathbf{W}) = \left\|\mathrm{offdiag}(\mathbf{W}\mathbf{W}^\top)\right\|_F$ *and* $\mu_W = \|\mathbf{W}\|_{\mathrm{op}}$*. Every such* $A = DQ$ *satisfies*

$$\|\mathbf{W} - A\|_F \geq \sqrt{\mu_W^2 + \gamma_{\mathrm{row}}(\mathbf{W})} - \mu_W. \tag{D.5}$$

Restricting Q *to* $\mathcal{Q}_{\mathrm{circ}}$ *cannot reduce the optimum error.*

Proof. $\mathbf{W} = DQ$ implies $\mathbf{W}\mathbf{W}^\top = D^2$. Conversely, normalize the nonzero orthogonal rows and complete an orthonormal row set at zero rows; row norms form D. For the bound, put $E = \mathbf{W} - A$. Since $AA^\top = D^2$,

$$\begin{aligned}\gamma_{\mathrm{row}}(\mathbf{W}) &\leq \left\|\mathbf{W}\mathbf{W}^\top - AA^\top\right\|_F = \left\|E\mathbf{W}^\top + AE^\top\right\|_F \\ &\leq \|E\|_F \left(\|\mathbf{W}\|_{\mathrm{op}} + \|A\|_{\mathrm{op}}\right) \leq \varepsilon(2\mu_W + \varepsilon), \qquad \varepsilon = \|E\|_F .\end{aligned} \tag{D.6}$$

Solving the quadratic inequality proves (D.5). A smaller circuit set cannot improve its minimum. □

Row orthogonality is not sufficient for a restricted-depth family. An $SO(n)$-only, positive-scale model needs the appropriate determinant sign unless an external sign operation is retained. Row geometry, circuit restrictions, and fitting error can all contribute; high-coverage error is not automatically an optimization failure.

Initial substitutions fit dense activations X and outputs Y using the nonnegative-clamped normalized quadratic objective

$$\frac{\mathrm{tr}(YY^\top) - 2\langle M, YX^\top \rangle + \mathrm{tr}(MXX^\top M^\top)}{\mathrm{tr}(YY^\top)}, \qquad M = \mathrm{diag}(e^s) Q_{\mathrm{BB}}. \tag{D.7}$$

Here $Q_{\mathrm{BB}} = Q_{\mathrm{BB}}(\theta) \in \mathcal{Q}_{\mathrm{circ}}$ denotes the implementation-specific parameterized orthogonal factor used in the activation fit. The retained records identify the logical stage count and aggregate parameterization but do not recover the within-stage pairing/permutation schedule, so no more specific circuit topology is inferred from the angle count. Angles and log scales start at zero. LBFGS uses learning rate 0.8, maximum 120 iterations/240 evaluations, history size 30, strong-Wolfe line search, gradient tolerance 10^{-10}, and change tolerance 10^{-13}. The selected `trunk.2` initialization is reused; `trunk.4` is fitted independently of coverage. This initialization is followed by supervised adaptation; the matched attribution study below separates structured-map-only from full-network updates.

D.3 Supervised adaptation, eligible coverage, and implementation cost

The parent is FI-DeepONet (archive alias E5), seed 7713, update 6600. QPF is the historical protocol identifier for QDO supervised adaptation, not an additional model family. The historical QPF protocol adapts all registered trainable parameters after structured initialization; its selected C94 instance for seed 4262703989 is update 1900. For each full 128-dimensional map, the archived structured parameterization contains eight logical Givens stages, totaling 3,584 rotation angles, together with 128 diagonal scale parameters and bias. The retained manuscript package does not recover the within-stage pairing/permutation schedule or any stage-to-stage reordering; accordingly, "stage" denotes a logical parameterization stage rather than routed circuit depth. QPF fits reconstructed pre-FAL normalized-price MSE on 410 original-training surfaces, holding out a disjoint training-only 102-surface subset for selection. Batches are uniform 32 surfaces by 512 queries. Learning rate stays 10^{-3} through update 800, then follows cosine decay to 10^{-5} at update 2500. The matched attribution experiment in Appendix D.4 reuses the same data/update/schedule semantics but freezes unreplaced parameters in its local-only arm. C100 is a separate 3,500-update endpoint diagnostic, not part of the four-candidate knee selection.

QDO admits only ten 128×128 hidden maps: $C_{\text{elig}} = 108,953,600$ versus $C_{\text{whole}} = 109,896,064$ dense-map MACs. C94 substitutes `trunk.2` and a selected subset of `trunk.4`, recorded in the archived summary as fourteen output blocks, with

$$\begin{aligned} C_{\text{repl}}^{94} &= 102,021,120, & \rho_{94} &= 0.9363721804511278, & \omega_{94} &= 0.9283418922082596, \\ C_{\text{repl}}^{99} &= 108,822,528, & \rho_{99} &= 0.998796992481203, & \omega_{99} &= 0.9902313516888103, \\ C_{\text{repl}}^{100} &= 108,953,600, & \rho_{100} &= 1, & \omega_{100} &= 0.9914240422659724. \end{aligned} \tag{D.8}$$

Here $f_{\text{elig}} = C_{\text{elig}}/C_{\text{whole}}$, $\rho = C_{\text{repl}}/C_{\text{elig}}$, and $\omega = f_{\text{elig}}\rho$. The symbol α remains reserved for the maturity-rescaling exponent. Every count uses the same semantic MAC convention. They are not percentages of all end-to-end computation.

The saved partial implementation evaluates both the full classical dense map and the structured map, then overwrites selected outputs. Replaced-MAC coverage therefore does not measure current runtime reduction. The denominator excludes reference-price evaluation, nonlinearities, latent contraction, FAL, PP, data movement, state preparation, readout, and other quantum overhead. Materialized DQ remains an ordinary dense matrix when evaluated classically.

Figure 12 separates exact matrix factorization, the restricted QDO family, and the denominators used for semantic MAC coverage. No arrow or area in this schematic represents measured latency.

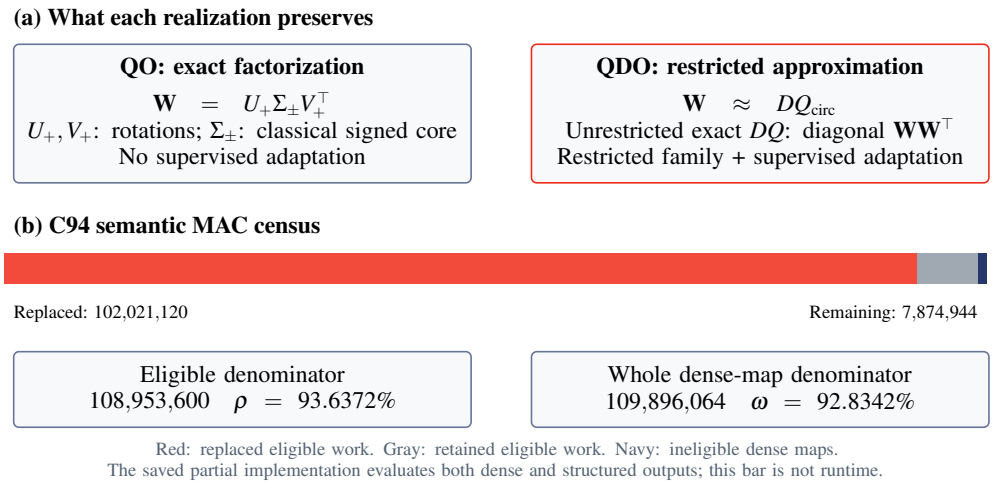


Figure 12: Quantum-compatible realization scope. (a) QO preserves the selected matrix through two orthogonal factors and a classical signed core; QDO restricts the matrix family and then adapts it. The diagonal-Gram condition is necessary and sufficient only for unrestricted $Q \in O(n)$ with nonnegative scales, and is merely necessary for the restricted circuit family. (b) C94 replaced-MAC accounting: the small navy portion is ineligible dense-map work, so eligible coverage ρ exceeds whole-map coverage ω. Counts refer to the same semantic dense-map workload, excluding preparation, measurement, carrier evaluation, nonlinearities, and other overhead. Neither the proportional bar nor QO exactness demonstrates hardware speedup.

D.4 Matched adaptation attribution

To separate recovery by the structured maps from compensation elsewhere in the network, we rerun C94 with the same parent, replacement support, 410/102 training-only fitting/selection split, mini-batch schedule, optimizer schedule, update budget, and checkpoint rule. The only intervention is the trainable parameter set. The frozen arm applies the fitted DQ substitution without supervised adaptation; the local arm updates only the substituted structured parameters (and their replaced biases where applicable); the full arm follows the historical QPF scope and updates all registered trainable parameters. Three paired QPF seeds are used for the trainable arms.

Table 28 compares fidelity recovery and unreplaced-parameter drift under this matched intervention.

Table 28: C94 matched adaptation attribution on the locked 32-surface synthetic evaluation. Local and full rows are three-seed mean ± sample SD. The recovery fraction is computed on global price RelL2 from the frozen substitution toward the full-adaptation result.

Arm	Global price RelL2	Unreplaced-parameter drift
Dense FI parent	0.000836816	—
DQ frozen	0.00808648	0
DQ local	0.00180191 ± 0.0000673	0
DQ full	0.00178871 ± 0.0000160	0.08698 ± 0.00200

Local-only adaptation accounts for 99.79% of the frozen-to-full global-RelL2 recovery. Full adaptation has a slightly lower mean error, and its unreplaced parameters move measurably (relative drift 0.08698 ± 0.00200; replaced-parameter drift 0.02010 ± 0.00091), but the extra full-over-local error reduction is only about 0.73% on the global-RelL2 scale. Parameter drift alone is not causal evidence; these results therefore support recoverability within the structured maps but do not support full-network compensation as necessary for the bulk of fidelity recovery. A historical block-sparse proxy lacks recoverable matched adaptation semantics, so no equal-adaptation DQ-versus-block-sparse claim is made.

D.5 Training-only score and knee selection

With $\varepsilon = 10^{-12}$, the six-component selection score is

$$J = 0.35\log(\varepsilon + L_g) + 0.20\log(\varepsilon + L_n) + 0.15\log(\varepsilon + L_\Delta) + 0.15\log(\varepsilon + L_\Gamma) + 0.10\log(\varepsilon + L_w) + 0.05\log(\varepsilon + L_p), \qquad R_{\mathrm{comp}} = e^{J - J_{\mathrm{E5}}}. \tag{D.9}$$

L_g is pooled price RelL2, L_n near-strike price p95, and L_Δ, L_Γ global finite-point AD absolute p95. For post-FAL, pre-PP price P, its repaired grid P_{PP}, and $C = C^{\mathrm{ref}}$, $L_p = (\|P_{\mathrm{PP}} - P\|_2 / (\|P - C\|_2 + \varepsilon))^2$. L_w is the maximum of six surface-averaged normalized weak-condition components described below. The earliest minimum (J, step) among evaluations every 100 updates selects the checkpoint. Primary $R_{\mathrm{comp}} = 1.305077932190532$ is a training-holdout composite, not a price-error ratio, a formal-validation statistic, or a hardware metric.

Four candidates C49, C87, C94, and C99 are represented by median selected R_{comp} over three paired seeds. Nondominance minimizes this cost while maximizing eligible coverage and $S_i = G_{\mathrm{ref}}(\eta = c = 0.01)$. With $\ell_i = \log S_i$, define

$$d_i = \left[\left(\frac{R_i - R_{\min}}{R_{\max} - R_{\min} + 10^{-30}} \right)^2 + \left(1 - \frac{\ell_i - \ell_{\min}}{\ell_{\max} - \ell_{\min} + 10^{-30}} \right)^2 \right]^{1/2}. \tag{D.10}$$

Ranges use all four candidates. Select the nondominated point minimizing d_i, breaking ties toward smaller R_i: C94 is selected. Accuracy/financial flags are evaluated afterward, not used as a general feasibility filter. C94 does not pass the frozen ID/OOD accuracy-preservation gates; it must not be described as a successful lossless migration. Independent replay of those gates requires their numerical tolerance/configuration records in addition to these summaries.

Weak-condition components are surrogate diagnostics. Projected prices are reversed in time and transferred by tensor-product linear interpolation from uniform auxiliary coordinates $x_{\mathrm{w}} \in [-1, 1]$, $\tau_{\mathrm{w}} \in [0, 1]$ to grids G1/G2/G4 of sizes $(17, 9)$, $(33, 17)$, and $(65, 33)$. With $b_{\mathrm{w}} = (1 - x_{\mathrm{w}}^2)_+^2 [\tau_{\mathrm{w}}(1 - \tau_{\mathrm{w}})]_+^2$, the three test functions are

$$b_{\mathrm{w}}(1 + 0.2x_{\mathrm{w}} + 0.1\tau_{\mathrm{w}}), \qquad b_{\mathrm{w}}(1 - 0.15x_{\mathrm{w}} + 0.3\tau_{\mathrm{w}} + 0.1x_{\mathrm{w}}\tau_{\mathrm{w}}), \qquad b_{\mathrm{w}}(1 + 0.25x_{\mathrm{w}}^2 - 0.2\tau_{\mathrm{w}} + 0.1\tau_{\mathrm{w}}^2). \tag{D.11}$$

Each grid component averages squared weak residuals integrating $\phi\,\partial_{\tau_{\mathrm{w}}} v + 0.02(\partial_{x_{\mathrm{w}}}\phi)(\partial_{x_{\mathrm{w}}} v) + 0.03\phi v$, with zero source/drift. On every grid, $\partial_{\tau_{\mathrm{w}}} v$, $\partial_{x_{\mathrm{w}}} v$, and $\partial_{x_{\mathrm{w}}}\phi$ use custom five-point fourth-order centered differences where two neighbors on each side are available. At the remaining points, `np.gradient(edge_order=2)` supplies second-order centered differences adjacent to the endpoints and second-order one-sided endpoint differences. Nested Simpson quadrature integrates first in x_{w}, then in τ_{w}. Three additional components integrate squared coarse/fine differences

by Simpson quadrature on the finer grid after interpolating the coarser field to that grid, for G1–G2, G1–G4, and G2–G4. Surface means are normalized as $(\varepsilon + \text{candidate mean})/(\varepsilon + \text{E5 mean})$; their maximum is L_w. This is a frozen surrogate weak-condition diagnostic, *not the local-volatility PDE residual*.

D.6 FIDELITY–COVERAGE FRONTIER

The frozen candidate grid is evaluated on the same 32-surface formal population. C49/C87/C94/C99 use the original three paired QPF seeds; C100 uses three different historical seeds and is descriptive only. The mean sequence is nonmonotone, so increasing replacement coverage is not identified with either monotone fidelity loss or monotone improvement.

Table 29 lists both coverage denominators alongside the paired synthetic means.

Table 29: Synthetic fidelity across the frozen coverage grid. C49–C99 are three-seed paired-QPF means ± sample SD. C100 uses a non-paired historical seed registry and is descriptive only. Coverage is analytical dense-map accounting, not measured runtime saving.

Candidate	ρ	ω	Global price $\mathrm{RelL2}$	Near-price p95
C49	0.4994	0.4951	0.0018105 ± 0.0000123	0.1137 ± 0.0062
C87	0.8739	0.8665	0.0017746 ± 0.0000267	0.1427 ± 0.0201
C94	0.9364	0.9283	0.0017887 ± 0.0000160	0.1318 ± 0.0160
C99	0.9988	0.9902	0.0017780 ± 0.0000278	0.1350 ± 0.0204
C100[†]	1.0000	0.9914	0.0017791 ± 0.0000543	0.1401 ± 0.0122

[†]Different historical training seeds; excluded from paired-seed inference. C94 remains the knee selected by the frozen training-only rule in Appendix D.5, not a universal accuracy-optimal coverage.

D.7 SELECTED-C94 FIDELITY AND UNMATCHED PROXY CONTROLS

Table 30: Paired synthetic fidelity on a separate 32-surface population; selected metrics are summarized in Table 2B. The adapted C94 QDO instance is compared with its own dense parent, not the principal five-seed mean. Delta and Gamma use AD on the formal grid. Ratios are QDO/dense; bold marks the lower error in each row.

Metric	Dense FI	QDO-FI	Ratio
Global price $\mathrm{RelL2}$	$\mathbf{0.0008368}$	0.0017702	2.115
Price RMSE	$\mathbf{0.05189}$	0.10978	2.116
Price MAE	$\mathbf{0.03049}$	0.09600	3.149
Near-strike price p95 error	$\mathbf{0.13147}$	0.13238	1.007
Global AD Delta p95 error	$\mathbf{0.01267}$	0.01481	1.169
Global AD Gamma p95 error	0.002441	$\mathbf{0.002237}$	0.916

Table 30 shows a fidelity trade-off: global price error roughly doubles, near-strike p95 changes little, and Gamma p95 decreases slightly. The selected C94 instance does not meet the fixed ID/OOD accuracy-preservation criteria. Formal synthetic Greeks use float64 AD on the formal grid; unlike training-only score evaluation, the formal evaluator does not discard nonfinite values, potentially giving nonfinite OOD summaries.

Table 31 retains the historical proxy-budget controls; their adaptation protocols are not matched.

The controls use the selected blocks and nested coverage levels, with no market rows in fitting or selection. QDO has lower listed price errors than these structured classical controls at the declared proxy budgets. Because equal trainable parameter sets, adaptation data, update budgets, and selection rules are not recoverable for these historical controls, this comparison is descriptive and does not isolate the effect of the DQ parameterization. The matched experiment in Appendix D.4 addresses adaptation scope within QDO but does not retrofit matched training semantics onto these controls. Coverage trends can be nonmonotone because layer perturbations reinforce or cancel; the trend alone cannot identify whether fitting was independent or joint.

Table 31: Structured-approximation price $\mathrm{RelL2}$ on the matched 32-surface population. Coverage columns use the QDO eligible denominator. QO is an exact algebraic reference, not a new measured hardware realization at each column. Budgets match declared arithmetic/storage proxies, not QPU latency. Lower is better; bold marks the lowest error in each coverage column, including the tied dense and exact-compilation references.

Realization / control	93.6%	99.9%	100%
Dense FI	**0.000836816**	**0.000836816**	**0.000836816**
QO exact control	**0.000836816**	**0.000836816**	**0.000836816**
QDO matrix, classical evaluation	0.001770246	0.001747465	0.001716484
Block sparse, arithmetic matched	0.0224863	0.0269211	0.0398960
Truncated SVD, arithmetic matched	0.137667	0.141756	0.169582
Truncated SVD, storage matched	0.153499	0.177486	0.219893
Diagonal + low rank, matched	0.160663	0.180123	0.221246

Representative whole-model proxy budgets at these three coverages are $(59,682,544,1,879,944)$, $(56,334,976,1,863,688)$, and $(56,270,464,1,052,680)$ in MACs and bytes. They define the matching envelope, not actual dense-materialized runtime/storage gains. The measured host-side materialization median for a full FI surface is 0.0155 ms, not evidence that readout dominates current CPU latency or determines QPU costs.

D.8 Perturbation propagation and conditional resource scenarios

For a sequential network, let κ_j bound activation Lipschitz constants, B_j bound the hybrid input to layer j, and M_j bound old/new matrix operator norms. For a *matrix-substitution comparison with the biases held fixed*, a one-layer-at-a-time hybrid-network argument gives

$$\left\|\widetilde{h}_L - h_L\right\| \le \sum_{j=0}^{L-1} \kappa_j B_j \left\|\widetilde{\mathbf{W}}_j - \mathbf{W}_j\right\|_{\mathrm{op}} \prod_{k=j+1}^{L-1} \kappa_k M_k. \tag{D.12}$$

At a substituted node the immediate error is then bounded by $\kappa_j B_j \left\|\widetilde{\mathbf{W}}_j - \mathbf{W}_j\right\|_{\mathrm{op}}$; propagating downstream and summing proves the result. If biases also differ, the immediate term becomes $\kappa_j\left(B_j \left\|\widetilde{\mathbf{W}}_j - \mathbf{W}_j\right\|_{\mathrm{op}} + \left\|\widetilde{b}_j - b_j\right\|\right)$ before the same downstream product is applied. Thus (D.12) is a fixed-bias perturbation bound and is *not* a bound for the full QDO adaptation, where all registered trainable parameters may change. The same construction applies nodewise to the branch–trunk graph. FAL with fixed bounds does not amplify the final scalar price perturbation, but Delta/Gamma require separate differentiation.

Let $\eta \in [0,1)$ denote an assumed cost fraction for replaced work and $c \ge 0$ overhead, normalized to the same reference cost as ω. Define

$$G_{\mathrm{ref}} = \frac{1}{1-\omega(1-\eta)+c}. \tag{D.13}$$

For positive denominator, $k > 1$, and $\omega > 0$, $G_{\mathrm{ref}} \ge k$ iff

$$\omega \ge 1 - 1/k, \quad 0 \le c \le \omega - (1-1/k), \quad 0 \le \eta \le 1 - \frac{1-1/k+c}{\omega}. \tag{D.14}$$

At $\omega = 1-1/k$, only $c = \eta = 0$ reaches equality. Thus $\omega = 0.9$ permits a tenfold gain only at the zero-cost/zero-overhead boundary. Table 32 evaluates the conditional cost formula and tenfold-gain feasibility intercepts.

C94's census reconciles $102,021,120$ replaced and $7,874,944$ remaining operations to $109,896,064$. This is a dense-linear-map decomposition, not complete model work. A full query returns 3,321 classical prices; selected market quotes are a different output workload.

An end-to-end model must include target device, preparation, communication, precision, routing, shots, recovery, and unreplaced work (Rebentrost et al., 2018; Stamatopoulos et al., 2020; Chakrabarti et al., 2021; Herman et al., 2026; Guseynov et al., 2026; Chen et al., 2026a). With quantum workload $\mathcal{W}_Q(H)$ and compatible throughput Θ_H,

$$T_{\mathrm{hybrid}}(H) = T_{\mathrm{nonQ}}(H) + \mathcal{W}_Q(H)/\Theta_H, \qquad S_{\mathrm{e2e}}(H) = T_{\mathrm{classical}}(H)/T_{\mathrm{hybrid}}(H). \tag{D.15}$$

Table 32: Analytical scenarios, not measured speedups. G_{ref} uses $\eta = c = 0.01$; intercepts describe $G_{\mathrm{ref}} \geq 10$. Eligible coverage ρ and whole dense-map coverage ω differ.

ρ	ω	G_{ref}	$c_{\max}$ at $\eta = 0$	$\eta_{\max}$ at $c = 0$
93.6%	0.9283419	10.996	0.0283419	0.0305296
99.9%	0.9902314	33.703	0.0902314	0.0911215
100%	0.9914240	35.100	0.0914240	0.0922149

For positive quantum work, $S_{\mathrm{e2e}} \geq k$ requires $T_{\mathrm{nonQ}} < T_{\mathrm{classical}}/k$ and

$$\Theta_H \geq \frac{\mathcal{W}_Q(H)}{T_{\mathrm{classical}}(H)/k - T_{\mathrm{nonQ}}(H)}. \tag{D.16}$$

A device throughput can be substituted only when workload, accuracy, and execution conventions match. Returning an explicit M-point classical surface needs $\Omega(M)$ scalar slots/materialization operations. A stronger $\Omega(M \log(B/\varepsilon))$ bit bound needs a full output cube $[-B, B]^M$, which is not asserted for financial surfaces. If a per-scalar output cost c_{out} is measured for the explicit-list task, $T_{\mathrm{hybrid}} \geq T_{\mathrm{fixed}} + c_{\mathrm{out}} M$. Returning an amplitude-encoded state, an expectation estimate, or a few prices is a different task; such quantum pricing bounds do not directly certify acceleration for explicit full-surface output.

E Real-Market Evaluation with Constant-Volatility Inputs

E.1 Data identity, quote provenance, and input mapping

The study uses Nasdaq-100 index option records identified as NDX/.NDX: an OptionMetrics extract through WRDS with LSEG-linked underlying/identifier fields. The observation window is 13 May–7 August 2026, covering 60 trading dates. The own-IV protocol contains 13,131 calls. The primary held-out-IV protocol begins on 26 May and contains 2,586 targets from 101 date–expiry chains on 52 dates. All 2,586 targets are uniquely recoverable by source-row identifier and a composite quote key. Raw option fields include `best_bid` and `best_offer`; bid, offer, mid, IV, and target identity are carried from the same daily source row, and frozen model predictions align to that row. Quote time is missing for all rows, so intraday timestamp synchrony is not established. The data are market-derived *constant-volatility inputs*, not reconstructed local-volatility fields.

Cached q is a decimal dividend-yield feature and cached r a decimal one-year zero-coupon-rate proxy shared across 7–90-day contracts on each date. For quote i, let

$$\tau_i = \mathrm{DTE}_i/365, \qquad F_i = S_i e^{(r_i - q_i)\tau_i}, \qquad S_i^{\mathrm{eff}} = S_i e^{-q_i \tau_i}. \tag{E.1}$$

The one-year zero-dividend model is queried via

$$K_i^{\mathrm{model}} = 100, \quad S_i^{\mathrm{model}} = 100 S_i^{\mathrm{eff}}/K_i^{\mathrm{mkt}}, \quad \bar{t}_i = 1 - \tau_i, \quad [K, r, \sigma_0, \beta, \gamma] = [100, r_i, \widehat{\sigma}_i, 0, 0]. \tag{E.2}$$

All 231 volatility sensors equal $\widehat{\sigma}_i$; the 101 payoff sensors use strike 100. Own-IV input sets $\widehat{\sigma}_i = \mathrm{IV}_i/100$; held-out input reconstructs it as below. Since $\beta = \gamma = 0$, the corresponding continuous model has spatially and temporally constant variance at the supplied rate/volatility pair. Under the effective-spot zero-dividend mapping in (E.2), the exact same-input solution is therefore the Black–Scholes carrier itself:

$$V = C^{\mathrm{ref}}, \qquad W = V - C^{\mathrm{ref}} = 0, \qquad R^* = 0 \quad (\tau > 0). \tag{E.3}$$

Thus this experiment checks whether the learned correction remains consistent on a known constant-volatility subfamily; it is not evidence of general local-volatility market transfer. Since $T_{\mathrm{model}} = 1$, $\bar{t}_i = 1 - \tau_i$ retains the actual remaining maturity, not a stretched one-year expiry. Training standardizers remain unchanged; no market scaler is fitted. For dimensionless network output $\widehat{c}_i = \widehat{V}_i^{\mathrm{model}}/K_i^{\mathrm{model}}$, recover

$$\widehat{C}_i^{\mathrm{mkt}} = K_i^{\mathrm{mkt}} \widehat{c}_i. \tag{E.4}$$

The dividend substitution is a constant-volatility quote mapping, not a general dividend-removal identity for every state-dependent local-volatility field.

E.2 HELD-OUT IV, SPREAD-AWARE ERROR MEASURES, AND NEURAL BASELINES

Within a date–expiry chain, strikes and source-row identifiers determine order; duplicate strikes retain the smallest source identifier. Endpoints stay in the fit. Internal positions 4, 9, 14, and so forth, under recorded one-based indexing, are withheld jointly. A natural cubic spline fits $(\log(K/F), \mathrm{IV}/100)$ using retained quotes, with extrapolation disabled. Recorded checks find no target in its own fit, no extrapolated target, and no invalid interpolation. The original eligibility rule used the target's own-IV domain flag before fitting, so predictions were target-value blind but eligibility was not logically fully target-selection blind. A sensitivity based only on reconstructed IV and target-independent metadata gives the *identical realized population*: 2,586 targets, 101 chains, 52 dates, zero additions/removals, and unchanged learned-model ordering.

For market mid $\mathbf{C}^{\mathrm{mkt}}$, same-input Black–Scholes values $\mathbf{C}^{\mathrm{BS}}$, and model $\widehat{\mathbf{C}}$, distinguish

$$\mathrm{rRMSE}_{\mathrm{mkt}} = \frac{\left\|\widehat{\mathbf{C}} - \mathbf{C}^{\mathrm{mkt}}\right\|_2}{\|\mathbf{C}^{\mathrm{mkt}}\|_2}, \qquad \mathrm{RelL2}_{\mathrm{op}} = \frac{\left\|\widehat{\mathbf{C}} - \mathbf{C}^{\mathrm{BS}}\right\|_2}{\|\mathbf{C}^{\mathrm{BS}}\|_2}. \tag{E.5}$$

With bid B_i, offer A_i, mid $M_i = (A_i + B_i)/2$, and half-spread $H_i = (A_i - B_i)/2$, define

$$E_i^{\mathrm{spr}} = \frac{|\widehat{C}_i - M_i|}{H_i}, \qquad D_i^{\mathrm{BS}} = \frac{|\widehat{C}_i - C_i^{\mathrm{BS}}|}{H_i}, \qquad I_i = \mathbf{1}\{B_i \le \widehat{C}_i \le A_i\}. \tag{E.6}$$

All 2,586 held-out-IV rows have finite positive spreads and pass this spread-eligibility gate. ATM-short p95 is measured against the same-input Black–Scholes operator, not market mid, using $|\log(K/F)| \le 0.03$ and 7–30 DTE. All quote predictions are before PP.

Table 33: Market-derived evaluation. Neural entries are five-seed mean ± sample SD. Operator error and ATM-short p95 use the same-input Black–Scholes reference; their Black–Scholes entries are zero by definition. "Normalized" denotes Normalized DeepONet. Lower is better.

Input	Configuration	Market rRMSE	Operator RelL2	ATM-short p95
Own IV	Black–Scholes	**0.54%**	**0**	**0**
	Normalized	$(19.23 \pm 4.21)\%$	0.1912 ± 0.0416	725.14 ± 109.87
	Normalized + FAL	$(4.14 \pm 0.62)\%$	0.0419 ± 0.0056	80.55 ± 17.85
	FI	$(1.23 \pm 0.15)\%$	0.0143 ± 0.0017	14.42 ± 1.98
Held-out IV	Black–Scholes	**0.98%**	**0**	**0**
	Normalized	$(19.64 \pm 4.29)\%$	0.1952 ± 0.0424	722.56 ± 110.59
	Normalized + FAL	$(4.26 \pm 0.61)\%$	0.0417 ± 0.0056	79.52 ± 18.02
	FI	$(1.53 \pm 0.13)\%$	0.0145 ± 0.0017	14.55 ± 2.02

Table 34: Spread-aware held-out-IV consistency on the unchanged 2,586-target population. Neural predictions are arithmetic means of the five frozen seed predictions, evaluated after averaging prices; they are not means of five separately computed metrics. Market error is relative to quote mid; BS deviation is relative to the exact same-input Black–Scholes operator. Spread quantities use half-spread H_i, so a market-mid error at most one half-spread is necessary and sufficient for lying inside the bid–offer interval.

Model	Market rRMSE	Operator RelL2	Inside spread	Median E^{spr}	Median D^{BS}
Black–Scholes	0.980%	0	93.74%	0.215	0
Normalized	19.56%	0.1943	18.72%	11.391	10.997
Normalized + FAL	4.05%	0.03958	39.48%	1.612	1.471
FI	1.519%	0.01444	50.04%	0.998	1.201

The ordering of learned models survives both input protocols. Black–Scholes remains more accurate against quoted prices with the same volatility information. FI nevertheless moves the learned approximation into a market-relevant scale: its median market-mid error is about one half-spread, while its median deviation from the exact same-input Black–Scholes operator is 1.201 half-spreads. The corresponding p95 values are 4.800 and 4.805 half-spreads, so an unqualified "small relative

to spread" claim is not supported. This is approximation consistency, not market local-volatility generalization or calibration.

Figure 13 separates the five-seed metric summary from the diagnostic computed on averaged seed predictions. This distinction explains the slightly different FI market rRMSE values 1.53% and 1.519%: the population is unchanged, but the aggregation is different.

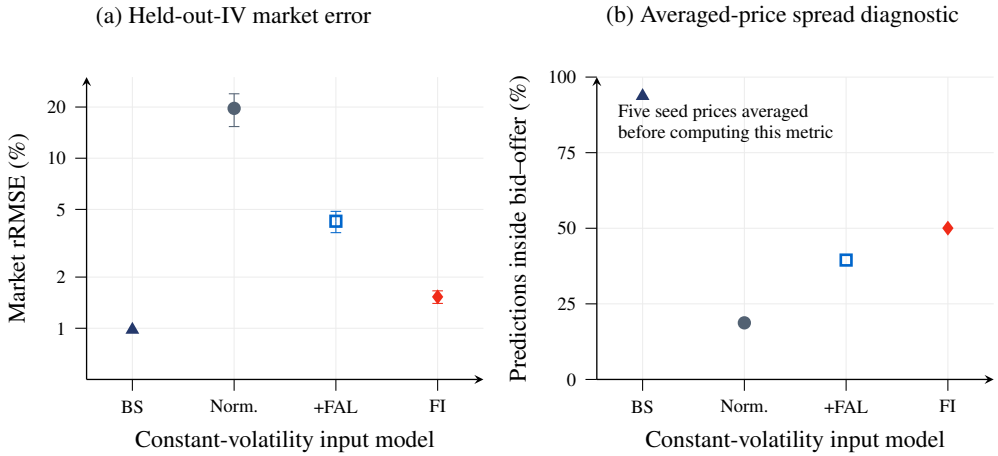


Figure 13: Two aggregation conventions on the same 2,586 held-out-IV quotes. (a) Per-seed market rRMSE, summarized by five-seed mean and sample SD, from Table 33; Black–Scholes is deterministic and has no seed error bar. The vertical axis is logarithmic. (b) Fraction inside the bid–offer interval, computed from the arithmetic average of five seed prices, from Table 34. No interval is supplied for (b). "Norm." denotes Normalized DeepONet and "+FAL" its FAL-only extension. FI improves on both neural baselines, while same-input Black–Scholes remains more accurate and more frequently inside the spread. The panels do not constitute local-volatility calibration or a trading-performance test.

E.3 Temporal and IV-reconstruction robustness

The following temporal and IV-reconstruction diagnostics use the arithmetic mean of the five frozen FI seed predictions, as in Table 34. They are distinct from both the per-seed statistics in Table 33 and the single-parent QDO comparisons. The 52 held-out-IV dates are frozen chronologically before evaluation: the final 16 dates (1,376 quotes) form a sealed robustness subset. No model retraining or checkpoint selection uses either period. On the sealed dates, FI market rRMSE is 1.452%, operator RelL2 is 1.444%, and inside-spread fraction is 55.60%; same-input Black–Scholes market rRMSE is 0.893% and inside-spread fraction 95.13%. The direction and scale of the operator-consistency result therefore persist late in the sample.

Table 35 holds the target population fixed while changing only the target-excluded IV interpolator.

Table 35: Held-out-IV reconstruction sensitivity on the identical 2,586 targets. Natural cubic is the frozen primary protocol; PCHIP and linear interpolation use the same withheld positions, target exclusion, endpoint treatment, and no extrapolation.

IV method	IV RMSE (decimal)	FI market rRMSE	FI operator RelL2	FI inside spread
Natural cubic	0.003206	1.519%	0.0144449	50.04%
PCHIP	0.002708	1.437%	0.0144411	50.08%
Linear	0.002651	1.427%	0.0144395	50.62%

Across these reconstruction methods the FI operator RelL2 range is only about 0.037% relative, supporting stability to reasonable target-excluded IV interpolation choices rather than a method-selection advantage.

E.4 MULTI-COVERAGE QDO/FI MARKET BRIDGE AND ERROR ALIGNMENT

All market-bridge evaluations are frozen inference: no training, fitting, calibration, coverage selection, seed selection, checkpoint selection, filter change, IV-method change, or temporal re-splitting uses market outcomes. C49/C87/C94/C99 each use the same three paired QPF seeds and descend from the same dense FI parent. C100 uses a different historical seed registry and is descriptive only. A frozen C94 replay reproduces the historical 2,586-row predictions to maximum absolute discrepancy 4.55×10^{-13} and exactly reproduces the reported market and operator metrics within floating-point tolerance.

Table 36 summarizes the paired coverage comparisons and their conditional chain/date intervals.

Table 36: Frozen quantum-to-market bridge. Full-sample ratios are geometric means over three paired QPF seeds; all 12 seed-level ratios are below one. Chain/date intervals use 5,000 cluster resamples conditional on the fixed trained realizations. Sealed ratios use the frozen 16-date subset. Perturbation scales are $|\mathrm{QDO} - \mathrm{FI}|$ divided by half-spread.

Coverage	Full RMSE ratio	Chain 95% CI	Date 95% CI	Sealed ratio	Wins	Median / p95 scale
C49	0.7999	[0.7507,0.8471]	[0.7433,0.8495]	0.8720	3/3	1.215 / 5.373
C87	0.7617	[0.7111,0.8101]	[0.7082,0.8074]	0.8192	3/3	1.137 / 4.877
C94	0.7694	[0.7172,0.8181]	[0.7147,0.8174]	0.8345	3/3	1.154 / 5.139
C99	0.7460	[0.6931,0.7938]	[0.6878,0.7934]	0.8018	3/3	1.045 / 4.793

The C94 market improvement is therefore not unique: all four paired coverages improve market RMSE in all three seeds, and all chain/date conditional intervals lie below one. The sequence is not strictly monotone in coverage, so these data do not support a claim that higher quantum-compatible coverage systematically improves market fit. C100's non-paired historical seeds give a descriptive geometric-mean RMSE ratio 0.7756, broadly consistent with the paired candidates but excluded from paired inference.

For an algebraic decomposition of the market change, write $f_Q = f_{\mathrm{FI}} + \delta_m$ and $e = f_{\mathrm{FI}} - y$, where y is the quote midpoint and $\mathbb{E}$ below is the equal-weight empirical average over the fixed quote population. Then

$$\mathrm{MSE}(f_Q) - \mathrm{MSE}(f_{\mathrm{FI}}) = 2\mathbb{E}[e\delta_m] + \mathbb{E}[\delta_m^2]. \tag{E.7}$$

Across C49/C87/C94/C99, the alignment term is negative in all 12 paired seed-level comparisons and its magnitude exceeds the perturbation energy in all 12. The identity closes to floating-point tolerance. Thus the observed market reduction is consistently associated with favorable alignment of the frozen QDO perturbation with the parent error on this market population. This is sample-dependent error cancellation, not causal evidence of quantum regularization. Synthetic global fidelity is only a partial descriptive predictor of the market ratio (12 paired points); synthetic near-price p95 has the opposite correlation direction, so no single synthetic metric is claimed to predict market performance.

E.5 FEATURE-CACHE PROVENANCE AND RELEASE LIMITS

The fixed cache uses a decimal dividend feature q and decimal one-year rate r, without a second division by 100. The associated rate series is FRED `THREEFY1`; the associated dividend source is `NASDAQNDXDIV`, a dividend-point series, not directly a yield. Original dividend-point observations and historical reset, rolling, point-to-yield, annualization, and compounding transformations are unavailable. Rate export/fill implementation is only partly documented. These gaps are not filled using plausible but unverified formulas.

Downstream quote matching uses exact cache dates and drops rows missing spot, rate, or dividend features. A source check reported a 31 July 2026 rate carried into 3–7 August, a 3–7-calendar-day lag, without recovering the fill code. Exact-date cache matching is therefore not proof of contemporaneous source observations. The raw contract multiplier and currency convention are not populated in the supplied extract; this does not block normalized comparisons because bid, offer, mid, Black–Scholes, and neural predictions share the same quoted price units, but it blocks an independently verified multiplier/currency claim. A replay of ten fixed rows and 90 scalars has maximum absolute difference 9.37×10^{-17} and maximum relative difference 1.68×10^{-14}. Quote time is unavailable, so same-row daily alignment is verified but intraday synchronization is not.

Selected exercise, settlement, correction, and IV-inversion metadata also remain incomplete. Accordingly, the mapping is reported from the fixed cache, not as a fully reconstructed source-to-model pipeline. Data licenses and anonymization must be checked before release; quote-level licensed records are excluded from lightweight manifests.

F Related Work and Reproducibility

Operator learning and financial surrogates. Universal operator approximation (Chen & Chen, 1995), DeepONet (Lu et al., 2021a), Fourier and integral-kernel operators (Li et al., 2021; Kovachki et al., 2023), and multiple-input constructions (Jin et al., 2022) provide ways to amortize families of PDE solutions. Neural pricing and hedging methods (Hutchinson et al., 1994), deep PDE solvers (Sirignano & Spiliopoulos, 2018; Han et al., 2018), Black–Scholes approximation analyses (Berner et al., 2020), and volatility-calibration surrogates (Horvath et al., 2021) establish the broader financial context. Our contribution concerns the representation of the coefficient-to-price map through input-dependent strike matching. Its empirical scope distinguishes held-out ID prediction, parameter-support extrapolation within the benchmark family, and off-family directional responses.

PDE loss and derivative accuracy. Physics-informed learning (Raissi et al., 2019) and its operator variants (Wang et al., 2021b; Li et al., 2024) incorporate known equations into training, with outcomes affected by loss balancing and optimization (Wang et al., 2021a; Krishnapriyan et al., 2021; Chen et al., 2026b; McClenny & Braga-Neto, 2023; Wang et al., 2022). Derivative-informed methods explicitly target sensitivities (O'Leary-Roseberry et al., 2024; Qiu et al., 2024; Luo et al., 2026; Noguer i Alonso, 2026; de Feo, 2026), while EquiNO and MD-PNOP illustrate other ways to incorporate equation structure (Eivazi et al., 2026; Cheng et al., 2026). Our representation uses an analytic carrier independently of a PDE loss. The controlled factorial separately evaluates how adding the PDE loss affects price error, PDE residual RMS, and derivative errors.

Financial constraints and analytic representations. Option-surface restrictions couple contracts (Merton, 1973; Breeden & Litzenberger, 1978; Carr & Madan, 2005), and learning-based calibration must account for their implications (Itkin, 2019). Second-order financial knowledge and smooth or arbitrage-constrained surface models address this structure (Dugas et al., 2000; Ackerer et al., 2020; Wiedemann et al., 2025; Buehler et al., 2026b;a; Aboussalah et al., 2026). Residual-learning methods refine analytic stochastic-volatility prices (Kim et al., 2025); ETCNN uses terminal masks and moneyness normalization (Zhang et al., 2026); and a recent 0DTE model learns a maturity-gated variance correction within a Black–Scholes representation (Sakuma, 2026). Analytic output transforms and hard-constraint constructions provide further precedents (Lagaris et al., 1998; Sukumar & Srivastava, 2022; Lu et al., 2021b). Building on these ideas, the distinction here is strike-line variance matching for a local-volatility operator, together with its correction order, critical scaling, and output-stage analysis. FAL imposes pointwise bounds; cross-contract shape restrictions remain a separate property.

Analysis and quantum-compatible realizations. The correction analysis uses diffusion, parametrix, and short-maturity methods (Corielli et al., 2010; Cheng et al., 2011; Gatheral et al., 2012; Pirjol & Zhu, 2023), together with interior regularity (Friedman, 1964; Lieberman, 1996). Exact compilation uses established Givens/RBS and orthogonal-network constructions (Kerenidis et al., 2021; Landman et al., 2022); the restricted-approximation analysis uses Procrustes geometry (Schönemann, 1966). Quantum DeepONet transfers a trained orthogonal model to circuit evaluation (Xiao et al., 2025), while other quantum neural operators use different structured parameterizations or spectral embeddings (Wang et al., 2026; Abid et al., 2026). Our starting point is an unconstrained dense FI parent, with frozen exact compilation separated from structured approximation and adaptation. Quantum derivative-pricing algorithms (Rebentrost et al., 2018; Stamatopoulos et al., 2020; Chakrabarti et al., 2021; Herman et al., 2026; Guseynov et al., 2026; Chen et al., 2026a) have their own input, output, and hardware assumptions, which must be included when comparing computational costs.

Evaluation scope. The principal five-seed ID comparison, matched five-seed carrier–rescaling ablation, complete parameter-support OOD evaluation, three-seed kink intervention, objective factorial, single-checkpoint off-family test, selected-model realization comparison, and selected-model

market pairing use distinct evaluation protocols. Each metric is specified by its model, population, output stage, mask, reference, and aggregation. Surface-bootstrap intervals condition on the trained models; they do not measure variation over newly trained seeds. The separate sensor-count study is likewise distinct from the multi-family JVP development study. The carrier–rescaling study uses paired initialization and sampling schedules across exponents and reports pre-FAL mechanism metrics separately from post-FAL inference-stage metrics.

Reproducibility scope. The retained synthetic-benchmark records support replay of selected-state evaluations through fixed data, selected checkpoints, saved predictions, evaluation masks, configurations, and checkpoint metadata. The current principal DeepONet baseline is a separate five-seed refit, and the complete parameter-support OOD study is maintained as a separate frozen-state replay. A fresh retraining of every historical principal run has not been performed, and the original split-generation stream remains incomplete. Accordingly, our reproducibility claims concern the fixed realized datasets, audited selected states, and the explicitly identified refits and replays. Licensed quote-level market records are not redistributed.