Variance Forecasting from Forced Flow: Liquidations, the Variance Risk Premium, and Crypto Volatility

This version: June 26, 2026

Abstract

The variance risk premium prices and forecasts variance, and cryptocurrency markets add a signal absent elsewhere: the exchange-published stream of forced perpetual-swap liquidations. We establish this forced-flow tape as a forecastable and tradable state variable for variance. A two-stage CNN–LSTM that times intraday liquidation bursts and forecasts forward variance, combined with a HAR, is the only model **never** excluded from the 75% Model Confidence Set under squared-error loss for Bitcoin or Ether, and its nonlinear encompassing increment over the HAR makes the signal's value irreducibly nonlinear. Trading a replicated variance swap, a liquidation-timing gate delivers net Sharpe ratios near two for ETH weekly and above one for BTC monthly. Methodologically, we reconstruct a daily 30-day implied variance series back to 2018–2019 from option liquidation prints—a free byproduct needing no pre-existing index—recovering Deribit's DVOL at correlations of 0.94 and 0.96 and supplying the long history this literature has lacked.

Keywords: variance risk premium; forced liquidations; realized variance; forecast combination; deep learning; variance swap; implied volatility index; cryptocurrency.

JEL: C53; C58; G12; G13; G17.

1 Introduction

The variance risk premium (VRP), the gap between option-implied and subsequently realized variance, predicts future variance and returns (Bekaert & Hoerova, 2014; Bollerslev, Tauchen, & Zhou, 2009; Carr & Wu, 2009) and is the economic engine of variance-swap payoffs (Demeterfi, Derman, Kamal, & Zou, 1999), including in cryptocurrency markets (Alexander & Imeraj, 2021; Du & Shen, 2025). Cryptocurrency markets also supply a signal that is unavailable elsewhere: the real-time stream of forced liquidations. Perpetual futures trade at up to $100\times$ leverage, so an adverse move triggers an exchange's risk engine to close positions with price-insensitive market orders, and because each such order consumes book liquidity and pushes the price further, the same way it can set off more liquidations, the digital-asset form of a funding-liquidity spiral (Brunnermeier & Pedersen, 2009) in which deleveraging is clustered, self-exciting (Aït-Sahalia, Cacho-Diaz, & Laeven, 2015), and one-sided. Every venue publishes these liquidations tick-by-tick, so the forced flow that is invisible in most markets is here a measured time series. Our idea is simple: the VRP prices *how much* variance risk is compensated, while liquidations reveal *when* that risk is converted into realized variance, so a forecaster that joins the

premium with its liquidation trigger should improve on either. We test this by combining a linear heterogeneous autoregression (HAR) with a deep network conditioned on the premium and the liquidation record.

This runs into a data constraint. The natural implied variance series for crypto is Deribit's DVOL index, but DVOL has been published only since March 2021. Any premium that uses it, and any flexible model that must learn the map from the premium and liquidations to variance, therefore has at most a year of history before the out-of-sample period begins, too little to span more than one volatility regime, and a premium-augmented HAR estimated on so short a sample carries a near-zero premium coefficient. The same constraint is sharper for any newly optioned coin: with no index of its own, a modeller must simply wait years for one to accumulate before the question can be posed at all. We address this directly.

The paper's first contribution is methodological: we *reconstruct* a long daily implied variance history. Although DVOL itself starts in 2021, Deribit's option liquidations reach back to 2018 for BTC and 2019 for ETH, and each option-liquidation print carries the contract's implied volatility, strike, expiry, and the underlying index. From these sparse, stress-clustered prints, we build a backward-looking, kernel-weighted daily at-the-money 30-day implied volatility proxy and calibrate it to DVOL over their overlap. The proxy recovers DVOL at a correlation of 0.94 (BTC) and 0.96 (ETH) and a root mean square error of 6.3/6.7 volatility points, with an error flat across calm and stressed days, and it extends the usable history back across the March 2020 cascade, a second stress regime, so that no result rests on the 2021 episode alone. The only input is a free, publicly broadcast byproduct, the liquidation tape every venue already streams, rather than the costly or proprietary historical option-surface snapshots a model-free index normally demands, so the construction is simple, cheap, and reproducible by anyone with the public feed. The proxy is noisier than the exchange's full-surface index, but its level and dynamics are recovered with enough precision to train on, which makes the recipe useful well beyond our setting: for any newly listed coin that has options but no index yet, for assets whose implied volatility history is too short, and even for delisted markets whose surface can no longer be sampled but whose liquidation record survives. We confirm the transfer on SOL and XRP, two coins with no published index, in Appendix C. We train on the reconstructed series and reserve the published DVOL for the test set and all trading, so the train/test boundary is the date of DVOL inception.

The second contribution is the forecaster and its economics. A two-stage multi-scale CNN–LSTM first predicts the timing of intraday liquidation bursts, predictable out of sample at an AUC of 0.78 (BTC) and 0.76 (ETH) with liquidation order flow as its dominant input, and then forecasts forward variance from the lagged VRP and the resulting daily liquidation risk score. The headline forecaster combines this deep model with HAR$(1, 7, 28)$ at a weight chosen, separately for each loss, on a held-out validation tail of the training sample, so the weight carries no look-ahead. Four results follow. (i) The combination is the only model **never** excluded from the 75% Model Confidence Set (Hansen, Lunde, & Nason, 2011) under squared-error loss at any of three horizons for both assets, and it improves significantly on each component (Diebold–Mariano $p \leq 0.001$ over the deep model). (ii) The liquidation signal's predictive content is genuinely *nonlinear*, which we show with a deliberately fair comparison. We give the *linear* HAR the same liquidation inputs as the deep model, the log liquidation notional, and the signed liquidation imbalance at $1/7/28$-day lags (the HAR-LIQ and HAR-VRP-LIQ models): they do not help and at longer horizons make the forecast *worse*, raising in-sample fit while degrading out-of-sample accuracy, since the individual liquidation coefficients are insignificant and the mapping is one

that a linear model cannot represent. A deep network given those identical inputs, by contrast, carries a strongly significant encompassing increment over the liquidation-aware HAR, at every horizon for BTC ($t = 6.0$–6.5) and for ETH at all three horizons ($t = 1.8$–2.8). The deep model's contribution is thus its nonlinear use of a liquidation signal both models can see, with an asset emphasis we trace to the market structure. (iii) Gating a replicated variance swap on liquidation or stress risk converts the statistical edge into cost-robust risk-adjusted returns whose horizon is asset-specific: the gated combination earns a net Sharpe near two for ETH at the weekly horizon (1.7–2.0 depending on the gate), while for BTC the gated overlay is strongest at the monthly horizon (around 1.1–1.8), on small non-overlapping samples that we read as economic corroboration of the forecast results. (iv) BTC and ETH differ systematically, which we trace to their leverage and liquidation intensity. We score forecasts by the MSE of log variance and the Jensen-corrected QLIKE loss of Patton (2011).

The remainder of the paper reviews the related literature (Section 2); assembles the Deribit market and liquidation mechanics, the implied variance reconstruction and its validation, and the data and descriptive statistics (Section 3); sets out the methodology (Section 4); presents the results across all three horizons (Section 5) and an extended robustness analysis (Section 6); discusses (Section 7); and concludes (Section 8).

2 Related literature

This study draws on two strands of work: the variance risk premium and the forecasting of realized variance, and the microstructure of forced liquidations in leveraged crypto markets. We review each in turn and locate our contribution at their intersection.

2.1 The variance risk premium and realized variance forecasting

Realized variance, the sum of squared high-frequency returns, is a consistent ex-post variance estimator (Andersen, Bollerslev, Diebold, & Labys, 2003) that is highly persistent and long-memoried; the HAR of Corsi (2009) captures this parsimoniously with daily, weekly, and monthly components and remains a benchmark that more elaborate models struggle to beat (Bollerslev, Hood, Huss, & Pedersen, 2018; Bollerslev, Patton, & Quaedvlieg, 2016), including in cryptocurrency markets, where it forecasts volatility well and typically dominates GARCH specifications (Bergsli, Lind, Molnár, & Polasik, 2022), while the same long-memory feature carries into Bitcoin option valuation (Siu, 2025). Option-implied variance carries the market's forward, risk-neutral assessment and forecasts realized volatility in cryptocurrency markets (Hoang & Baur, 2020); the wedge between it and realized variance, the variance risk premium, forecasts future variance and the equity premium (Bekaert & Hoerova, 2014; Bollerslev et al., 2009; Carr & Wu, 2009). We take the premium as the economic foundation of crypto variance forecasting and ask what an observable microstructure trigger adds to it. A related strand stresses that the maturity at which the premium is measured shapes its predictive power (Plíhal & Mampouya, 2025), and that volatility forecasts are sensitive to the prevailing regime and to external uncertainty shocks (Plíhal, Mampouya, Deev, & Linnertová, 2025); our reconstructed index is anchored at a single 30-day tenor, a restriction we return to where it bears on the results.

On the modelling side, neural networks can represent nonlinear, threshold dynamics that a linear HAR cannot, and machine-learning methods now feature prominently in asset pricing (Gu, Kelly, &

Xiu, 2020) and in learning directly from order flow (Sirignano & Cont, 2019); recurrent (LSTM) architectures (Hochreiter & Schmidhuber, 1997) are natural for sequential volatility data, and a multi-scale convolutional design fits the heterogeneous, multi-horizon structure of volatility. Neural networks can improve realized-variance forecasts (Bucci, 2020), but flexible models pay for their capacity in estimation variance, and across the broader literature the out-of-sample evidence against the HAR is mixed. This is the setting in which forecast combination helps: when two predictors have imperfectly correlated errors, averaging reduces loss, and a parsimonious weighting typically outperforms a fully estimated optimal-weight vector whose estimation error swamps the diversification gain, the forecast-combination puzzle (Bates & Granger, 1969; Granger & Ramanathan, 1984; Timmermann, 2006). We exploit this by averaging a low-variance linear model with a higher-variance nonlinear one at a single weight selected on held-out training data, the minimal disciplined hybrid, and we quantify each component's contribution with Chong–Hendry encompassing regressions (Chong & Hendry, 1986). Relative to the debate between the HAR and machine learning (Bollerslev et al., 2018; Gu et al., 2020), our reading is that the premium's predictive content is nonlinear and that the linear cascade and the nonlinear liquidation response are complementary, so the value lies in the conditioning information and the hybrid rather than in the modelling machinery.

2.2 Liquidations, liquidity spirals, and the crypto setting

Order flow moves prices (Cont, Kukanov, & Stoikov, 2014; Kyle, 1985), and forced order flow does so most violently: when levered positions are liquidated, the resulting price-insensitive selling depresses prices and triggers further margin calls, a funding-liquidity spiral (Brunnermeier & Pedersen, 2009). Such deleveraging is clustered and self-exciting, well described by mutually exciting (Hawkes-type) processes (Aït-Sahalia et al., 2015), and one-sided because leverage is asymmetric; this one-sidedness is the order-flow counterpart of the signed-jump, good/bad volatility asymmetry in returns (Patton & Sheppard, 2015). Order-flow imbalance and aggressive trading already forecast price jumps in the Bitcoin market (Scaillet, Treccani, & Trevisan, 2020), and we study the variance-side analogue. The difference from most markets is observability: forced selling elsewhere must be inferred from prices and volumes, whereas in crypto, every venue publishes it tick-by-tick. Cryptocurrency derivatives are dominated by perpetual swaps at extreme leverage, and liquidation cascades, episodes in which hundreds of millions of dollars of positions are force-closed within hours, are a recurring, first-order feature of the market, one that equilibrium models of Bitcoin option prices now incorporate through correlated jumps and liquidity risk (Li, 2026). Deribit is the venue for crypto options and publishes the DVOL implied volatility index (Alexander & Imeraj, 2021); variance swaps, however, are not listed, so a variance view must be expressed through an option replication (Demeterfi et al., 1999). We bring the published liquidation record and the option-implied premium together, contributing to this literature the observation that an exchange's published forced-flow stream is a tradable, forecastable state variable for variance, and we address the short-history problem that has constrained the literature by reconstructing the implied variance series itself, the first use of liquidation-borne implied volatilities to extend a crypto volatility index.

3 Data, the Deribit market, and the implied variance reconstruction

Our analysis draws on tick-level liquidation and option records and the DVOL implied volatility index from Deribit, the dominant venue for cryptocurrency options and a major venue for BTC and ETH perpetual swaps, together with a 5-minute price feed. Two features of this market shape everything that follows. The exchange publishes its forced liquidations tick-by-tick, so the destabilizing forced selling flow that is invisible in most markets is here a measured time series, the signal at the heart of the paper. Deribit also publishes DVOL, its 30-day model-free implied volatility index, only from March 2021, leaving any premium or liquidation-conditioned model barely a year of history before the out-of-sample period begins. We turn the first feature into the paper's main predictor and address the second by reconstructing the implied variance series backward from the option-liquidation prints that predate the index.

3.1 Forced liquidations and the DVOL index

Deribit is the dominant venue for cryptocurrency options and a major venue for perpetual swaps on BTC and ETH. A perpetual swap is a futures contract without expiry, kept near the index by a periodic funding payment, and is offered at leverage up to $100\times$; positions are margined either in the coin (inverse) or in a stablecoin (USDC). When a position's margin falls below maintenance, the exchange's risk engine liquidates it with a market order and prints a flagged liquidation trade carrying the timestamp, side, size, mark and index prices, and, for option liquidations, the contract's implied volatility, strike, and expiry. A long position liquidation is forced *selling*, a short position liquidation forced *buying*; in a structurally long, levered market, the forced-selling direction is the destabilizing one, and the clustering of these prints within minutes is the measurable footprint of the Brunnermeier and Pedersen (2009) spiral.

The two margining conventions differ in a way that matters for cascades. Figure 1 plots the payoff per unit of posted margin against the price move: a linear, stablecoin-margined contract pays off linearly in price, whereas an inverse, coin-margined contract is *convex*, because its margin is denominated in the same coin whose price is moving. A long inverse position, therefore, loses value faster as the price falls and reaches its maintenance margin at a smaller drop than the linear equivalent, which steepens the downside and feeds the forced-selling spiral; the mirror-image convexity exposes inverse shorts on the upside.

We merge the inverse and USDC perpetual liquidation books for BTC and ETH, restricting to perpetual contracts, which dominate flagged liquidations and are the most actively traded leveraged instruments. The DVOL index is Deribit's 30-day, constant-maturity, model-free implied volatility, constructed from the option surface in the manner of the VIX; its square DVOL^2 is the natural strike for a 30-day variance swap and the implied leg of the variance risk premium. As noted, DVOL is published only from 2021-03-24; the liquidation books, including their option prints, reach back to 2018-03 (BTC) and 2019-03 (ETH), and we exploit this asymmetry to reconstruct the index.

3.2 Reconstructing a long implied variance history

The deep model must learn a nonlinear map from the premium and liquidations to forward variance, and the premium-augmented HAR must estimate a premium coefficient; both require a history in which the implied variance series exists. With DVOL beginning in 2021 and the out-of-sample period beginning

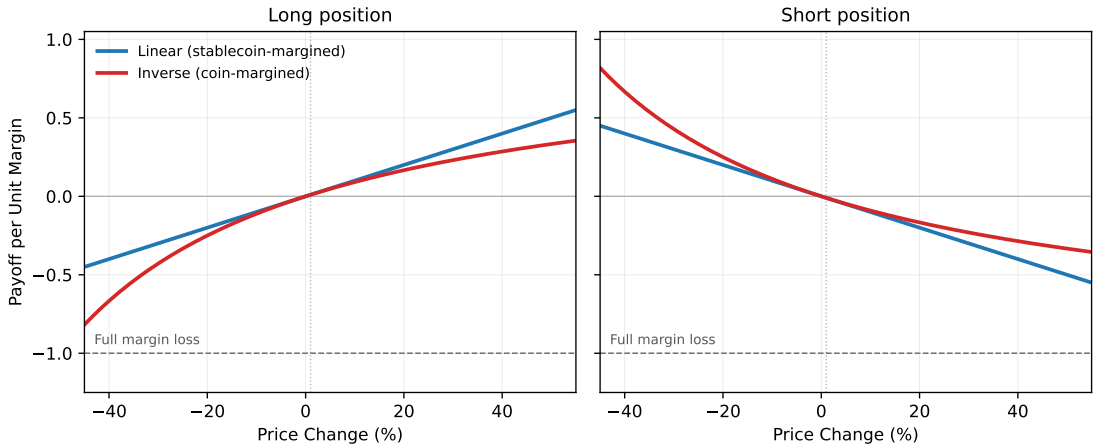


Figure 1: Payoff per unit of posted margin against the price change, for linear (stablecoin-margined) and inverse (coin-margined) perpetual contracts, at unit leverage and entry price normalized to one. *Left:* a long position; *right:* a short position. The inverse contract is convex because its margin is denominated in the coin whose price is moving, so a long loses faster as the price falls (and a short as it rises), reaching full margin loss at a smaller move than the linear contract.

at the same date, an unaided design gives the network roughly one year of training and forces the HAR-VRP coefficient toward zero. It also leaves the economic case resting on the 2021 leverage unwind. We therefore reconstruct the implied variance series backward from data that exist before 2021.

Only option liquidations carry an implied volatility, and they are sparse and clustered in stress. We turn this into a daily series in three steps. (i) Within each day, we collapse repeated prints of the same option contract to their mean implied volatility, mark-price moneyness, and time to expiry, which removes the over-weighting of a single cascaded contract. (ii) Across that day's distinct contracts, we form a kernel-weighted average that pulls toward the at-the-money, 30-day point that DVOL prices,

$$\widehat{\mathrm{IV}}_t = \frac{\sum_j w_j \,\mathrm{IV}_j}{\sum_j w_j}, \qquad w_j = \exp\!\Big(-\frac{m_j^2}{2h_m^2}\Big)\exp\!\Big(-\frac{(\tau_j-30)^2}{2h_\tau^2}\Big), \tag{1}$$

where $m_j = \log(K_j/S_t)$ is log-moneyness, τ_j is the contract's maturity in days, and the bandwidths, $h_m = 0.10$ and $h_\tau = 25$, set how fast the weight decays away from the at-the-money, 30-day point that DVOL prices. They are chosen to admit a usable cross-section of the thin pre-2021 days while keeping the average near that point: $h_m = 0.10$ in log-moneyness gives near-full weight to strikes within roughly $\pm 10\%$ of spot, where implied volatility is closest to the index level and least distorted by the smile, and $h_\tau = 25$ days lets contracts from about one week to two months inform the 30-day estimate, the maturity range is routinely populated even on sparse days. Contracts with $|m_j| > 0.40$ or $\tau_j \notin [3, 90]$ are discarded: deep out-of-the-money strikes are priced largely off the volatility skew and proxy the at-the-money level poorly, while options shorter than three days are dominated by expiry microstructure, and those beyond ninety days sit too far from the 30-day target to be informative. (iii) We forward-fill the occasional empty day onto a complete daily grid and smooth idiosyncratic sparsity with a short *backward* exponential moving average (span nine days), then map the proxy to the index with an affine calibration $\mathrm{DVOL}_t \approx a + b\,\widehat{\mathrm{IV}}_t$ estimated by ordinary least squares on the 2021-onward overlap where both exist. The affine coefficients are *fixed* (not re-estimated walk-forward), so the only cross-sample information they carry is a constant level and scale, never the prediction target: the calibration

is a level alignment that uses no information about future realized variance. Pre-2021 it is therefore an out-of-sample extrapolation, and the Stage-2 network in any case standardizes each input window, which absorbs a residual level error.

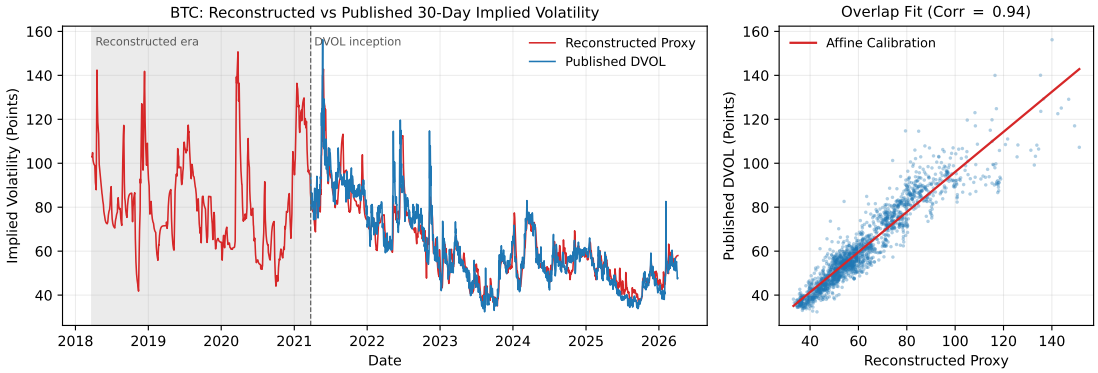
Table 1 reports the replication quality over the overlap, and Figure 2 plots the reconstructed and published series. The proxy tracks the published index closely: correlation 0.94 (BTC) and 0.96 (ETH), affine-calibrated R^2 of 0.89 and 0.91, and a root-mean-square error of 6.3 and 6.7 volatility points against indices that average in the sixties, on a median of only two to three option contracts per day. Two properties make the series usable as a predictor. The mean absolute percentage error is essentially flat across calm and stressed terciles (it does not deteriorate when it matters, because liquidation prints are, if anything, denser in stress), and it is stable when the calibration is fit on an early sub-sample and tested on a later one. The reconstruction reaches back to 2018-03 (BTC), and 2019-04 (ETH) and so spans the March-2020 COVID cascade, a stress regime entirely outside the 2021 episode. We attempted richer variants, projecting the full smile by skew and term structure and a validation-selected estimator grid, and neither improved on Eq. (1): with two to three contracts a day, the extra parameters add noise. The residual error is structural, set by print sparsity, and would fall only with a dense historical option chain; for our purpose, which is to give the models a long, regime-rich training history while reserving the *real* index for all evaluation and trading, the proxy is adequate. Two limitations bound the construction. First, the replication can be validated only on the 2021-onward overlap, yet the proxy feeds the models *before* 2021, where no ground truth exists; we therefore rest pre-inception fidelity on the two indirect checks above, the error being flat across calm and stressed terciles and stable under a time split calibration, and we confirm in Section 6 that the headline does not hinge on the reconstructed era, since a walk-forward scheme that excludes the 2020 COVID window from training reproduces the combination's standing (Table 6). Second, assembled from only two to three sparse, stress-clustered prints a day, the proxy is a noisier estimator than the exchange's full-surface index, carrying a visibly higher day-to-day variance than the smooth published DVOL, so the ≈ 6 volatility-point RMSE is largely this excess dispersion rather than bias; this is the price of a thin historical cross-section, tolerable here because the proxy feeds the *training* premium only, where the network standardizes each input window and absorbs much of the extra noise. Evaluation and trading use the real published DVOL throughout; the reconstructed proxy is never used there.

Table 1: Reconstruction of the 30-day implied variance index from option-liquidation implied volatilities.

Asset	Reconstructed from	Overlap days	Median opts/day	Corr.	Calib. R^2	RMSE (vol pts)
BTC	2018-03	1,840	2–3	0.941	0.886	6.26
ETH	2019-04	1,840	2–3	0.956	0.914	6.69

Notes. The daily at-the-money 30-day implied volatility proxy of Eq. (1), affine-calibrated to the published DVOL by OLS on the overlap and compared to it. "Corr." is the correlation of the raw proxy with DVOL; "Calib. R^2" and RMSE are after the affine calibration $\mathrm{DVOL} \approx a + b\,\widehat{\mathrm{IV}}$ (BTC $a = 4.88, b = 0.91$; ETH $a = 1.03, b = 0.99$). Volatility points are annualized volatility units (DVOL averages in the sixties). The mean absolute percentage error is ≈ 6–8% and is flat across calm/stress terciles and stable under a time-split of the calibration. The proxy feeds the variance risk premium in *training only*; the real published DVOL is used for the test set and all trading.

Figure 2: Reconstructed versus published 30-day implied volatility.



Notes. Left: the calibrated proxy (from option-liquidation implied volatilities) and the published DVOL over the overlap, with the pre-2021 reconstruction, including the March 2020 COVID cascade, to its left. *Right:* proxy versus DVOL scatter over the overlap. The proxy enters the premium in training only; the real index is used for all evaluation and trading.

3.3 Data and descriptive statistics

We combine prices from a commercial vendor with two series taken directly from the exchange. Five-minute OHLCV prices for BTC and ETH are from FirstRateData, a commercial market-data provider; the daily DVOL implied volatility index and the tick-level liquidation books, with full per-event attributes, come from the public Deribit API. We use prices from 2018-03 (BTC) and 2019-03 (ETH) onward, the span over which the liquidation record and the reconstructed premium are available; the DVOL index itself begins 2021-03-24 and is extended backward by the reconstruction of Section 3.2. Bars are reindexed to a complete 5-minute grid, with absent bars carrying a zero return. Realized variance is the sum of squared 5-minute log returns within day t, annualized (Andersen et al., 2003); it is the forecast *target* and, at lags $1/7/28$, the HAR's regressors. The Parkinson (Parkinson, 1980) range estimator, $\mathrm{PV}_t = (4\ln 2)^{-1}\,[\ln(H_t/L_t)]^2$ from the day's high H_t and low L_t and annualized, is a separate, noise-robust volatility *input* to the deep model, never the target; it complements RV by reading the intraday range rather than the return path. For the H-day horizon, the target is the average forward log variance,

$$\mathrm{RV}_t = \textstyle\sum_i r_{t,i}^2, \qquad y_t^{(H)} = \log\Big(\frac{1}{H}\sum_{j=1}^{H}\mathrm{RV}_{t+j}\Big), \tag{2}$$

with $r_{t,i}$ the i-th 5-minute log return on day t. Following Bollerslev et al. (2009) we use the lagged variance risk premium, measured as the *log* variance ratio so that it is always defined, stationary, and on the scale of the other (log) inputs,

$$\mathrm{VRP}_t = \log \mathrm{DVOL}_t^2 - \log \mathrm{RV30}_t, \qquad \mathrm{RV30}_t = \tfrac{1}{30}\textstyle\sum_{j=0}^{29}\mathrm{RV}_{t-j}, \tag{3}$$

where $\mathrm{RV30}_t$ is the *trailing* 30-day realized variance, matched to DVOL's constant 30-day tenor, so VRP_t is known at t and predicts $t{+}1,\ldots,t{+}H$.

The liquidation record enters through two daily variables, both fully observed at the close of day t. The log liquidation notional, $\mathrm{LN}_t = \log\big(1+\sum_k V_{k,t}\big)$, aggregates the US-dollar size $V_{k,t}$ of every forced trade k on day t and measures the *intensity* of forced flow. The liquidation imbalance, $\mathrm{IMB}_t = (S_t - B_t)/(S_t + B_t) \in [-1,1]$, measures its *direction*: S_t and B_t are the day's forced sells (long liquidations) and forced buys (short liquidations), so a positive value denotes net forced selling, the

destabilizing side. We compute IMB_t both notional-weighted (the headline, with S_t, B_t in dollars) and count-weighted (with S_t, B_t event counts; Appendix D). The deep model additionally reads the Parkinson range and the same-day liquidation notional of the *other* asset (the cross-asset input, ETH for BTC and BTC for ETH, since the two deleverage together), alongside the intraday order-flow detail set out in Section 4.2.

All forecasting and trading are evaluated on the *common window* 2021-03-24 to 2026-01-26 (1,770 days), the published DVOL sample on which every variable is observed. Models are estimated by an expanding walk-forward, refit quarterly on all history available before each forecast date, which reaches back to 2018/2019 through the reconstructed premium; rolling-window and COVID-excluded schemes are examined in Section 6.

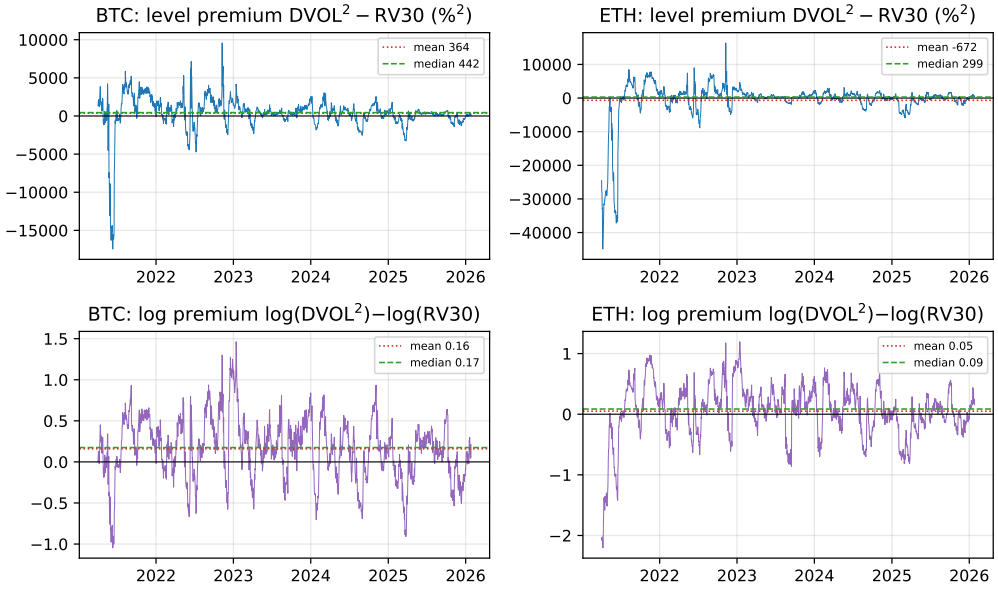
Table 2 reports descriptive statistics in the log form in which the variables enter the models. Realized variance is highly persistent; its autocorrelation is still substantial at the monthly lag ($\rho(28) = 0.45$ BTC, 0.44 ETH), motivating a long-memory specification. Every variable in Table 2 has a positive mean. The log premium in particular is mildly skewed and positive on average for both assets (0.16 BTC, 0.04 ETH); it nonetheless turns negative on a third to two-fifths of individual *days*, when realized variance exceeds implied, and those liquidation-driven spikes are what ex-ante implied variance underprices, precisely the nonlinearity the deep model targets. We use this log ratio rather than the raw *level* difference $\mathrm{DVOL}^2 - \mathrm{RV30}$ (which does not enter the models and is not tabulated) precisely because the level difference is dominated by a heavy left tail of such spikes (on ETH's worst day, realized variance exceeded implied more than tenfold): its arithmetic mean is dragged below zero, even though its median is positive, whereas the log ratio compresses the tail and enters the models with the well-behaved positive mean reported in the table (Fig. 3). Liquidation activity does *not* decay over the week, its lag-7 autocorrelation matching or exceeding its lag-1, consistent with multi-day deleveraging.

Table 2: Descriptive statistics of the model variables over the common evaluation window (2021-03-24 to 2026-01-26, 1,770 days).

Asset	Variable	Mean	SD	Skew	Kurt	$\rho(1)$	$\rho(7)$	$\rho(28)$
BTC	$\log \mathrm{RV}$	7.69	0.99	0.14	0.80	0.67	0.59	0.45
	$\log \mathrm{PV}$	7.50	1.22	−0.04	0.32	0.44	0.48	0.39
	VRP	0.16	0.37	−0.17	0.40	0.96	0.75	0.10
	$\log \mathrm{DVOL}^2$	8.16	0.58	0.24	−0.63	0.99	0.93	0.81
	log liq. notional	22.42	4.34	−3.38	15.00	0.22	0.24	0.22
ETH	$\log \mathrm{RV}$	8.20	1.01	0.26	1.03	0.71	0.58	0.44
	$\log \mathrm{PV}$	8.09	1.19	0.03	0.56	0.49	0.44	0.36
	VRP	0.04	0.48	−1.33	4.07	0.97	0.79	0.24
	$\log \mathrm{DVOL}^2$	8.57	0.60	−0.11	0.34	0.99	0.94	0.82
	log liq. notional	18.69	4.50	−2.60	8.21	0.19	0.20	0.17

Notes. Statistics over the common daily window (the published DVOL sample), which is also the out-of-sample evaluation period. Rows are the variables that enter the models: $\log \mathrm{RV}$ is the log realized variance (the HAR target and, at lags $1/7/28$, its regressors); $\log \mathrm{PV}$ the log Parkinson range variance (Parkinson, 1980); VRP the log variance risk premium of Eq. (3); $\log \mathrm{DVOL}^2$ the log squared implied volatility index (the swap strike); and log liquidation notional, the daily liquidation channel. $\rho(\ell)$ is the sample autocorrelation at lag ℓ days. The full autocorrelation evidence motivating the $(1, 7, 28)$ lags is in Appendix B.

Figure 3: The variance risk premium, level versus log.



Notes. Top: the level difference $\mathrm{DVOL}_t^2 - \mathrm{RV30}_t$, with arithmetic mean (dotted) and median (dashed); ETH's mean is dragged below zero by a heavy left tail of realized variance spikes. *Bottom:* the log premium we use is well-behaved and positive on average for both assets. Negative log-premium days (31–40% of the sample) are liquidation-driven episodes in which realized variance exceeds implied.

4 Methodology

We forecast the H-day forward log variance of Eq. (2) at horizons $H \in \{7, 14, 30\}$ days. This section sets out the ladder of forecasters we compare, the two-stage deep model that carries the liquidation signal, the combination that joins the linear and nonlinear views, the losses by which we score them, and the replicated variance-swap strategy that trades the resulting forecast. Every model is estimated walk-forward, on the history available before each forecast date.

4.1 The model ladder

We compare a ladder of forecasters of the H-day forward log variance (2). The linear benchmark is the HAR, which regresses $\log \mathrm{RV}$ on its trailing daily, weekly, and monthly averages,

$$\hat{y}_t^{\mathrm{HAR}} = \beta_0 + \beta_1 \overline{\mathrm{RV}}_t^{1} + \beta_7 \overline{\mathrm{RV}}_t^{7} + \beta_{28} \overline{\mathrm{RV}}_t^{28}, \qquad \overline{\mathrm{RV}}_t^{\ell} = \tfrac{1}{\ell} \textstyle\sum_{i=0}^{\ell-1} \log \mathrm{RV}_{t-i}, \tag{4}$$

with $\overline{\mathrm{RV}}_t^{\ell}$ the trailing ℓ-day average log variance. Equity practice uses $1/5/22$ *trading* days; because crypto trades $24/7$ we use the calendar lags $1/7/28$. The autocorrelation evidence supports this: realized variance and liquidation persistence are higher at lags 7 and 28 than at 5 and 22 in every case (Appendix B), and a Diebold–Mariano test (Diebold & Mariano, 1995) finds $\mathrm{HAR}(1, 7, 28)$ significantly more accurate than $\mathrm{HAR}(1, 5, 22)$ (HAC $p < 0.002$ for ETH at every horizon and for BTC at the monthly horizon). The remaining linear models add regressors to (4). HAR-VRP adds the lagged premium, $+\gamma\, \mathrm{VRP}_t$; because the reconstruction gives the premium a multi-year pre-sample, its coefficient is genuinely estimated rather than mechanically zero. To make the comparison with the deep model *fair*, HAR-LIQ and HAR-VRP-LIQ give the linear HAR the liquidation channel directly, adding the log

liquidation notional LN and the signed imbalance IMB of Section 3.3 at the same trailing $1/7/28$-day lags,

$$+\sum_{\ell\in\{1,7,28\}}\left(\phi_\ell\,\overline{\mathrm{LN}}_t^{\ell}+\psi_\ell\,\overline{\mathrm{IMB}}_t^{\ell}\right), \tag{5}$$

so the linear model and the deep network see the identical liquidation information set; IMB is notional-weighted in the headline and count-weighted in the variant of Appendix D. We benchmark GARCH(1,1) (Bollerslev, 1986), $\sigma_t^2=\omega+\alpha\,\epsilon_{t-1}^2+\beta\,\sigma_{t-1}^2$, and EWMA/RiskMetrics (J.P. Morgan, 1996), $\sigma_t^2=(1-\lambda)\,\epsilon_{t-1}^2+\lambda\,\sigma_{t-1}^2$. All linear models are estimated by walk-forward OLS on log variance, for which the log transform stabilizes the level-proportional measurement error of realized variance and makes OLS efficient (Andersen et al., 2003; Bollerslev et al., 2016).

4.2 Deep model: two stages, architecture, and combination

The premium and its trigger operate at different frequencies, so the deep model is split into two stages that answer different questions and feed one another.

Stage 1 forecasts when *forced flow will arrive.* A liquidation *burst* is a 5-minute bar b whose liquidation notional exceeds the trailing 30-day rolling 99th percentile of bar notional, computed from bars up to and including b, a rare event at about 1% of bars. A multi-scale CNN–LSTM reads an L_1=36-bar window of 13 order-flow features[1] and outputs the burst probability

$$p_b=\sigma\Big(w^{\top}\,\mathrm{LSTM}\big[\,\mathrm{IN}\big(\|_{\kappa\in\{3,6,12\}}\mathrm{ReLU}(\mathrm{Conv}_\kappa X_b)\big)\big]+c\Big), \tag{6}$$

where $\|$ concatenates parallel convolutions of kernel widths κ, IN is instance normalization, and σ the logistic function. It is trained by class-weighted cross-entropy; its quality is the area under the ROC curve (AUC), a threshold-free, class-imbalance-robust measure suited to rare-event labels. The daily *liquidation risk score* aggregates the bar probabilities, $\lambda_t=n_t^{-1}\sum_{b\in t}p_b$.

Stage 2 forecasts how much *variance follows.* Because variance is persistent, a second CNN–LSTM (kernels $\{3,5,22\}$) predicts the deviation from a random-walk baseline over an L_2=22-day window Z_t,

$$\hat{y}_t=b_t+g_\theta(Z_t),\qquad b_t=\log\Big(\tfrac{1}{H}\textstyle\sum_{j=0}^{H-1}\mathrm{RV}_{t-j}\Big). \tag{7}$$

Deep-Base uses $Z_t=[\log\mathrm{PV}_t,\mathrm{VRP}_t]$, with PV_t the Parkinson range variance, a noise-robust volatility proxy complementing RV, and VRP_t the lagged premium; *Deep-Liq* additionally ingests λ_t and daily liquidation channels (taker share, mark–index basis, and cross-asset liquidation notional). The two stages are complementary: Stage 1 converts irregular intraday order flow into a single daily state variable λ_t that a daily variance model could not extract from prices alone, and Stage 2 maps that state, with the premium, into the multi-week variance, the swap pays on. Using both is what links the microstructure trigger to the traded horizon.

Turning to the architecture and the combination, the design is dictated by the data rather than chosen for capacity. Multi-scale convolutions extract short-range and long-range patterns without committing to one lag, mirroring the HAR's heterogeneous components; the LSTM (Hochreiter & Schmidhuber, 1997) summarizes the sequence into a fixed state; instance normalization stabilizes training across calm

[1]The thirteen per-bar features are the 5-minute log return and Parkinson range; traded volume; the liquidation count, notional, imbalance, and slippage; the taker and perpetual shares of liquidation flow; a signed tick-pressure measure; the mark–index basis; a liquidation-clustering indicator; and the cross-asset liquidation notional (the other coin's bar notional).

and stressed scale shifts, and the residual parameterization in (7) embeds the random-walk benchmark, so the network learns only the predictable deviation, which regularizes the problem and makes the deep and linear forecasts directly comparable. The two-stage architecture is sketched in Figure 10, and the hyperparameters, fixed ex ante, are in Appendix A. Each deep model is trained from seven independent random initializations, and we report the average of their forecasts throughout.[2] Our headline forecaster is a convex combination of the HAR and the deep model in log space,

$$\hat{y}_t^{\text{Combo}} = w^\star \, \hat{y}_t^{\text{HAR}} + (1 - w^\star) \, \hat{y}_t^{\text{Deep-Liq}}, \tag{8}$$

whose weight $w^\star$ is chosen from the data rather than imposed. We sweep $w \in \{0.01, 0.02, \ldots, 0.99\}$ and take the w that minimizes each loss, evaluated only on a held-out validation tail of the training history (the final 365 days before the out-of-sample boundary), so the weight is fixed before any test date and each forecast entering the selection comes from a model trained only on data preceding it. Because the squared error and QLIKE objectives need not agree, we select a separate weight for each: $w^\star_{\text{MSE}}$ governs the squared error evaluation and $w^\star_{\text{QLIKE}}$ the QLIKE evaluation and the trade rule (Section 5.2 and Fig. 5). The selected weights land near, but not at, the even split (HAR weight 0.53–0.73 for BTC, falling to 0.09–0.63 for ETH as the horizon lengthens and the deep model earns more weight), confirming the economic rationale for combining: the HAR is a low-variance estimator of the linear volatility cascade, while the deep model captures the convex, threshold response of variance to forced deleveraging, negligible in calm states and explosive in stressed ones (Aït-Sahalia et al., 2015), so their errors are imperfectly correlated and a near-even weighting cancels offsetting mistakes (Section 5). Holding the weight fixed at $\frac{1}{2}$ changes the headline negligibly (the selection curves are flat near their optima), so the result does not hinge on the weight, only on the act of combining.

4.3 Forecast losses and evaluation

Every model forecasts the conditional mean of *log* variance, $\hat{y}_t = \widehat{\mathbb{E}}[\log \text{RV}_t^{(H)} \mid \mathcal{F}_t]$, but the QLIKE loss and the variance-swap payoff live in variance *levels*, where the forecast is not $\exp(\hat{y}_t)$. Under conditional log-normality, with $\hat{\sigma}_t^2$ the variance of the log variance forecast error,

$$\widehat{\mathbb{E}}[\text{RV}_t^{(H)} \mid \mathcal{F}_t] = \exp\big(\hat{y}_t + \tfrac{1}{2}\hat{\sigma}_t^2\big), \tag{9}$$

the Jensen (log-normal) correction, which we apply wherever a level forecast is needed: for QLIKE, $\hat{\sigma}_t^2$ is each model's out-of-sample log variance forecast error variance; for the trade rule, a backward-looking expanding estimate from past errors only.

We score forecasts with two out-of-sample losses. The first is the MSE of the log variance forecast, computed in log space and so invariant to the Jensen correction. The second is the variance-proxy-robust QLIKE of Patton (2011),

$$\text{QLIKE}_t = \frac{\text{RV}_t^{(H)}}{\hat{v}_t} - \log \frac{\text{RV}_t^{(H)}}{\hat{v}_t} - 1, \qquad \hat{v}_t = \exp\big(\hat{y}_t + \tfrac{1}{2}\hat{\sigma}_t^2\big), \tag{10}$$

[2] The seven seeds are independent random initializations of the network weights and training order; averaging their forecasts removes the run-to-run noise of stochastic optimization, so the reported numbers do not rest on a single lucky or unlucky training run.

which penalizes under-prediction of variance more heavily than over-prediction, the costly error for a risk manager or short-volatility desk, and uses the Jensen-corrected level $\hat{v}_t$. Both losses are non-negative and smaller is better, and their occasional disagreement is itself diagnostic (Section 5). Forecasts are compared pairwise with the Diebold–Mariano statistic (Diebold & Mariano, 1995) using the Harvey, Leybourne, and Newbold (1997) small-sample correction and a HAC variance truncated at the horizon, and jointly with the Model Confidence Set (Hansen et al., 2011) under both losses (size 0.25, stationary bootstrap). The combination is decomposed by Bates–Granger inverse-MSE weights (Bates & Granger, 1969) and Chong–Hendry encompassing regressions (Chong & Hendry, 1986) with Newey–West (Newey & West, 1987) standard errors.

4.4 Replicated variance-swap strategy

A variance swap is not listed in crypto, so we trade its Demeterfi et al. (1999) replication: a static strip of options plus delta hedging whose payoff per unit of capital is the finite-strip-bounded variance surprise. Because DVOL^2 is a model-free (VIX-style) squared-volatility index, it understates the fair variance-swap strike by a convexity term, so the measured premium and the short-variance returns reported below are, if anything, conservative. With strike $K_t^2 = \mathrm{DVOL}_t^2$, the Jensen-corrected mean-variance forecast $\hat{v}_t$ of (9), and $\mathrm{RV}_t^{(H)}$ the realized variance over the H-day hold, the trade direction and net return are

$$d_t = \operatorname{sign}\big(\hat{v}_t - K_t^2\big), \qquad r_t = \phi\, w_t\, d_t\, \frac{\mathrm{RV}_t^{(H)} - K_t^2}{K_t^2} - c_t, \tag{11}$$

where $w_t = \min(\sigma_{\mathrm{tgt}}/\hat{\sigma}_t,\, \bar{w})$ is a volatility-targeted, capped weight, $\hat{\sigma}_t = \sqrt{\hat{v}_t}$ the forecast volatility, $\sigma_{\mathrm{tgt}} = 0.60$ the annualized target and $\bar{w} = 1.5$ the cap; ϕ is the capital fraction at risk, and c_t the Deribit cost (Deribit, 2026): option fees, a 1.25 volatility-point replication spread, and collateral funding. The directional rule trades on the sign of forecast-minus-strike, so that, unlike a confidence band, *every* model trades on the same dates, and the comparison is not confounded by differing trade counts. On top of this, we apply a *gate*, a filter that decides on each date whether to take that day's trade at all; Figure 4 summarizes the decision. The *liquidation gate* takes the trade only when the day's liquidation risk score λ_t is in its top tercile; the *stress gate* only on days when liquidation notional exceeds its trailing 252-day 80th percentile. Both gates read only information available at t, and their role is to concentrate the strategy in the states where the liquidation signal is informative, leaving capital idle otherwise. Positions compound a \$10,000 account and are non-overlapping (held H days), which avoids the error autocorrelation overlapping horizons induce but caps the sample at 58 trades for H=30 (252 for H=7); inference is therefore drawn primarily from the daily forecast tests, with the gated returns read as economic illustrations. The deep network is costlier to train than the HAR, but that cost is paid once, at training time, and not at each forecast. The pipeline is, in any case, computationally light: although Stage 1 ingests high-frequency 5-minute order flow, the forecasting and trading horizons are long, one to four weeks, so the intraday convolutions are evaluated only to summarize each day into the single state variable λ_t, and the book is rebalanced once every H days. The network is trained once and then queried daily, and the strategy runs on commodity hardware across a multi-asset panel.

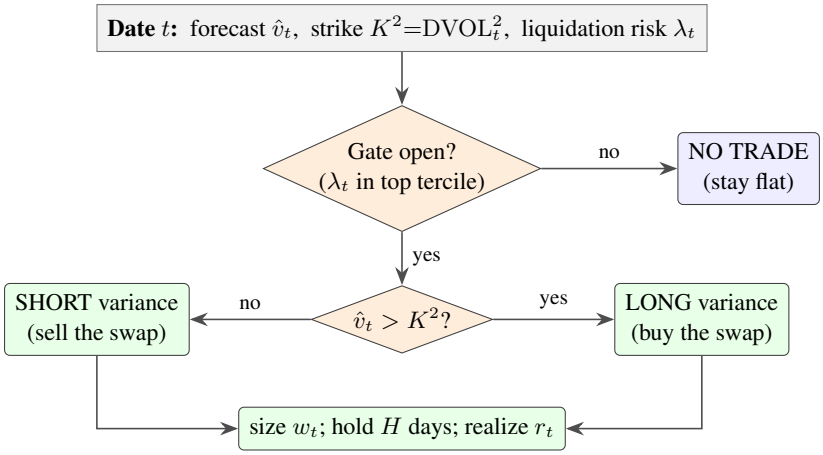


Figure 4: Directional trade rule of Eq. (11) with the liquidation gate. On each date the gate first decides whether to trade at all, taking the trade only when liquidation risk λ_t is in its top tercile; if it trades, the sign of the Jensen-corrected forecast $\hat{v}_t$ of Eq. (9) minus the strike K^2 sets the direction, and the position is sized by w_t, held H days, and realized. The stress gate replaces the gate condition with a trailing-percentile threshold on liquidation notional.

5 Empirical results

We present the forecasting results first, then decompose the combination, contrast BTC with ETH, and report the trading performance.

5.1 Liquidation timing and forecast accuracy

Stage 1 attains an out-of-sample AUC of $\mathbf{0.78}$ (BTC) and $\mathbf{0.76}$ (ETH) against a $\sim 1\%$ burst rate. Permutation importance is decisive: shuffling the liquidation-flow features lowers AUC by 0.11 (ETH) to 0.14 (BTC), roughly eight to nine times more than any other feature group (≤ 0.015), so λ_t is genuinely liquidation-driven. Much of this is self-excitation, liquidations begetting liquidations, the cascade mechanism the strategy intends to exploit.

Turning to the variance forecasts, Table 3 reports Model Confidence Set p-values at $H \in \{7, 14, 30\}$ under both losses; the underlying raw MSE and QLIKE levels are in Appendix B (Table 9), and their cumulative paths relative to the HAR over time are plotted in Figures 6 (squared error) and 7 (QLIKE). We report MCS p-values rather than losses because they convert each loss column into a direct, multiple-comparison-aware statement of which models are statistically indistinguishable from the best, marking the 75% set with † and the wider 90% set with ‡. The two losses, by design, need not agree, and the reading reflects that. Under **squared error** the combination is the **only model never excluded from the** 75% **Model Confidence Set at any horizon for either asset**: it is the *sole* member of the set for BTC at all three horizons ($p = 1.00$, every rival $p \leq 0.04$) and a comfortable member for ETH ($p = 0.69/0.51/0.76$). The asset asymmetry is itself informative. For **BTC** the combination is the sole member of the 75% squared-error set at every horizon and the best model under QLIKE as well ($p = 1.00$), sole at the monthly horizon while the premium-augmented HAR remains in the 75% QLIKE set at the shorter horizons. For **ETH** the parsimonious HAR is the single best model under squared error at every horizon ($p = 1.00$) and under QLIKE at the longer horizons, its long reconstructed training history making it a formidable linear baseline, and the combination is comfortably retained for ETH

under both losses (squared-error $p = 0.69/0.51/0.76$, QLIKE $0.32/0.60/0.71$). Two further patterns are visible. For ETH the liquidation-aware HAR-LIQ and HAR-VRP-LIQ sit in the 75% squared-error set alongside the plain HAR but never beat it, the first sign that the liquidation channel adds nothing a linear model can use; and the *standalone* deep models, Deep-Base and Deep-Liq, are excluded from the 75% set everywhere, their value realized only in combination, never alone. We read the combination's MCS membership cautiously: an average is, by construction, hard to exclude from a confidence set because averaging shrinks error variance, so membership is necessary but not sufficient evidence of a superior *forecaster*. The sharper, less mechanical evidence that the liquidation network forecasts better is the BTC encompassing increment reported next.

By a Diebold–Mariano test the combination significantly improves on the standalone deep model at every horizon for both assets ($p \leq 0.001$), and it dominates the HAR for BTC on the confidence set itself, the sole 75% member at every horizon while the HAR carries $p \leq 0.03$ (Table 3). The decisive evidence on the liquidation channel, however, comes from the *fair* comparison in which the linear HAR is given the same liquidation inputs as the deep model, the log notional and the signed imbalance at $1/7/28$ lags. Its out-of-sample loss then does not fall but *rises*: HAR-VRP-LIQ is worse than the plain HAR for BTC at every horizon and sharply so at the monthly one (MSE $0.36 \to 0.46$, Table 9), with no gain for ETH. The reason is overfitting. The liquidation regressors *raise* the in-sample fit (BTC monthly R^2 $0.59 \to 0.65$) while degrading the out-of-sample forecast, and their fitted loadings describe no sensible forward map: the imbalance terms are insignificant ($|t| \leq 0.8$) and the notional loads small and *wrong-signed* (a counterintuitive negative coefficient; Appendix B), so the in-sample gain is fitted noise rather than a stable linear liquidation signal. The deep network, given those identical inputs, tells the opposite story. In a Chong–Hendry encompassing regression the deep increment over the *liquidation-aware* HAR is strongly significant at every horizon for BTC ($\lambda = 0.42$–0.63, $t = 6.0$–6.5) and significant for ETH at all three horizons ($\lambda = 0.14$–0.38, $t = 1.8$–2.8), where the increment over the *plain* HAR was weak. So controlling for the linear liquidation channel *sharpens* the deep contribution rather than weakening it, and reveals it for ETH as well as BTC. In conclusion, the liquidation signal's predictive content is irreducibly nonlinear: a linear model cannot extract it, and is harmed by it, even under ridge, LASSO, or elastic-net penalization, whereas a deep network can. The result survives even when the linear model is handed the network's *full* daily feature set and their interactions: a regularized 134-regressor design shrinks the liquidation terms toward zero and still trails the network (Appendix E). The gain therefore reflects the network's nonlinear use of a signal both models can see. We headline the most accurate of three deep specifications differing only in regularization; Section 6 shows the combination's losses are essentially unchanged across all three, so the result reflects neither over-fitting nor the training protocol.

5.2 Decomposition and weight selection of the combination

Table 4 decomposes the combination at *all three* horizons. The HAR and Deep-Liq errors correlate $\rho \approx 0.77$–0.83 (BTC) and 0.83–0.88 (ETH), leaving room for each model to correct the other, and Bates–Granger inverse-MSE weights (Bates & Granger, 1969) are close to even (HAR weight 0.49–0.55), confirming that a near-even combination is well-founded; the loss-optimal selected weights of Section 5.2 refine this without overturning it. The encompassing regressions make the nonlinearity precise. The deep increment over the *plain* HAR is significant for BTC at every horizon ($\lambda = 0.34$–0.37, $t = 4.0$–5.5) but insignificant for ETH ($t \leq 1.3$), the asset asymmetry a single-model comparison

Table 3: Model Confidence Set p-values across horizons.

		Squared-error MCS p			QLIKE MCS p (Jensen)		
Asset	Model	H=7	H=14	H=30	H=7	H=14	H=30
BTC	HAR	0.00	0.02	0.01	$0.14^{\ddagger}$	$\mathbf{0.41}^{\dagger}$	0.01
	HAR-VRP	0.01	0.02	0.01	$\mathbf{0.31}^{\dagger}$	$\mathbf{0.41}^{\dagger}$	0.01
	HAR-LIQ	0.01	0.01	0.01	$0.14^{\ddagger}$	0.06	0.01
	HAR-VRP-LIQ	0.00	0.01	0.01	$0.12^{\ddagger}$	0.06	0.01
	GARCH	0.01	0.01	0.01	0.03	0.04	0.01
	EWMA	0.01	0.02	0.01	0.07	0.06	0.02
	Deep-Base	0.00	0.00	0.00	0.06	0.02	0.01
	Deep-Liq	0.01	0.01	0.04	$0.24^{\ddagger}$	$0.15^{\ddagger}$	$0.16^{\ddagger}$
	Combo	$\mathbf{1.00}^{\dagger}$	$\mathbf{1.00}^{\dagger}$	$\mathbf{1.00}^{\dagger}$	$\mathbf{1.00}^{\dagger}$	$\mathbf{1.00}^{\dagger}$	$\mathbf{1.00}^{\dagger}$
ETH	HAR	$\mathbf{1.00}^{\dagger}$	$\mathbf{1.00}^{\dagger}$	$\mathbf{1.00}^{\dagger}$	$\mathbf{0.73}^{\dagger}$	$\mathbf{1.00}^{\dagger}$	$\mathbf{1.00}^{\dagger}$
	HAR-VRP	$\mathbf{0.68}^{\dagger}$	$\mathbf{0.46}^{\dagger}$	$\mathbf{0.27}^{\dagger}$	$\mathbf{0.50}^{\dagger}$	$\mathbf{0.44}^{\dagger}$	$0.16^{\ddagger}$
	HAR-LIQ	$\mathbf{0.48}^{\dagger}$	$\mathbf{0.35}^{\dagger}$	$\mathbf{0.44}^{\dagger}$	$\mathbf{1.00}^{\dagger}$	$\mathbf{0.71}^{\dagger}$	$\mathbf{0.33}^{\dagger}$
	HAR-VRP-LIQ	$\mathbf{0.68}^{\dagger}$	$\mathbf{0.46}^{\dagger}$	$\mathbf{0.44}^{\dagger}$	$\mathbf{0.73}^{\dagger}$	$\mathbf{0.71}^{\dagger}$	$\mathbf{0.33}^{\dagger}$
	GARCH	0.01	0.02	$0.12^{\ddagger}$	0.04	0.04	0.09
	EWMA	0.01	0.02	0.07	0.04	0.09	$0.24^{\ddagger}$
	Deep-Base	0.00	0.00	0.01	0.04	0.00	$0.16^{\ddagger}$
	Deep-Liq	0.00	0.01	0.03	0.00	0.00	0.02
	Combo	$\mathbf{0.69}^{\dagger}$	$\mathbf{0.51}^{\dagger}$	$\mathbf{0.76}^{\dagger}$	$\mathbf{0.32}^{\dagger}$	$\mathbf{0.60}^{\dagger}$	$\mathbf{0.71}^{\dagger}$

Notes. Model Confidence Set (Hansen et al., 2011) p-values, out-of-sample period 2021-03 to 2026-01 (1,770 days), one panel per loss: the squared error of the log-variance forecast and the Jensen-corrected QLIKE of Patton (2011), Eq. (10). The MCS p-value is the largest test size at which a model is retained in the set; the best model has $p = 1.00$. † (**bold**) marks membership of the 75% set ($p > 0.25$), ‡ membership of the 90% but not the 75% set ($0.10 < p \leq 0.25$). Deep and Combo are 7-seed means; baselines are deterministic; HAR-classic$(1, 5, 22)$ is in Appendix B. The underlying raw MSE and QLIKE *levels* are tabulated in Appendix B (Table 9); all losses and tests are computed in log space and are invariant to the Jensen correction. The combination is the only model in the 75% squared-error set at *every* horizon for *both* assets (sole member for BTC); for ETH the parsimonious HAR is the single best model, the liquidation-aware HAR-LIQ and HAR-VRP-LIQ are statistically indistinguishable from it but no better, and the combination is comfortably retained.

would report. Once the HAR is given the liquidation inputs, however, the increment over the *liquidation-aware* HAR is *larger and more significant*, BTC $\lambda = 0.42$–0.63 ($t = 6.0$–6.5) and ETH $\lambda = 0.14$–0.38 ($t = 1.8$–2.8) at all three horizons, because the liquidation-aware HAR is the worse forecast and the deep model corrects more of its error. Controlling for the linear liquidation channel thus uncovers a significant nonlinear deep contribution for ETH as well as BTC. Neither model encompasses the other, so the combination is statistically warranted.

Table 4: Combination decomposition and the fair encompassing test (HAR⊕Deep-Liq), all horizons, out-of-sample. The deep increment is reported over the plain HAR and over the liquidation-aware HAR-VRP-LIQ.

			inv-MSE weight		Deep increment λ (t)	
Asset	H	ρ	HAR	Deep	over HAR	over HAR-VRP-LIQ
BTC	7	0.83	0.52	0.48	**0.37 (5.5)**	**0.42 (6.3)**
	14	0.81	0.52	0.48	**0.34 (4.2)**	**0.47 (6.0)**
	30	0.77	0.49	0.51	**0.36 (4.0)**	**0.63 (6.5)**
ETH	7	0.88	0.55	0.45	0.06 (0.8)	**0.14 (1.8)**
	14	0.88	0.55	0.45	0.07 (0.7)	**0.22 (1.9)**
	30	0.83	0.54	0.46	0.18 (1.3)	**0.38 (2.8)**

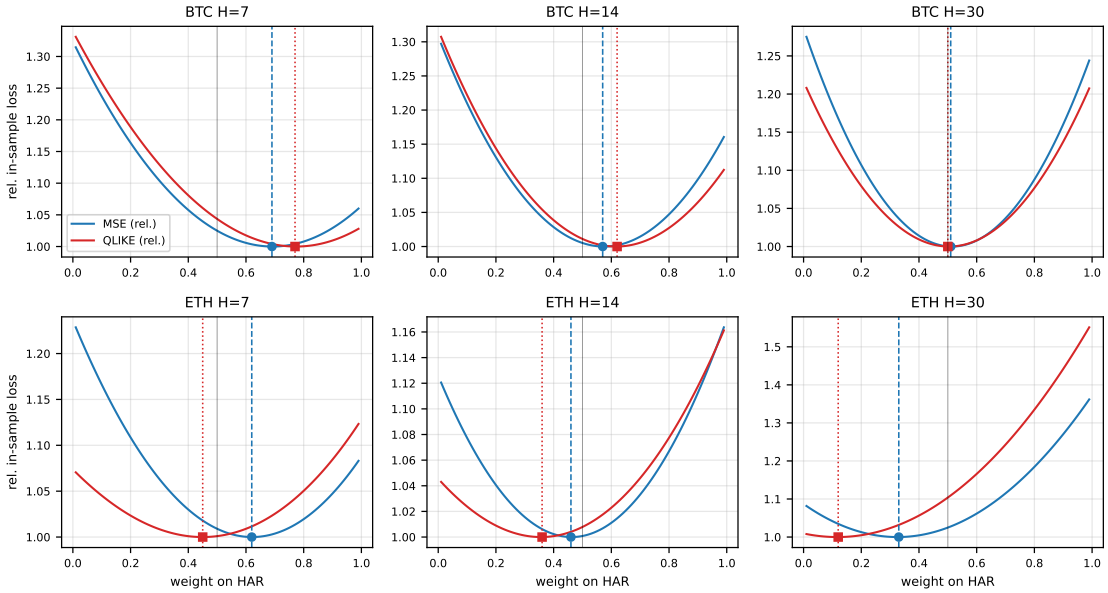
Notes. ρ is the correlation between the HAR and Deep-Liq out-of-sample forecast errors. Inverse-MSE (Bates–Granger) weights are $w_{\mathrm{HAR}} \propto 1/\mathrm{MSE}_{\mathrm{HAR}}$, normalized to one, stable under collinearity, unlike unconstrained optimal weights. The encompassing coefficient λ comes from regressing the *baseline* model's out-of-sample forecast error on the gap between the deep and baseline forecasts (Chong & Hendry, 1986): for "over HAR", $y_t - \hat{y}_t^{\mathrm{HAR}} = \lambda\,(\hat{y}_t^{\mathrm{Deep}} - \hat{y}_t^{\mathrm{HAR}}) + \varepsilon_t$, and "over HAR-VRP-LIQ" replaces the HAR baseline with the liquidation-aware HAR. A large, significant λ means the deep forecast explains part of the baseline's error, so it carries information the baseline lacks; the second column is the *fair* test, isolating the deep model's nonlinear use of the liquidation inputs the linear model already has. HAC t-statistics (truncated at the horizon) in parentheses, on the seed-mean out-of-sample path. **Bold** marks increments significant at the 10% level.

We turn to how the combination weight is selected. Figure 5 plots, for each asset and horizon, the squared-error and QLIKE losses of the combination as the HAR weight w sweeps from 0.01 to 0.99, evaluated on the held-out validation tail of training (Section 4.2); the dotted lines mark the loss-minimizing $w^{\star}_{\mathrm{MSE}}$ and $w^{\star}_{\mathrm{QLIKE}}$ that are then *frozen* and applied out of sample. Three things are visible. First, the curves are convex with interior minima, so a genuine optimal mix exists and the combination is not merely an artifact of averaging. Second, the minima are shallow, with the loss flat near the optimum, so the even split is close to optimal and the headline is insensitive to the exact weight, which is reassuring given that the weight is estimated. Third, and most informative, the optimum migrates with the asset and horizon in the same direction as the encompassing evidence: for BTC it sits at a HAR weight of 0.65/0.61/0.55 (MSE) and 0.73/0.66/0.53 (QLIKE) across $H = 7/14/30$, a balanced mix that tilts toward the deep model as the horizon lengthens; for ETH it falls from a near-even 0.63 (MSE) / 0.48 (QLIKE) at the weekly horizon to 0.27 / 0.09 at the monthly horizon, where the deep model earns most of the weight. That the data choose to weight the liquidation network most heavily at ETH's monthly horizon, precisely where its encompassing increment turns positive (Table 4), is independent corroboration that the selection is tracking real, horizon-specific forecasting content rather than noise.

5.3 Differences between BTC and ETH

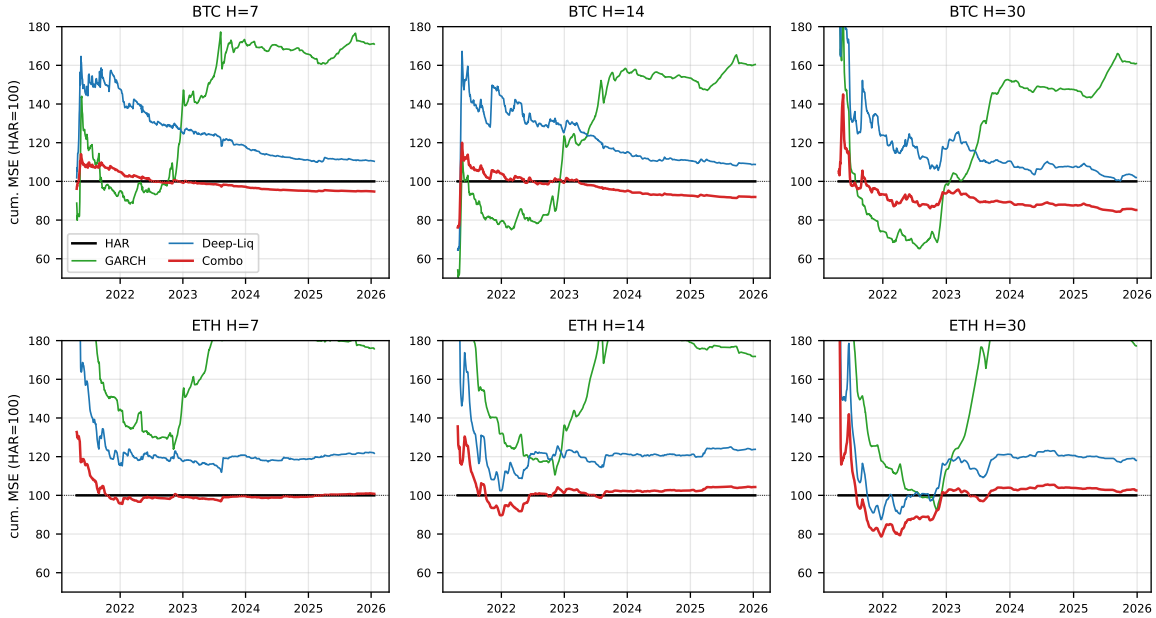
The two markets place different weight on the linear and nonlinear components, and the evidence is now sharper than a single statistic. For **BTC** the liquidation-aware deep model adds genuine *forecast* value:

Figure 5: Loss-optimal selection of the combination weight.



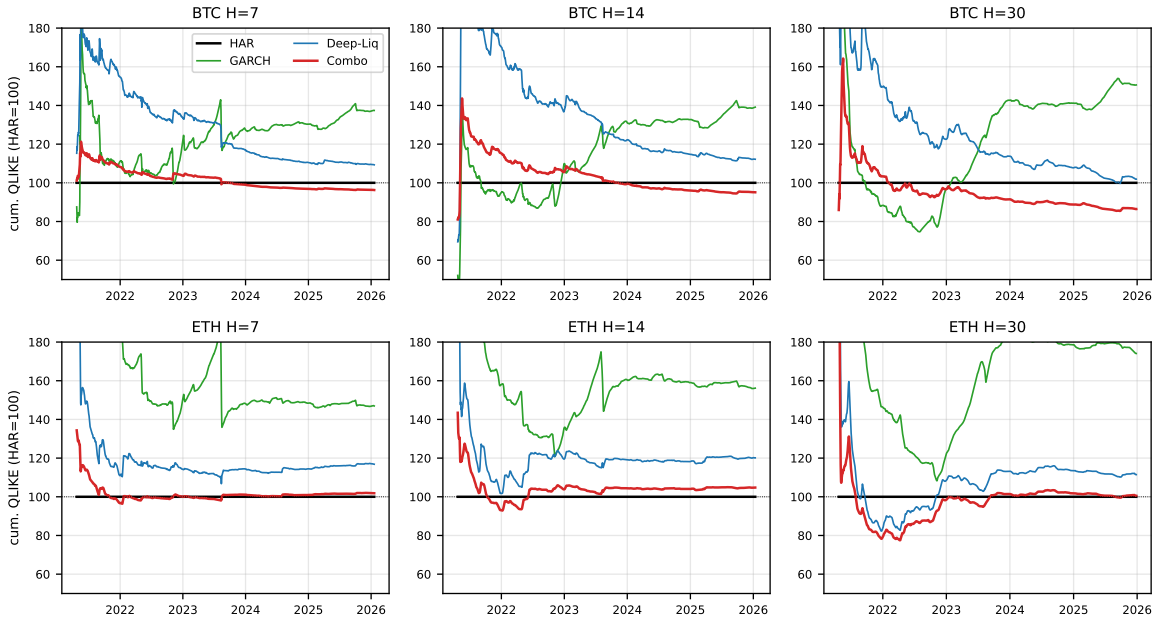
Notes. By asset (rows) and horizon (columns). Each panel shows the combination's squared-error (red) and QLIKE (blue) loss, relative to its own minimum, as the HAR weight w in Eq. (8) sweeps $[0.01, 0.99]$; dotted verticals mark the selected $w^{\star}_{\mathrm{MSE}}$ and $w^{\star}_{\mathrm{QLIKE}}$. Losses are evaluated only on the held-out final-year validation tail of the training sample, so $w^{\star}$ carries no look-ahead; the selected weights are then frozen for all out-of-sample evaluation and trading.

Figure 6: Cumulative out-of-sample squared-error loss over time, relative to HAR.



Notes. By asset (rows) and horizon (columns), with HAR$=100$. Each line is a model's running cumulative MSE divided by the HAR's, $\times 100$; below 100 means lower cumulative squared-error loss than the HAR benchmark. The first 30 days are omitted so that early-sample proportionality does not distort the scale, and the y-axis is common across panels. The combination (red) ends below 100 in every BTC panel and tracks the HAR closely for ETH, marginally above at the weekly horizon and at par by the monthly one, consistent with the squared-error MCS in Table 3.

Figure 7: Cumulative out-of-sample QLIKE loss over time, relative to HAR.



Notes. By asset (rows) and horizon (columns), Jensen-corrected, with HAR$=100$. Construction as in Fig. 6; below 100 is lower cumulative QLIKE than HAR. The HAR family is strongest under QLIKE, especially for ETH and at the short horizons, consistent with the QLIKE Model Confidence Set in Table 3; the combination stays close to the HAR throughout and below it for BTC.

it lowers the deep loss significantly (Deep-Liq vs. Deep-Base, $p \leq 0.034$), enters the combination as a significant encompassing contributor over both the plain and the liquidation-aware HAR (Table 4), and lifts the combination to the lowest MSE at every horizon. For **ETH** the long-memory HAR is a formidable point forecaster, numerically the lowest MSE and QLIKE at the monthly horizon, which is why a single-model comparison reports the deep contribution as insignificant. The *fair* encompassing test corrects that impression: once the HAR carries the same liquidation inputs, the deep increment over it is significant for ETH at all three horizons (Table 4), so the nonlinear channel does add forecast value for ETH, on top of the larger and more regime-robust *economic* edge it delivers through the trade gate (Section 5.4). Table 2 rationalizes the split: ETH's $\log \mathrm{RV}$ is less weekly-persistent ($\rho(7) = 0.58$ with a fitted weekly HAR coefficient an order of magnitude below its monthly one, Appendix B), so its variance is more jump- and event-driven than smoothly autoregressive, the regime in which liquidation cascades carry information but also in which a parsimonious monthly-anchored HAR is hard to beat on average. This fits ETH being the higher-beta, more leverage-driven asset, whose forced deleveraging drives a larger share of variance, while BTC's more institutional flow leaves both a stronger linear signal and a cleaner marginal role for the liquidation network. The contract mechanics reinforce the channel: the inverse, coin-margined perpetuals that dominate these books are convex in the price move (Fig. 1), so a long position reaches maintenance margin at a smaller drop than a linear contract would and the forced-selling leg of the cascade steepens, the threshold response the linear model cannot price and the deep network can.

5.4 Trading performance

Table 5 reports the combination's net Sharpe ratio for every horizon and trade rule and Fig. 8 plots it against the benchmarks. We summarize trading performance by the annualized net Sharpe ratio, the natural risk-adjusted measure for a volatility-targeted overlay: it scores return per unit of the variance risk the desk bears and stays comparable across horizons whose holding periods and trade counts differ. Ungated directional trading is weak. Gating on liquidation or stress risk is what generates performance, and it does so in a *horizon-specific* pattern that mirrors the forecasting evidence. For **ETH the edge is at the weekly horizon**: the gated combination earns a net Sharpe near two at $H{=}7$ (1.97 liquidation-gated, 1.74 stress-gated) and the standalone Deep-Liq model reaches 1.62 there, fading to 0.99 and 0.78 at the biweekly and monthly horizons. For **BTC the edge is at the monthly horizon**: the combination's liquidation-gated Sharpe is 1.18 at $H{=}30$ and the standalone Deep-Liq overlay reaches 1.75 at the liquidation gate, above one but below the ETH weekly figure, while BTC's weekly gated Sharpe (1.13) is unremarkable. The **biweekly $H{=}14$ horizon is weak for both assets and every gate**, a consistent dead spot. This short-versus-long split is itself a finding: ETH's variance, being more jump- and liquidation-driven, rewards the timing gate at the horizon where liquidation cascades resolve fastest, while BTC's smoother, more persistent variance rewards it at the monthly horizon where the deep model also earns its largest forecast weight.

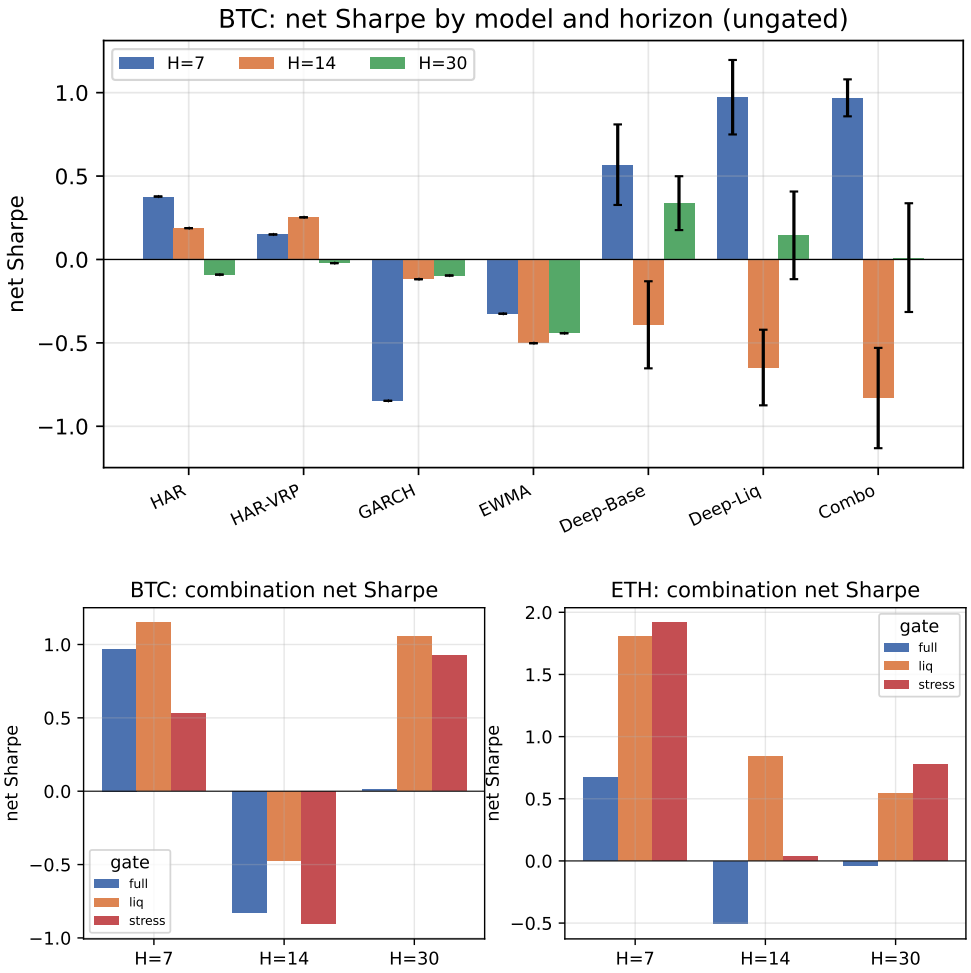
These gated figures rest on small, non-overlapping samples (e.g. 58 monthly and ~ 17 monthly stress-gated trades) across a grid of two assets, three horizons, and three gates, so we read them as economic corroboration of the daily forecast tests (MCS, Diebold–Mariano) rather than as standalone significance. The premium is, on average, a thin edge that the liquidation signal tells the strategy *when* to take. Across the full model–gate matrix (Fig. 9) the liquidation gate lifts every model above its ungated rule, and the stress gate does so in most cells, so the timing gain is shared across forecasters, while the combination's and deep model's advantage is on the forecast side (the MCS and encompassing results above). The gate quantile is a design choice, and the result is not knife-edge in it: sweeping the cutoff from the median to the top quintile ($q = 0.5/0.67/0.8$) the strong-horizon Sharpe rises monotonically (Appendix B, Table 10), as a stricter gate concentrates capital in fewer, more informative states at the cost of trade count; we report the top tercile as the balance between selectivity and sample size. Doubling all Deribit costs leaves the gated Sharpe positive.

Table 5: Combination net Sharpe by horizon and trade rule.

Asset	Gate	$H{=}7$	$H{=}14$	$H{=}30$
BTC	full	1.09	-0.88	0.18
	liq	**1.13**	$\mathbf{-0.67}$	**1.18**
	stress	0.74	-1.18	1.07
ETH	full	0.88	-0.45	0.07
	liq	**1.97**	**0.99**	**0.78**
	stress	1.74	-0.04	0.73

Notes. Net Sharpe ratio (annualized) of the replicated variance-swap strategy of Eq. (11) on a $10,000 account, net of Deribit costs, 7-seed means; **bold** is the best gate per asset $\times$ horizon. "full" trades every period; "liq" only when λ_t is in its top tercile; "stress" only on stress days. Trade counts shrink with the gate (full/liq/stress $= 58/20/17$ at $H{=}30$, versus $252/99/56$ at $H{=}7$), so gated figures are economic illustrations rather than significance claims. The trade rule compares the Jensen-corrected mean-variance forecast of Eq. (9) to the strike. The full $2 \times 8 \times 3$ model–gate matrix is in Fig. 9.

Figure 8: Net Sharpe by model, horizon, and trade rule.



Notes. Top: ungated net Sharpe by model and horizon, BTC (mean ± s.d. over seeds). *Bottom:* the combination's net Sharpe by trade rule and horizon, both assets. The liquidation and stress gates lift the ungated rule, and the edge is horizon-specific: strongest at $H{=}7$ for ETH and at $H{=}30$ for BTC, with $H{=}14$ weak for both.

6 Robustness

We first check that the headline does not stem from network capacity, training three deep specifications differing only in regularization: heavily regularized (in-sample $R^2 \approx 0.5$), moderate (≈ 0.75), and near-unconstrained (≈ 0.99). The resulting combination's out-of-sample MSE and QLIKE are essentially identical across all three at every horizon, differing by at most about 0.02 in each loss at H=7, 14, and 30 (Table 10), so the headline does not stem from network capacity or in-sample over-fitting. We report the near-unconstrained net, marginally the most accurate out of sample.

We next probe sensitivity to the estimation protocol, re-estimating the entire pipeline under three walk-forward schemes (Table 6): the headline *expanding* anchored window, a *rolling two-year* window, and a window that *excludes the 2020 COVID period* from training, all with quarterly refits. Stage 1 is held fixed across schemes, fit once on data preceding each scheme's first forecast, because the intraday relationship between order flow and liquidation bursts changes far more slowly over time than the variance level itself; only Stage 2, the HAR, and the variance filters are re-estimated each quarter. Three findings stand out. (i) The combination's forecast standing is robust: it beats the standalone Deep-Liq model by a Diebold–Mariano test at every horizon under every scheme, and it is retained in the squared-error MCS in all six asset–horizon cells under both the headline expanding window and the COVID-excluded window, dropping out of only one cell under the more demanding rolling two-year window. (ii) The parsimonious HAR remains the standout for ETH under all three schemes, retained in both confidence sets at every horizon, so the asset asymmetry is not a protocol artifact. (iii) The economic edge survives at each asset's strong horizon: the BTC liquidation-gated combination posts a *monthly* net Sharpe of 1.18, 2.34, and 1.90 across the three schemes, and the ETH gate a *weekly* Sharpe of 1.97, 2.34, and 2.45, so neither the forecast result nor the (horizon-specific) trading edge depends on the inception-anchored protocol or on the 2020/2021 stress episodes. The off-peak horizons are more variable (Table 6): the weekly BTC and monthly ETH gates remain positive under every scheme, but the biweekly gate is weak or negative for both assets, the same dead spot the forecast tests show.

Table 6: Walk-forward robustness of the combination (near-unconstrained net).

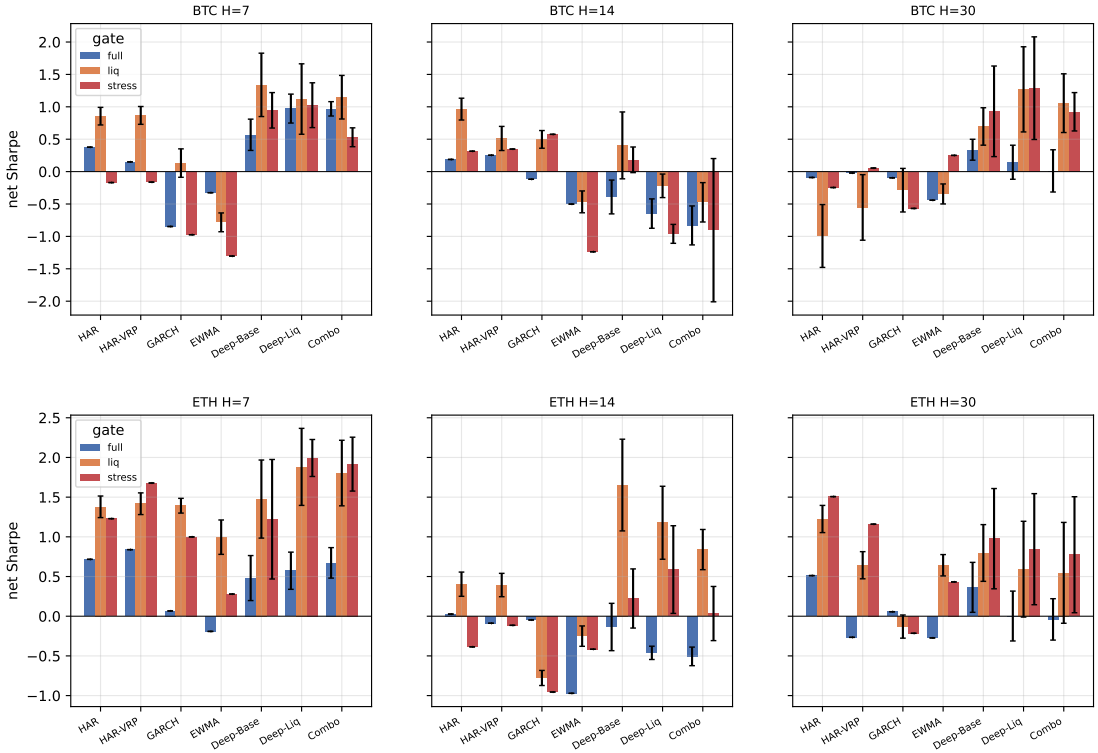
Scheme	MSE-MCS	QLIKE-MCS	BTC liq $S_{7/14/30}$	ETH liq $S_{7/14/30}$
Expanding (headline)	6/6	6/6	$1.13/-0.67/1.18$	$1.97/0.99/0.78$
Rolling 2-year	5/6	4/6	$1.60/-0.45/2.34$	$2.34/1.04/0.39$
COVID-excluded	6/6	6/6	$1.73/-0.40/1.90$	$2.45/0.73/0.71$

Notes. The pipeline re-estimated under three walk-forward schemes with quarterly refits: an expanding anchored window (the body's headline), a rolling two-year window, and a window excluding the 2020 COVID period from training. The two MCS columns count the asset×horizon cells (of six) in which the combination is retained in the 75% squared-error and Jensen-corrected QLIKE Model Confidence Sets. It is retained in all six cells under the expanding and COVID-excluded windows; under the rolling window it leaves one squared-error cell (BTC weekly) and two QLIKE cells (BTC weekly and biweekly). The last two columns give the liquidation-gated combination's net Sharpe at the weekly/biweekly/monthly horizons. Under every scheme the combination beats the standalone Deep-Liq model (Diebold–Mariano, all horizons) and the parsimonious HAR is itself retained in both confidence sets at every horizon for ETH, so the BTC-favouring forecast asymmetry is not a protocol artifact.

For completeness of the robustness testing we also examine how the economic edge depends on the market regime. Splitting the trade record at end-2021, the BTC stress-gated edge is concentrated *before* 2022 (the 2021 cascade, pre-2022 stress-gated net Sharpe 3.2 versus a weak post-2022 figure), whereas the ETH stress-gated edge *survives* into the post-2022 sample (post-2022 stress-gated mean

net Sharpe positive but noisy), and each asset's strong-horizon liquidation-gated edge, BTC monthly and ETH weekly, persists under quarterly walk-forward refitting (Table 6). Stress days, defined by the trailing-percentile rule, recur in every year of the sample, not only in 2021, so the stress gate is not a single-episode artifact. The cost and gate-quantile sensitivities, the architecture-robustness check, and the post-2022 figures are collected in Appendix B (Table 10): doubling all Deribit costs leaves the headline monthly BTC gated Sharpe positive ($1.75 \to 1.47$), and tightening the gate raises the Sharpe at each asset's strong horizon ($q = 0.5/0.67/0.8$: BTC monthly $0.9/1.8/1.9$, ETH weekly $0.7/1.6/2.3$), while the combination's losses are essentially flat across the three regularization levels, so the headline stems from neither cost assumptions nor network capacity.

Figure 9: Net Sharpe of every model under the full, liquidation, and stress gates.



Notes. BTC (top) and ETH (bottom), one panel per horizon $H \in \{7, 14, 30\}$. Bars are seed-mean net Sharpe ratios. The gates lift every model, which isolates the timing gain from the forecast gain.

7 Discussion

What do these results say about how crypto variance is generated? They point to two regimes with different dynamics. In normal states variance is a smooth, persistent cascade that a linear HAR captures well, the long-memory behaviour common across asset classes. In stressed states it is driven by forced deleveraging, a convex, threshold response in which a funding-liquidity spiral turns leverage into realized variance (Brunnermeier & Pedersen, 2009): negligible while leverage holds, explosive when it breaks. Our forecasting evidence is the statistical signature of this split. A linear model handed the liquidation channel cannot use it, and is harmed by it, because the map from forced flow to variance is convex and a single linear coefficient must blend the calm and stressed states into a number that fits neither; the deep network, which can represent the threshold, extracts a significant nonlinear increment

over the very same linear information set, for ETH as well as BTC. The premium and the liquidation tape are complementary in this picture: the premium prices how much variance risk costs, while the liquidations reveal when the convex channel fires. That is why conditioning on both improves on either, and why the tradable edge comes from a timing gate that concentrates the position in exactly the states where the channel is active rather than from the point forecast alone. The two channels also explain the asset split we report below: the network's *forecast* value, its nonlinear use of the liquidation signal, is strongest for BTC, whereas for ETH the same signal pays off mainly through the *timing* gate.

Two results ran against our prior expectation and are worth dwelling on. The first is the biweekly dead spot: at $H{=}14$ the gated overlay is weak for both assets and every gate, even though the weekly and monthly horizons are strong. We read this as the two economic mechanisms passing each other, with the fast liquidation cascade largely resolved by two weeks, so the timing gate has little left to exploit, while the slow, persistent-variance channel that rewards the monthly horizon has not yet taken over. Part of the weakness may be measurement rather than economics: the premium is anchored at a single 30-day tenor (below), so it is least aligned with the forecast horizon at the intermediate $H{=}14$. The second is the asset asymmetry in forecasting: we had expected the liquidation network to help both coins, but its standalone forecast edge is a BTC phenomenon, and for ETH the parsimonious HAR is hard to beat on point loss. The descriptive evidence rationalizes this, since ETH's variance is more jump- and event-driven and less weekly-persistent, so a monthly-anchored HAR captures most of its predictable variation while the liquidation signal pays off through the trade gate rather than the point forecast. We cannot fully separate this structural reading from a data one: ETH's option market is younger and thinner than BTC's, so part of the HAR's relative strength for ETH may reflect a noisier reconstructed premium rather than a difference in variance dynamics alone. A further limitation bears on the weaker cells directly: our reconstructed index, like DVOL, is fixed at a single 30-day tenor, whereas the forecast and trade horizons range from one week to one month, and recent evidence that aligning the premium's maturity to the forecast horizon sharpens its predictive power (Plíhal & Mampouya, 2025; Plíhal et al., 2025) suggests that a term structure of implied variance, which the sparse liquidation prints cannot yet support, would likely improve the off-tenor horizons once a denser historical option cross-section is available.

The findings are bounded. The reconstructed proxy carries a ≈ 6 volatility-point error and a visibly higher day-to-day variance than the published index, both of which a dense historical option chain would reduce; the DVOL^2 strike is a model-free squared-volatility index that understates the fair variance-swap strike by a convexity term, so the measured premium and short-variance returns are, if anything, conservative; crypto variance swaps are synthetic, traded through an option replication; the gated-trade samples are small, so the economic figures are illustrations rather than significance claims; and the evidence covers two assets on one venue. Our walk-forward checks bound, but do not cleanly isolate, the reconstruction's pre-2021 contribution: the COVID-excluded scheme still trains on the 2019–2021 reconstructed premium and only withholds the 2020 window, so it shows the headline does not hinge on the COVID episode without separately identifying the marginal value of the reconstructed era, which a dense historical option chain would permit. One channel is visible, however: without the reconstruction the premium-augmented HAR is estimated on barely a year of pre-sample and its VRP coefficient is mechanically near zero, whereas the multi-year reconstructed premium lets that coefficient be genuinely estimated, so for the linear premium channel the reconstruction's value is a lower bound we can

point to rather than merely assert. Extending the design to more venues and instruments, and using the reconstruction to bring further optioned coins into the panel, is the natural next step.

Two of our results generalize beyond the specific assets and venue. The reconstruction technique applies to *any* market whose liquidation or trade feed prints option-implied volatilities: where a volatility index is young or absent, the historical implied volatility tape embedded in the liquidation record can extend it, at an error we have quantified and that a denser chain would shrink. The method needs no pre-existing index at all, so any cryptocurrency with a liquid option market whose feed carries implied volatilities can be handed a synthetic implied variance history built directly from those prints, even where no volatility index has ever been published, and without waiting years for a native index to accumulate. We demonstrate this in two stages: the model-free DVOL methodology rebuilt from full option surfaces reproduces the published BTC and ETH indices at correlations of 0.997 and 0.90, and applied to SOL and XRP, two coins with no published index, the liquidation tape alone recovers that index at correlations of 0.79 and 0.90 (Appendix C). And the economic object, a published *forced-flow* stream used as a conditioning state variable, is not crypto-specific in principle, only in availability: any venue that discloses its forced deleveraging supplies a leading signal for variance that is useful for volatility targeting, derivative pricing, and, through public liquidation intensity, real-time market-fragility monitoring by exchanges and regulators.

8 Conclusion

This paper forecasts and prices cryptocurrency realized variance by pairing the variance risk premium with a signal unique to these markets, the exchange-published stream of forced liquidations. The exercise is constrained by a short implied variance history, a limitation this literature has largely accepted; we address it directly. The contribution has a methodological and an economic component, each of which extends beyond the two assets and single venue studied here.

The methodological contribution is a way to construct a volatility index where none long enough exists. The option-implied volatilities embedded in the liquidation tape, sparse and previously discarded, can be turned by a kernel-weighted projection onto the at-the-money 30-day point and a fixed affine calibration into a daily implied variance series that tracks the published Deribit DVOL at a correlation of 0.94/0.96 and reaches back across the March-2020 cascade. Because the index is recovered from the freely published forced-trade record itself, the recipe needs no pre-existing index and transfers to any optioned asset, at the cost of a series noisier than the exchange's. This is the first use of liquidation-borne implied volatilities to extend a crypto volatility index.

The economic contribution, which is where the paper's weight lies, refines rather than overturns the long-running contest between the linear heterogeneous autoregression and flexible machine learning. The liquidation signal's predictive content is irreducibly *nonlinear*, which we establish with a fair comparison rather than an assertion: given the same liquidation features, the log notional and signed imbalance at $1/7/28$ lags, a linear HAR cannot extract them and is harmed by overfitting (its in-sample fit rises while its out-of-sample loss worsens), whereas a deep network carries a strongly significant encompassing increment over that liquidation-aware HAR, at every horizon for BTC ($t = 6.0$–6.5) and significantly for ETH ($t = 1.8$–2.8), where a comparison against the plain HAR had found nothing. The HAR remains hard to beat, as the literature has long held; the lesson is that the linear cascade and the

nonlinear liquidation response are *complementary*, each correcting errors the other makes, so only their combination is never excluded from the squared-error Model Confidence Set for either asset. The novelty is in the *conditioning information*, a published forced-flow state variable, and in the disciplined hybrid that exploits it rather than in the modelling machinery, a concrete instance of the linear-plus-nonlinear template that the volatility-forecasting literature has called for but rarely grounds in economics.

Across all three horizons the two effects separate cleanly. The deep model's *forecast* value, its nonlinear use of liquidations, is significant for both assets once the linear channel is controlled (strongest for BTC at all horizons); the *economic* edge from gating on liquidation risk is sharply horizon-specific and asset-specific, weekly for ETH, monthly for BTC, weak at the intermediate horizon for both, which we trace to the two markets' differing leverage and liquidation intensity. Separating the short-horizon from the long-horizon contribution is itself a contribution: it disciplines the conclusions to the level of generality the evidence supports and shows that a liquidation signal can be informative at the horizon where cascades resolve (ETH, weekly) or where they accumulate into persistent variance (BTC, monthly), depending on market structure.

The practical implications follow directly. A published forced-flow stream is a leading, real-time state variable for variance that is useful for volatility targeting, for setting and hedging variance-swap and option strikes, and, through public liquidation intensity, for market-fragility monitoring by exchanges and regulators; the reconstruction makes the historical version of that signal available wherever the live feed exists. The broader message is simple and, we hope, immediately usable: in markets that publish their forced flow, that flow is a tradable, forecastable state variable conditioning the variance risk premium, and a disciplined, horizon-aware combination of a linear premium model with a nonlinear liquidation-aware network, trained on a history that the liquidation tape itself can supply, exploits it. Extending the design to more venues and instruments, and using the reconstruction to bring further optioned coins into the panel, is the natural and now low-cost next step.

A Architecture, hyper-parameters, and estimation

Every forecast uses only information dated up to t, and the target is the *forward* average realized variance over $t{+}1, \ldots, t{+}H$. $\mathrm{RV30}_t$ and the burst thresholds are trailing quantities, and DVOL_t enters as the same-day variance price (the swap strike), which is not circular because the target is future realized variance. Stage-2 input windows end at $t{-}1$, and standardization, the burst thresholds, and the validation split use training-block statistics only. The liquidation regressors of HAR-LIQ and HAR-VRP-LIQ are trailing $1/7/28$-day averages of day-t quantities observed at the close of day t. In the reconstruction, the proxy of Eq. (1) uses backward gap-fill and a backward exponential moving average and is calibrated on the 2021-onward overlap; the real published DVOL is used for the test set and all trading, the reconstructed proxy feeding training only.

The per-seed estimation procedure is as follows. We standardize the Stage-1 features on the pre-2021 block and train the burst classifier of Eq. (6), score all bars out of sample, and form λ_t. For each horizon H we build the Stage-2 input Z_t and train the forecaster walk-forward, refitting every 91 days (a calendar quarter). We form the HAR by walk-forward OLS and the selected-weight combination of Eq. (8), the weight fixed on the training-tail validation year (Section 4.2), and we trade by Eq. (11).

The headline-net hyper-parameters are as follows. Stage 1 uses kernels $\{3, 6, 12\}$, 32 channels,

LSTM width 64, and dropout 0.3; Stage 2 uses kernels $\{3, 5, 22\}$, 24 channels, LSTM width 48, and dropout 0.3. Optimization is Adam at a learning rate of 10^{-3}; the horizons are $\{7, 14, 30\}$. The strategy uses capital \$10,000, a capital fraction of 0.15, a volatility target of 0.60, a leverage cap of 1.5, and a top-tercile gate; costs are 0.03% per leg, a 1.25-volatility-point replication spread, 4% funding, and 15% collateral; the Model Confidence Set uses size 0.25 via the stationary bootstrap, and P&L significance uses a moving-block bootstrap.

The three Stage-2 specifications of the capacity-robustness check (Section 6) differ only in regularization: a *regularized* net (16/32 channels/LSTM width, dropout 0.45, weight decay 10^{-4}, early stopping, in-sample $R^2 \approx 0.5$); a *moderate* net (20/40, dropout 0.38, weight decay 5×10^{-5}, early stopping, $R^2 \approx 0.65$); and the *headline near-unconstrained* net (24/48, dropout 0.30, weight decay 10^{-5}, no early stopping, 120 epochs, $R^2 \approx 0.98$). The early-stopped variants halt when the held-out validation loss fails to improve for 15/25 epochs (cap 250); each net is trained from seven random seeds and the forecasts are averaged.

Reproducibility. Every result comes from a single notebook run on public Deribit data (DVOL and the liquidation books) plus a 5-minute price feed, with the seven seeds and library versions fixed; the code and the reconstructed series will be released on publication.

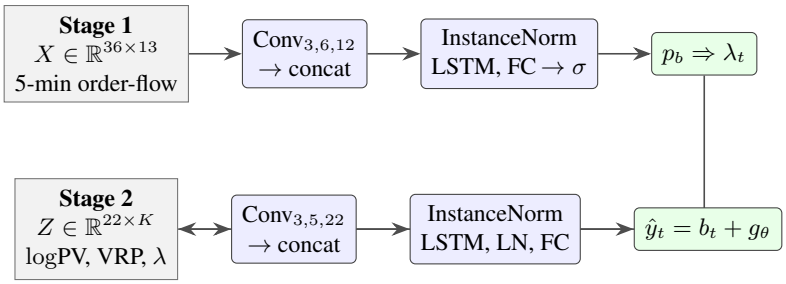


Figure 10: Two-stage multi-scale CNN–LSTM. Stage 1's daily liquidation risk score λ_t conditions the Stage 2 variance forecaster; the headline forecast is the selected-weight combination of Stage 2's DeepLiq output with HAR$(1, 7, 28)$.

B Supporting results

Table 7 reports autocorrelations at the competing HAR lags: the crypto-calendar lags $(7, 28)$ exceed the equity-convention lags $(5, 22)$ in every cell for both realized variance and liquidation notional, supporting the $(1, 7, 28)$ specification. Table 8 reports the Newey–West HAR$(1, 7, 28)$ coefficients at all three horizons. The weight shifts from the daily toward the monthly component as the horizon lengthens, the standard heterogeneous-cascade pattern, and the monthly term dominates for both assets, increasingly so for ETH (its monthly coefficient rises from 0.41 at the weekly horizon to 0.64 at the monthly), consistent with ETH's more jump- and event-driven variance and with the weekly-versus-monthly split in the economic results (Section 5.3).

The linear liquidation channel overfits for a clear reason. Augmenting the HAR with the liquidation regressors (HAR-VRP-LIQ) *raises* the in-sample fit at every horizon, for BTC from $R^2 = 0.80/0.72/0.59$ to $0.81/0.74/0.65$ at $H = 7/14/30$, yet the out-of-sample MSE *worsens* (Table 9). The Newey–West coefficients explain why: the liquidation loadings describe no sensible forward map. The signed imbalance at every lag is insignificant ($|t| \leq 0.8$), and the log notional loads small and wrong-

Table 7: Sample autocorrelation at the competing HAR lags. **Bold** = larger of the classic $(5, 22)$ vs. crypto $(7, 28)$ pair.

		Weekly		Monthly	
Asset	Variable	lag 5	lag 7	lag 22	lag 28
BTC	$\log \mathrm{RV}$	0.778	**0.807**	0.633	**0.644**
	$\log$ liq. notional	0.187	**0.275**	0.115	**0.248**
ETH	$\log \mathrm{RV}$	0.712	**0.740**	0.631	**0.661**
	$\log$ liq. notional	0.200	**0.256**	0.154	**0.218**

Notes. Sample autocorrelations of the log series at the equity-convention weekly/monthly lags $(5, 22)$ versus the crypto-calendar lags $(7, 28)$, over the full estimation history; **bold** marks the larger of each pair. The crypto lags are larger in every cell, supporting the $(1, 7, 28)$ specification.

Table 8: HAR$(1, 7, 28)$ coefficients at all horizons (full sample), Newey–West t in parentheses.

Asset	H	const	daily (1)	weekly (7)	monthly (28)	fit R^2
BTC	7	1.05 (4.9)	0.33 (7.5)	0.41 (6.7)	0.16 (2.4)	0.79
	14	1.33 (4.5)	0.29 (6.2)	0.31 (4.4)	0.26 (3.0)	0.74
	30	1.66 (4.3)	0.22 (5.0)	0.24 (3.3)	0.37 (4.0)	0.68
ETH	7	0.88 (3.8)	0.23 (9.4)	0.28 (5.0)	0.41 (7.5)	0.77
	14	1.13 (3.8)	0.18 (7.9)	0.18 (3.2)	0.53 (8.3)	0.76
	30	1.43 (3.9)	0.12 (6.4)	0.11 (2.2)	0.64 (10.0)	0.77

Notes. OLS regression of $\log \mathrm{RV}^{(H)}$ on a constant and the trailing 1-, 7- and 28-day averages of $\log \mathrm{RV}$, over the full sample, at each horizon H. Newey–West (Newey & West, 1987) t-statistics (HAC, truncated at the horizon) in parentheses. As H rises the weight migrates from the daily to the monthly term; the monthly component dominates at the monthly horizon for both assets and is largest for ETH, consistent with its more jump-driven variance. The in-sample fit R^2 is reported only for these linear benchmarks, where the regressor count is well defined.

signed, its daily lag carrying a counterintuitive negative coefficient. There is thus no stable linear map from liquidations to forward variance; the regressors fit sample-specific noise that does not generalize, exactly the divergence between in-sample and out-of-sample fit that signals overfitting and the reason the informative, nonlinear liquidation response is recovered only by the deep network (Section 5.1).

Table 9 reports the raw out-of-sample loss *levels* underlying the Model Confidence Set p-values of Table 3, and Table 10 the cost, gate-quantile, and architecture sensitivities discussed in Section 6.

Table 9: Out-of-sample forecast losses (levels) underlying the MCS p-values of Table 3. Lower is better; per-column minimum in **bold**.

		MSE (log variance)			QLIKE (Jensen-corrected)		
Asset	Model	H=7	H=14	H=30	H=7	H=14	H=30
BTC	HAR	0.37	0.38	0.36	0.23	0.22	0.19
	HAR-VRP	0.37	0.39	0.37	0.23	0.22	0.19
	HAR-LIQ	0.38	0.41	0.44	0.24	0.23	0.23
	HAR-VRP-LIQ	0.38	0.42	0.46	0.24	0.24	0.24
	GARCH	0.63	0.61	0.59	0.32	0.31	0.28
	EWMA	0.45	0.46	0.44	0.28	0.26	0.23
	Deep-Base	0.50	0.54	0.50	0.29	0.29	0.26
	Deep-Liq	0.43	0.46	0.40	0.27	0.26	0.21
	Combo	**0.35**	**0.36**	**0.32**	**0.22**	**0.21**	**0.17**
	Combo-VL	0.36	0.38	0.35	0.23	0.22	0.18
ETH	HAR	**0.36**	**0.35**	**0.32**	0.24	**0.21**	**0.16**
	HAR-VRP	0.36	0.37	0.35	0.24	0.21	0.18
	HAR-LIQ	0.37	0.38	0.35	**0.23**	0.21	0.18
	HAR-VRP-LIQ	0.37	0.38	0.36	0.24	0.21	0.18
	GARCH	0.62	0.61	0.56	0.35	0.32	0.28
	EWMA	0.51	0.50	0.45	0.33	0.28	0.23
	Deep-Base	0.51	0.49	0.46	0.33	0.27	0.22
	Deep-Liq	0.46	0.45	0.40	0.29	0.25	0.19
	Combo	0.36	0.37	0.34	0.25	0.22	0.18
	Combo-VL	0.38	0.39	0.35	0.26	0.23	0.18

Notes. Out-of-sample MSE of the log variance forecast (Jensen-invariant) and Jensen-corrected QLIKE (Patton, 2011), 7-seed means; the corresponding multiple-comparison MCS p-values and set membership are in Table 3. The combination has the lowest MSE for BTC at every horizon; the parsimonious HAR is numerically lowest for ETH. Adding the liquidation channel *linearly* (HAR-LIQ, HAR-VRP-LIQ) raises the loss for BTC, sharply so at the monthly horizon ($0.36 \rightarrow 0.46$ MSE), and does not help ETH, the out-of-sample signature of overfitting discussed in Section 5.1. Combo-VL is the analogous equal-weight combination using the fully loaded HAR-VRP-LIQ in place of the plain HAR; its uniformly higher loss than Combo confirms that the parsimonious HAR is the better linear partner for the deep model.

Table 11 reports the Diebold–Mariano tests of equal predictive accuracy that the body summarizes. The combination is significantly more accurate than its standalone deep component at every horizon for both assets (BTC $t = 2.7$–3.4, ETH $t = 4.0$–4.4), and the liquidation-aware deep model improves significantly on the premium-only Deep-Base, the liquidation ablation, for BTC and marginally for ETH.

C Generalization to assets without a published volatility index

The reconstruction of Section 3.2 needs no pre-existing index, so in principle it transfers to any optioned asset whose feed prints liquidation-borne implied volatilities. We establish this in two stages over 2024-12 to 2025-12, the window for which we hold full Deribit option surfaces. The window is far too short for the forecasting and trading exercise of the main paper, so we use it only to ask whether the construction *recovers a volatility index*.

Table 10: Sensitivity of the liquidation-gated Deep-Liq net Sharpe across horizons (seed means).

	BTC			ETH		
	H=7	H=14	H=30	H=7	H=14	H=30
Gate $q = 0.5$	1.41	−0.32	0.85	0.67	0.57	0.75
Gate $q = 0.67$ (headline, $1\times$ cost)	1.05	−0.41	1.75	1.62	1.45	0.78
Gate $q = 0.8$	0.79	−0.57	1.92	2.31	1.04	1.21
Headline gate, $2\times$ cost	0.71	−0.63	1.47	1.30	1.12	0.62

Notes. Net Sharpe (annualized, seed means) of the liquidation-gated standalone Deep-Liq variance-swap overlay, by asset and horizon. The gate quantile q sets how selective the liquidation gate is (higher q = fewer, higher-risk days; $q = 0.67$ is the headline top-tercile gate); the last row applies the headline gate under a doubled Deribit cost schedule. Tightening the gate raises performance at the monthly horizon for BTC and across horizons for ETH, and the gated Sharpe stays positive under doubled costs at each asset's strong horizon. In a separate architecture-robustness check the combination's out-of-sample losses are essentially identical across three deep networks differing only in regularization (in-sample $R^2 \approx 0.5/0.75/0.99$) at *every* horizon. Combo MSE (reg/moderate/unconstrained) is $0.35/0.35/0.35$, $0.37/0.36/0.36$, $0.32/0.32/0.32$ for BTC at H=$7/14/30$ and $0.36/0.36/0.36$, $0.36/0.37/0.38$, $0.33/0.34/0.34$ for ETH; Combo QLIKE is likewise flat (BTC $0.23/0.21/0.17$ across horizons, ETH $0.25/0.23/0.19$), so the headline does not stem from network capacity.

Table 11: Diebold–Mariano tests of equal predictive accuracy.

Asset	Comparison	H=7	H=14	H=30
BTC	Combo vs. Deep-Liq	3.4^{***}	3.2^{***}	2.7^{***}
	Deep-Liq vs. Deep-Base (liq. ablation)	2.6^{**}	2.1^{**}	2.5^{**}
ETH	Combo vs. Deep-Liq	4.4^{***}	4.0^{***}	4.0^{***}
	Deep-Liq vs. Deep-Base (liq. ablation)	1.9^{*}	1.8^{*}	2.3^{**}

Notes. Diebold–Mariano t-statistics on the seed-mean out-of-sample squared-error loss, with the Harvey–Leybourne–Newbold small-sample correction and a HAC variance truncated at the horizon; a positive statistic means the first model has the lower loss. "Combo vs. Deep-Liq" tests the combination against its deep component; "Deep-Liq vs. Deep-Base" is the liquidation ablation (the deep model with versus without the liquidation channel). *, **, *** denote significance at the 10, 5, and 1% level.

The first stage confirms that DVOL itself can be rebuilt from option data, by applying Deribit's own model-free, VIX-style methodology to the full option surface: for each listed expiry we recover the forward from put-call parity at the strike minimizing the call–put price gap, integrate the out-of-the-money option prices,

$$\sigma_T^2 = \frac{2}{T}\sum_i \frac{\Delta K_i}{K_i^2}\, Q(K_i) - \frac{1}{T}\Big(\frac{F}{K_0} - 1\Big)^2, \tag{12}$$

discarding options below a five-percent delta, and interpolate the two expiries bracketing 30 days to the constant 30-day tenor. Compared with the *published* Deribit DVOL over the window, this model-free index reproduces it at a correlation of 0.997 for BTC and 0.895 for ETH (calibrated R^2 of 0.99 and 0.80, mean levels 50 versus 47 and 73 versus 70 volatility points; Table 12, Fig. 11). The construction is therefore Deribit's own index methodology, which we now use as ground truth for coins that have no published index.

The second stage shows that the liquidation tape alone recovers the index for two coins that have none. Solana (SOL) and Ripple (XRP) have no published DVOL. We build their model-free DVOL from the full surface as above, then run the liquidation-only reconstruction of Eq. (1) on their option-*liquidation* IVs alone and compare. Option liquidations on these thinner markets are sparse, with qualifying prints on 268 (SOL) and 118 (XRP) days of the window at a median of one to two contracts per active day; the backward exponential moving average of Eq. (1) carries them across the intervening days into a continuous daily series, exactly as in the main sample. The reconstruction recovers the model-free DVOL at a correlation of 0.79 (SOL) and 0.90 (XRP), with calibrated R^2 of 0.62 and 0.81 (Table 12, Fig. 11). Because no published index exists for these coins, this is an *internal-consistency* check: it asks whether the sparse liquidation-borne implied volatilities reproduce the full-surface model-free index, both built with the same kernel projection, rather than validating against an external benchmark. The external check is the first stage, where the model-free index matches the *published* DVOL at 0.997/0.895. These correlations sit a little below the 0.94/0.96 of the main BTC/ETH sample, as expected from far sparser prints and a one-year window, but they show the liquidation tape alone reproduces the index for coins that have none of their own. The construction is a general recipe rather than a BTC/ETH artifact; assembling the multi-year history the forecasting design requires for such coins awaits a longer liquidation record.

D Count- versus notional-weighted liquidation imbalance

The signed liquidation imbalance used in the body is *notional*-weighted, the net forced-selling share of daily liquidation *value*, $(\text{sell}-\text{buy})/(\text{sell}+\text{buy})$ of notional. A natural alternative weights the two sides by the *count* of liquidation events rather than their dollar size, $(n_{\text{sell}} - n_{\text{buy}})/(n_{\text{sell}} + n_{\text{buy}})$ with n the number of forced sells and buys. We re-estimate every liquidation-bearing model, the linear HAR-VRP-LIQ and the deep Deep-Liq, with the count-weighted imbalance in place of the notional one, leaving everything else unchanged. Table 13 reports the out-of-sample losses. The two definitions are nearly interchangeable: across assets and horizons they differ by at most 0.02 in MSE and 0.01 in QLIKE, and each variant holds the same Model Confidence Set memberships as its headline counterpart. The count-weighted imbalance is, if anything, marginally the better of the two for the deep model at the monthly horizon (BTC MSE $0.42 \to 0.41$, ETH essentially unchanged), and worse nowhere that matters. The

Table 12: Two-stage validation of the reconstruction over 2024-12 to 2025-12. Stage 1 rebuilds the model-free DVOL from the full option surface and compares it to the published index; Stage 2 reconstructs the index from *liquidations* alone and compares it to the model-free DVOL.

Asset	Comparison	Days	Corr.	Calib. R^2
Stage 1: model-free DVOL (full surface) vs published DVOL				
BTC	model-free vs published	364	0.997	0.995
ETH	model-free vs published	355	0.895	0.802
Stage 2: liquidation-only reconstruction vs model-free DVOL				
SOL	liquidations vs model-free (med. 2 opts/day)	356	0.789	0.622
XRP	liquidations vs model-free (med. 1 opt/day)	327	0.899	0.809

Notes. Stage 1: the model-free DVOL of Eq. (12) built from the full option surface, against the published Deribit DVOL (BTC inverse and ETH option chains). Stage 2: the liquidation-only ATM/30-day reconstruction of Eq. (1) from SOL and XRP option-*liquidation* implied volatilities alone, affine-calibrated to and compared with their model-free DVOL. "Days" is the number of overlapping daily observations in the continuous Eq. (1) series; option liquidations actually print on 268 (SOL) and 118 (XRP) of those days, the backward EMA carrying them across the gaps. "med. opts/day" is the median distinct option-liquidation contracts per active day. Stage-2 RMSE is 7.7 (SOL) and 9.3 (XRP) volatility points.

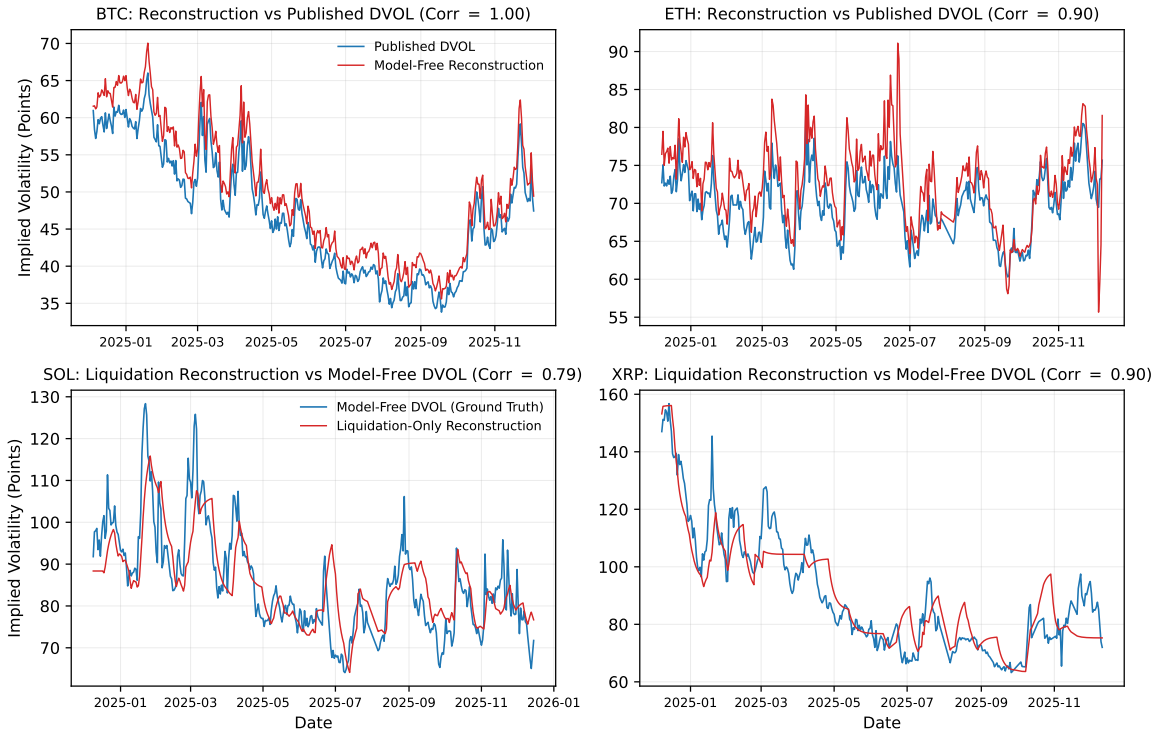


Figure 11: Two-stage validation over 2024-12 to 2025-12. *Top:* for BTC and ETH, the model-free DVOL rebuilt from the full option surface (red) against the published Deribit DVOL (blue). *Bottom:* for SOL and XRP, which have no published index, the liquidation-only reconstruction (red) against the model-free DVOL ground truth (blue). All series are affine-aligned over the window.

paper's conclusion, that the liquidation signal is usable only nonlinearly, is therefore not an artifact of how the imbalance is weighted: the linear HAR-VRP-LIQ overfits under either definition and the deep model's nonlinear encompassing increment survives both. We retain notional weighting in the body because dollar size is the economically meaningful measure of forced flow.

Table 13: Out-of-sample losses under notional- versus count-weighted liquidation imbalance.

			MSE			QLIKE		
Asset	Model	Weighting	H=7	H=14	H=30	H=7	H=14	H=30
BTC	HAR-VRP-LIQ	notional	0.38	0.42	0.46	0.24	0.24	0.24
		count	0.38	0.42	0.46	0.24	0.24	0.24
	Deep-Liq	notional	0.44	0.46	0.42	0.27	0.27	0.22
		count	0.44	0.47	0.41	0.27	0.27	0.21
ETH	HAR-VRP-LIQ	notional	0.37	0.38	0.36	0.24	0.21	0.18
		count	0.36	0.38	0.36	0.24	0.21	0.18
	Deep-Liq	notional	0.46	0.46	0.40	0.29	0.26	0.20
		count	0.46	0.46	0.40	0.29	0.26	0.19

Notes. Out-of-sample MSE of log variance and Jensen-corrected QLIKE, 7-seed means, for the liquidation-bearing models under the two imbalance definitions: notional-weighted (the body) and count-weighted. The HAR-LIQ models behave identically and are omitted for space. The two weightings differ by at most 0.02 (MSE) and 0.01 (QLIKE) and never change a model's 75% MCS membership.

E Penalized linear models do not rescue the liquidation channel

A natural question is whether the linear overfitting of Section 5.1 is an artifact of unpenalized OLS: would a *regularized* linear model extract a usable linear liquidation signal? We re-estimate HAR-VRP-LIQ by ridge, LASSO, and elastic-net regression, standardizing the regressors and selecting the penalty by five-fold cross-validation on each walk-forward training block, otherwise identical to the OLS specification. Table 14 reports the out-of-sample losses. Penalization barely moves the OLS result and never recovers the plain HAR: at the monthly horizon the best penalized HAR-VRP-LIQ carries an MSE of 0.455 (BTC) and 0.362 (ETH), against the HAR's 0.364 and 0.316. The reason is that the cross-validated penalty is mild, because the liquidation regressors lower the in-sample (training-fold) error, so the data-driven shrinkage retains them, and the in-sample signal they carry does not survive out of sample. A linear model thus cannot turn the liquidation tape into out-of-sample forecasting power even with regularization, the linear counterpart of the nonlinear encompassing increment the deep network earns (Section 5.1). QLIKE behaves identically.

We push the linear control further by handing it the network's *entire* daily feature set rather than only the liquidation channel. We form a high-dimensional design from the HAR lags and the premium, all thirteen order-flow features (Section 4.2) aggregated to daily means at $1/7/28$ lags, and their squares and pairwise interactions, 134 regressors in all, and estimate it by OLS and by ridge, LASSO, and elastic-net. Table 15 reports the result. Unregularized, the design overfits catastrophically: its in-sample R^2 reaches 0.84 while its out-of-sample MSE rises one to two orders of magnitude above the HAR's. Regularization contains the blow-up but recovers nothing useful. Ridge, LASSO, and elastic-net stay at or above the plain HAR at the monthly horizon and never reach the deep network, and the sparse estimators retain only six to thirteen of the 134 regressors, discarding the liquidation and interaction terms in favour of

Table 14: Out-of-sample MSE of the liquidation-aware HAR estimated by OLS versus ridge, LASSO, and elastic-net penalization (penalty cross-validated on each training block), against the plain HAR. 7-seed-independent (deterministic) walk-forward, out-of-sample period as in Table 9.

			HAR-VRP-LIQ estimated by				
Asset	H	HAR	OLS	Ridge	LASSO	Elastic-net	Combo
BTC	7	0.366	0.382	0.379	0.379	0.378	**0.348**
	14	0.383	0.422	0.419	0.421	0.418	**0.359**
	30	0.364	0.462	0.458	0.460	0.455	**0.316**
ETH	7	**0.356**	0.366	0.366	0.365	0.365	0.362
	14	**0.355**	0.379	0.379	0.382	0.382	0.371
	30	**0.316**	0.362	0.362	0.375	0.378	0.336

Notes. OOS MSE of log variance. The OLS, ridge, LASSO, and elastic-net columns are all HAR-VRP-LIQ, which adds the log liquidation notional and the signed imbalance at $1/7/28$ lags; the penalized columns standardize the regressors and select the penalty by five-fold cross-validation on each walk-forward training block. HAR is the plain HAR and Combo the headline HAR$\oplus$Deep-Liq combination. The penalized HAR-VRP-LIQ stays close to OLS and well above the plain HAR at every horizon (the liquidation signal is not linearly extractable even under regularization), while the combination is best for BTC and the plain HAR for ETH (**bold**, the per-row minimum).

the realized variance and range signal the HAR already carries. A regularized, interaction-rich linear model with the full feature set is therefore no substitute for the network: the liquidation signal's value lies in its nonlinear use, not in the breadth of the linear design.

Table 15: Out-of-sample MSE of the high-dimensional linear model (all thirteen daily order-flow features at $1/7/28$ lags plus their squares and pairwise interactions, 134 regressors) under ridge, LASSO, and elastic-net, against the plain HAR, the combination, and the deep network.

Asset	Model	H=7	H=14	H=30
BTC	HAR	0.37	0.38	0.36
	HAR-VRP-LIQ	0.38	0.42	0.46
	All-feature,ridge	0.40	0.55	0.87
	All-feature,LASSO	0.36	0.41	0.55
	All-feature,elastic-net	0.36	0.41	0.47
	Deep-Liq (deep net)	0.43	0.46	0.40
	Combo	**0.35**	**0.36**	**0.32**
ETH	HAR	**0.36**	**0.35**	**0.32**
	HAR-VRP-LIQ	0.37	0.38	0.36
	All-feature,ridge	0.46	0.46	0.36
	All-feature,LASSO	0.40	0.42	0.42
	All-feature,elastic-net	0.40	0.44	0.44
	Deep-Liq (deep net)	0.46	0.45	0.40
	Combo	0.36	0.37	0.34

Notes. OOS MSE of log variance. The all-feature design adds, to the HAR lags and the premium, all thirteen order-flow features (Section 4.2) aggregated to daily means at $1/7/28$ lags and their squares and pairwise interactions (134 regressors), estimated by ridge, LASSO, and elastic-net with the penalty cross-validated per training block. Unpenalized OLS (omitted) overfits catastrophically, its OOS MSE one to two orders of magnitude above the HAR's. The penalized estimators retain only six to thirteen of the 134 regressors, stay at or above the plain HAR at the monthly horizon, and never reach the deep network or the combination (**bold**, the per-row minimum: the combination for BTC, the plain HAR for ETH).

References

Aït-Sahalia, Y., Cacho-Diaz, J., & Laeven, R. J. A. (2015). Modeling financial contagion using mutually exciting jump processes. *Journal of Financial Economics*, *117*(3), 585–606.

Alexander, C., & Imeraj, A. (2021). The Bitcoin VIX and its variance risk premium. *The Journal of Alternative Investments*, *23*(4), 84–109.

Andersen, T. G., Bollerslev, T., Diebold, F. X., & Labys, P. (2003). Modeling and forecasting realized volatility. *Econometrica*, *71*(2), 579–625.

Bates, J. M., & Granger, C. W. J. (1969). The combination of forecasts. *Operational Research Quarterly*, *20*(4), 451–468.

Bekaert, G., & Hoerova, M. (2014). The VIX, the variance premium and stock market volatility. *Journal of Econometrics*, *183*(2), 181–192.

Bergsli, L. Ø., Lind, A. F., Molnár, P., & Polasik, M. (2022). Forecasting volatility of Bitcoin. *Research in International Business and Finance*, *59*, 101540.

Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*, *31*(3), 307–327.

Bollerslev, T., Hood, B., Huss, J., & Pedersen, L. H. (2018). Risk everywhere: Modeling and managing volatility. *The Review of Financial Studies*, *31*(7), 2729–2773.

Bollerslev, T., Patton, A. J., & Quaedvlieg, R. (2016). Exploiting the errors: A simple approach for improved volatility forecasting. *Journal of Econometrics*, *192*(1), 1–18.

Bollerslev, T., Tauchen, G., & Zhou, H. (2009). Expected stock returns and variance risk premia. *The Review of Financial Studies*, *22*(11), 4463–4492.

Brunnermeier, M. K., & Pedersen, L. H. (2009). Market liquidity and funding liquidity. *The Review of Financial Studies*, *22*(6), 2201–2238.

Bucci, A. (2020). Realized volatility forecasting with neural networks. *Journal of Financial Econometrics*, *18*(3), 502–531.

Carr, P., & Wu, L. (2009). Variance risk premiums. *The Review of Financial Studies*, *22*(3), 1311–1341.

Chong, Y. Y., & Hendry, D. F. (1986). Econometric evaluation of linear macro-economic models. *The Review of Economic Studies*, *53*(4), 671–690.

Cont, R., Kukanov, A., & Stoikov, S. (2014). The price impact of order book events. *Journal of Financial Econometrics*, *12*(1), 47–88.

Corsi, F. (2009). A simple approximate long-memory model of realized volatility. *Journal of Financial Econometrics*, *7*(2), 174–196.

Demeterfi, K., Derman, E., Kamal, M., & Zou, J. (1999). A guide to volatility and variance swaps. *The Journal of Derivatives*, *6*(4), 9–32.

Deribit. (2026). *Fee schedule and DVOL index methodology.* https://www.deribit.com/kb/fees. (Accessed 2026)

Diebold, F. X., & Mariano, R. S. (1995). Comparing predictive accuracy. *Journal of Business & Economic Statistics*, *13*(3), 253–263.

Du, L., & Shen, J. (2025). Pricing cryptocurrency options with volatility of volatility. *Journal of Futures Markets*, *45*(11), 2066–2091.

Granger, C. W. J., & Ramanathan, R. (1984). Improved methods of combining forecasts. *Journal of Forecasting*, *3*(2), 197–204.

Gu, S., Kelly, B., & Xiu, D. (2020). Empirical asset pricing via machine learning. *The Review of Financial Studies*, *33*(5), 2223–2273.

Hansen, P. R., Lunde, A., & Nason, J. M. (2011). The model confidence set. *Econometrica*, *79*(2), 453–497.

Harvey, D., Leybourne, S., & Newbold, P. (1997). Testing the equality of prediction mean squared errors. *International Journal of Forecasting*, *13*(2), 281–291.

Hoang, L. T., & Baur, D. G. (2020). Forecasting Bitcoin volatility: Evidence from the options market. *Journal of Futures Markets*, *40*(10), 1584–1602.

Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Neural Computation*, *9*(8), 1735–1780.

J.P. Morgan. (1996). *RiskMetrics technical document* (4th ed.; Tech. Rep.). J.P. Morgan/Reuters.

Kyle, A. S. (1985). Continuous auctions and insider trading. *Econometrica*, *53*(6), 1315–1335.

Li, J. (2026). Equilibrium pricing of Bitcoin options with stochastic volatility, jumps, and liquidity risk. *Journal of Futures Markets*, *46*(1).

Newey, W. K., & West, K. D. (1987). A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica*, *55*(3), 703–708.

Parkinson, M. (1980). The extreme value method for estimating the variance of the rate of return. *The Journal of Business*, *53*(1), 61–65.

Patton, A. J. (2011). Volatility forecast comparison using imperfect volatility proxies. *Journal of Econometrics*, *160*(1), 246–256.

Patton, A. J., & Sheppard, K. (2015). Good volatility, bad volatility: Signed jumps and the persistence of volatility. *The Review of Economics and Statistics*, *97*(3), 683–697.

Plíhal, T., & Mampouya, J. O. (2025). Maturity alignment matters: The predictive power of the variance risk premium. *Available at SSRN 6765993*.

Plíhal, T., Mampouya, J. O., Deev, O., & Linnertová, D. V. (2025). Volatility forecasting under the political uncertainty of the second Trump presidency. *Finance Research Letters*, *86*, 108754.

Scaillet, O., Treccani, A., & Trevisan, C. (2020). High-frequency jump analysis of the Bitcoin market. *Journal of Financial Econometrics*, *18*(2), 209–232.

Sirignano, J., & Cont, R. (2019). Universal features of price formation in financial markets: Perspectives from deep learning. *Quantitative Finance*, *19*(9), 1449–1459.

Siu, T. K. (2025). Market consistent valuation for Bitcoin options with long memory in conditional volatility and conditional non-normality. *Journal of Futures Markets*, *45*(8), 917–945.

Timmermann, A. (2006). Forecast combinations. In *Handbook of economic forecasting* (Vol. 1, pp. 135–196). Elsevier.