AIMS Mathematics, 10(11): 27314–27337.
DOI: 10.3934/math.20251201
Received: 19 September 2025
Revised: 09 November 2025
Accepted: 13 November 2025
Published: 24 November 2025

https://www.aimspress.com/journal/Math

Research article

Uniform one-sided conformal bands for forward realized volatility curves

Çağlar Sözen*

Department of Finance and Banking, Görele School of Applied Sciences, Giresun University, Giresun, Turkey

* **Correspondence:** Email: caglar.sozen@giresun.edu.tr;
ORCID: https://orcid.org/0000-0002-3732-5058.

Abstract: We introduce uniform, one-sided conformal prediction bands for forward realized volatility (FRV) paths that control the entire trajectory up to a fixed horizon H with finite-sample, distribution-free marginal validity. The construction is model-agnostic: A monotone (isotonic) baseline across horizons and robust per-horizon scales are fitted to the training data; a scaled sup-norm score calibrated on a chronological holdout yields the uniform envelope. To address serial dependence, we employ block-maxima calibration; for regime sensitivity, we add group conditional (Mondrian) variants based on training-only state variables. On eight liquid assets, the method achieves conservative uniform coverage while adapting the width to tail risk: Bands are widest for crypto and oil, and tightest for broad equities and treasuries. A simple operational law emerges, namely that the mean one-sided width grows approximately with $\sqrt{H}$, turning the horizon design into a transparent safety-tightness trade-off. Practical guidance is given as follows: $\alpha \approx 0.05$ is a reliable default for thinner tails, while $\alpha \in [0.05, 0.025]$ increases safety where bursts are frequent. Relative to parametric benchmarks such as heterogeneous autoregressive realized volatility (HAR-RV) and generalized autoregressive conditional heteroskedasticity (GARCH) models, our bands remain valid across regimes and stay width competitive in calmer markets. The algorithm is linear in nH and agrees with deployment diagnostics. Overall, uniform FRV envelopes provide an operationally transparent, model-agnostic tool for pathwise volatility control, with tunable conservatism and simple extensions for dependence, covariate shift, and cross-split stability.

Keywords: conformal prediction; forward realized volatility (FRV); uniform upper bands; isotonic regression; robust scaling; block-maxima calibration; Mondrian calibration; risk management
Mathematics Subject Classification: 62M10, 62P05

1. Introduction

Volatility is a central object in empirical finance and risk management. Modern practice constructs ex post measures such as realized volatility and deploys forecasting models to anticipate future variation for trading, margining, and stress testing [1–3]. While point or interval forecasts at a single horizon are standard, many applications require pathwise control over a forward horizon: One seeks guarantees that the realized volatility trajectory remains below a safe envelope for all steps up to a prespecified H. This is intrinsically a functional problem: The target is a curve indexed by the horizon. Here, H denotes the forward horizon measured in trading days, and $h = 1, \ldots, H$ indexes the cumulative window length; operationally, margining or stop policies may trigger at any intermediate h, so uniform control over h (rather than at a single endpoint) is the relevant objective.

We propose a simple, model-agnostic solution: Uniform, one-sided prediction bands for forward realized volatility (FRV) curves via split-conformal prediction [4–6]. The bands control the entire future path simultaneously, not just pointwise, and enjoy finite-sample, distribution-free marginal validity under the exchangeability of the calibration and test blocks. By "uniform", we mean a guarantee of the form $\mathbb{P}\{X_\star(h) \le U(h)\ \forall h \le H\} \ge 1 - \alpha$, which is strictly stronger than separate pointwise coverage at each horizon. For consistency with Section 2, we denote this envelope by $\widehat{U}$ when it is constructed from training and calibration data. We henceforth use FRV (forward realized volatility) to refer to the forward path X_t, reserving RV (realized volatility) for single-horizon realized volatility if needed. The construction exploits two domain features. First, the forward RV baseline is nondecreasing in the horizon (it cumulates future squared returns), which we encode with isotonic regression across horizons [7, 8]. Second, empirical returns are heavy-tailed with clustered extremes; robust, training-only per-horizon scales stabilize the conformal scores [9–11]. Calibrating a scaled sup-norm nonconformity score then yields an upper envelope with uniform (pathwise) coverage [5, 6].

Our work connects three strands. (i) In realized volatility modeling, heterogeneous autoregressive (HAR) summaries capture long memory while stochastic volatility (SV) models specify the latent variance dynamics [2, 3]. These typically deliver single-horizon targets and are not distribution-free. (ii) Conformal prediction provides finite-sample coverage for regression and quantile tasks, with refinements such as jackknife+ and covariate shift adjustments [5, 6, 12–14]. Functional variants operating in the sup-norm naturally deliver uniform control across an index set. (iii) Functional data analysis offers shape constraints and inference for curve-valued data [15, 16]; our monotone baseline is a minimal, domain-correct prior consistent with cumulated L^2 construction. This shape prior enforces coherence across h and prevents spurious long horizon dips that would be economically problematic when h serves as a decision clock.

We target an upper envelope aligned with the risk priorities. Monotonicity across horizons reduces variance and prevents economically incoherent long horizon dips [7]. Robust per-horizon scales dampen bursts and jumps [9, 10]. Validity is marginal and finite-sample: Conditional on training-only estimators, calibration and test curves are exchangeable [5]. In serially dependent data, chronological splits and blockwise scoring mitigate dependence in ranks; rolling or cross-split aggregation can further stabilize performance [12, 17]. Economically, uniform control over h supports conservative capital planning: The band width increases with h but, empirically, does so sublinearly, enabling horizon choices that balance safety and tightness.

We (i) formulate FRV path control as a functional prediction task and construct uniform, one-sided

bands with distribution-free, finite-sample validity; (ii) leverage a minimal yet domain-correct shape prior (a monotone baseline across horizons) together with robust, training-only per-horizon scales to stabilize efficiency under heavy tails and volatility clustering; (iii) characterize the efficiency level horizon trade-offs, documenting a near $\sqrt{H}$ law for one-sided widths that enables transparent operational tuning; and (iv) provide a linear time implementation in nH with practical diagnostics (hit/miss paths and cumulative/rolling coverage). Throughout, we set H=30 (approximately six trading weeks) as the default and report sensitivity to H; the near $\sqrt{H}$ law is observed consistently across-assets and regimes. On eight liquid assets—BTC-USD (Bitcoin versus US dollar), ETH-USD (Ether versus US dollar), SPY (S&P 500 ETF), QQQ (Nasdaq-100 ETF), GLD (gold ETF), GDX (gold miners ETF), TLT (20+ year US Treasury bond ETF), and USO (crude oil ETF)—over 2019–2024, the primary specification (monotone baseline + robust scaling + split-conformal calibration) achieves conservatively valid uniform coverage while adapting the width to the cross-sectional tail behavior and volatility clustering: The envelopes are widest for crypto and oil (USO), intermediate for precious metals/miners (GLD/GDX), and tightest for broad equities/treasuries (SPY/QQQ/TLT). In controlled simulations under SV-AR(1) (a stochastic volatility model with AR(1) log variance) and log-HAR-RV (log heterogeneous autoregressive realized volatility) scaled to the same daily standard deviation, the bands preserve conservative coverage, and the widths are larger under SV-AR(1), consistent with stronger persistence and heavier effective tails. A level sweep corroborates the expected monotonic coverage width trade-off induced by order statistic calibration [5, 6].

2. Methods

We develop uniform one-sided functional conformal bands for FRV curves built from asset returns. For brevity and consistency with the introduction, we refer to the forward path simply as FRV and use FRV to denote $X_t(\cdot)$ throughout. The construction follows the split-conformal template, which provides finite-sample, distribution-free marginal validity via calibration order statistics [4–6]. Concretely, (i) A training block produces training-only shape-constrained location and robust per-horizon scale summaries, (ii) a calibration block yields the conformal quantile, and (iii) the resulting band covers the entire test curve uniformly over the forecast horizon. Throughout, we enforce a leak-safe chronology (strict train $\to$ calibration $\to$ test) and report the empirical pathwise coverage on held-out windows. We refer to the baseline as UIR–SC (monotone upper baseline via isotonic regression + split-conformal calibration with a scaled sup-norm score).

2.1. Functionalization

Let P_t be the close price and define the daily log returns $r_t = \log P_t - \log P_{t-1}$, which are the standard input for realized volatility analyses [1, 2]. Fix $H \in \mathbb{N}$ and define the forward RV curve as

$$X_t(h) \;=\; \Big(\sum_{j=1}^{h} r_{t+j}^2\Big)^{1/2}, \qquad h = 1, \ldots, H, \tag{2.1}$$

which is nondecreasing in h by construction (cumulative L^2 aggregation of future realized variance). Computing (2.1) for all feasible (t, h) is $O(nH)$ using prefix sums of r_t^2 [3]. We do not normalize r_t or X_t; scale invariance is achieved at the score level by dividing the residuals by robust $\widehat{\sigma}(h)$. All computations

are performed on the asset's own trading calendar; horizons count trading days only, and no zero returns are imputed across market closures, so the realized sums remain additive on the business day grid. We henceforth refer to X_t as the FRV curve indexed by h. By convention, we set $X_t(0) = 0$ (the nonrandom baseline at h=0), so the FRV plots begin at zero and the band attaches naturally to the origin.

We partition indices chronologically into training I_{tr}, calibration I_{cal}, and a final test index $\star$. The training curves $\{X_i : i \in I_{\mathrm{tr}}\} \subset \mathbb{R}^H$ feed the estimators $(\widehat{\mu}, \widehat{\sigma})$, while $\{X_i : i \in I_{\mathrm{cal}}\}$ is reserved for calibration [5]; $X_\star$ denotes the held-out test curve. We use $(0.7, 0.15, 0.15)$ or $(0.6, 0.2, 0.2)$ splits; Section 3 documents that the 0.7/0.15/0.15 split yields narrower bands because the larger training block sharpens $(\widehat{\mu}, \widehat{\sigma})$ enough to reduce the calibration scores (despite a slightly coarser quantile grid). All splits and estimators are computed assetwise in the primary specification (no cross-asset pooling), unless explicitly noted in grouped extensions.

2.2. Training-based statistics

Because $X_t(h)$ is cumulative in h, a nondecreasing baseline across horizons is natural. We estimate it via isotonic regression as follows:

$$\widehat{\mu} \in \arg\min_{\mu(1)\le\cdots\le\mu(H)} \sum_{h=1}^{H} \{\overline{X}(h) - \mu(h)\}^2, \quad \overline{X}(h) = |I_{\mathrm{tr}}|^{-1} \sum_{i\in I_{\mathrm{tr}}} X_i(h), \tag{2.2}$$

computed by the pool-adjacent-violators (PAV) algorithm [7, 8]. The shape constraints of this form are classical and stabilize the long horizon tail of the RV curve. Monotonicity is enforced only on the training summaries; the test curve is never projected, ensuring valid uncertainty quantification around the unaltered path. The baselines $\widehat{\mu}$ and scales $\widehat{\sigma}$ are computed separately for each asset; cross-asset pooling is reserved for the optional grouped/Mondrian variants.

Let $R_i(h) = X_i(h) - \widehat{\mu}(h)$. We estimate a strictly positive per-horizon dispersion $\widehat{\sigma}(h)$ using either the median absolute deviation (MAD) or Huber's M-scale,

$$\widehat{\sigma}(h) = \begin{cases} 1.4826\,\mathrm{median}_{i\in I_{\mathrm{tr}}}|R_i(h)|, & \text{(MAD) [10]} \\ \text{Huber-}s, & \text{(Huber) [9]} \end{cases} \tag{2.3}$$

with a small training-only floor (we use 10^{-6} in RV units) for numerical stability [11]. For Huber's M-scale, we use Proposal 2 with the tuning $c = 1.345$, solving $|I_{\mathrm{tr}}|^{-1} \sum_i \psi(R_i(h)/s) = 0$ (equivalently $\frac{1}{|I_{\mathrm{tr}}|} \sum \rho(R_i(h)/s) = b$) with the usual b for Fisher consistency. When $\widehat{\sigma}(h)$ is extremely small on quiet horizons, the positive floor avoids division by zero and prevents undue inflation of q driven by numerical noise.

Discretized curves $X_t \in \mathbb{R}^H$ may be interpreted as functional observations under mesh refinement; the functional data analysis (FDA) literature provides the ambient spaces and regularity language [15, 16]. Our bands operate on the grid and, under mild continuity, conservatively approximate continuous time bands. All reported widths are in raw RV units; a two-sided equivalent width equals $2q$ (see §2.3).

2.3. Scores and bands

We conformalize using a one-sided, scaled sup-norm score tailored to the upper tail control,

$$S(x) = \max_{1\le h\le H} \frac{(x(h) - \widehat{\mu}(h))_+}{\widehat{\sigma}(h)}, \qquad (u)_+ = \max\{u, 0\}. \tag{2.4}$$

Let $S_i = S(X_i)$ for $i \in I_{\text{cal}}$ and $m = |I_{\text{cal}}|$. For a target $\alpha \in (0, 1)$, define

$$k = \lceil (m + 1)(1 - \alpha) \rceil, \quad q = S_{(k)}, \tag{2.5}$$

the k-th order statistic of calibration scores [5]; ties are broken conservatively toward larger values of q.

Unless noted otherwise, we set H=30 and α=0.05, with q defined by (2.5).

The uniform upper band is

$$\widehat{U}(h) = \widehat{\mu}(h) + q\,\widehat{\sigma}(h), \qquad h = 1, \ldots, H. \tag{2.6}$$

Unless stated otherwise, the reported widths are relative to the baseline, i.e.,

$$\text{width} := \frac{1}{H} \sum_{h=1}^{H} (\widehat{U}(h) - \widehat{\mu}(h)),$$

and we also summarize efficiency by the two-sided equivalent $2q$. Throughout, we use q generically for the calibration quantile, so that in the block-maxima setting q coincides with q_B from Proposition 2.1.

Lemma 2.1. *For any $c > 0$, replacing $\{X_i\}$ by $\{cX_i\}$ multiplies $(\widehat{\mu}, \widehat{\sigma}, \widehat{U})$ by c and leaves S_i and q unchanged.*

Both MAD and Huber-s are scale equivariant; thus $R_i \mapsto cR_i$ and $\widehat{\sigma} \mapsto c\widehat{\sigma}$, so S_i is invariant and the order statistic q is unchanged.

The score in (2.4) is also translation equivariant relative to $\widehat{\mu}$, ensuring comparability across-assets and enabling cross-asset coverage/width summaries. We focus on upper one-sided control, since the capital, margins, and risk limits are driven by right tail excursions of forward RV; lower deviations are economically benign.

If a training-only grouping $G : \mathbb{R}^H \to \{1, \ldots, K\}$ (e.g., terciles of recent volatility) is available, one may compute groupwise quantiles q_g and use $\widehat{U}_g(h) = \widehat{\mu}(h) + q_g \widehat{\sigma}(h)$ when $G(X_\star) = g$, preserving finite-sample validity conditional on the group ("Mondrian" conformal prediction) [4, 5]. In our implementation, G uses training-only features (the recent RV level and slope) with fixed tercile thresholds; at the test time, the current features route the curve to a group g, and we report both the global q and q_g.

2.4. Uniform one-sided bands: Assumptions and guarantees

Assumption 2.1. *$\widehat{\mu}, \widehat{\sigma}$ depend only on $\{X_i : i \in I_{\text{tr}}\}$ (i.e., are $\mathcal{F}_{\text{tr}}$-measurable) [5].*

Assumption 2.2. *$\widehat{\sigma}(h) > 0$ for all h (ensured by robust scales and a small floor) [11].*

Assumption 2.3. *$\widehat{\mu}(1) \le \cdots \le \widehat{\mu}(H)$ (isotonic fit) [7, 8].*

Assumption 2.4. *Conditional on $\mathcal{F}_{\text{tr}}$, $\{X_i : i \in I_{\text{cal}}\} \cup \{X_\star\}$ are exchangeable; equivalently, the scores are exchangeable given $\mathcal{F}_{\text{tr}}$ [4, 5].*

Lemma 2.2. *For any $x \in \mathbb{R}^H$, $S(x) \le q$ if and only if $x(h) \le \widehat{\mu}(h) + q\,\widehat{\sigma}(h)$ for all h.*

Lemma 2.3. *Under Assumption 2.4, conditional on $\mathcal{F}_{\text{tr}}$, $\mathbb{P}(S(X_\star) \le S_{(k)} \mid \mathcal{F}_{\text{tr}}) \ge k/(m + 1) \ge 1 - \alpha$.*

Theorem 2.1. *Under Assumptions 2.1–2.4, the band $\widehat{U}$ satisfies $\mathbb{P}\{X_\star(h) \le \widehat{U}(h)\ \forall h\} \ge 1 - \alpha$.*

Proof sketches and additional technical details are collected in Appendix A.3. We report empirical coverage as the proportion of held-out curves that remain below $\widehat{U}$ on all horizons; the widths are summarized by the mean raw width and the two-sided equivalent $2q$. Formally, for a test index set $\mathcal{I}_{\mathrm{te}}$, we have

$$\widehat{\mathrm{cov}} \;:=\; \frac{1}{|\mathcal{I}_{\mathrm{te}}|}\sum_{i\in\mathcal{I}_{\mathrm{te}}} \mathbf{1}\{X_i(h) \le \widehat{U}(h) \;\; \forall h \le H\},$$

and we accompany $\widehat{\mathrm{cov}}$ with 95% Wilson binomial confidence intervals based on the test-block proportion of hits; these can be supplemented by moving–block bootstrap intervals if one wishes to account more explicitly for serial dependence.

To relax exchangeability, calibrate on the block-maxima so the between block dependence vanishes asymptotically while the within block dependence is absorbed by the max [17]. Intuitively, maxima compress within block serial correlation into a single tail statistic; ranks across blocks then behave nearly identically and independently distributed (i.i.d.), enabling order statistic calibration.

Assumption 2.5. *$\{X_t : t \in I_{\mathrm{cal}} \cup \{\star\}\}$ is strictly stationary and β-mixing with coefficients $\beta(\ell) \to 0$. Partition I_{cal} into $m_B = \lfloor |I_{\mathrm{cal}}|/B \rfloor$ consecutive blocks J_b (with the last block truncated if needed) with $B = B_{|I_{\mathrm{cal}}|} \to \infty$, $m_B \to \infty$, and $\beta(B) \to 0$ as $|I_{\mathrm{cal}}| \to \infty$.*

Define $M_b = \max_{i\in J_b} S(X_i)$, let $k_B = \lceil (m_B + 1)(1-\alpha) \rceil$, and set $q_B = M_{(k_B)}$. Form $\widehat{U}_B(h) = \widehat{\mu}(h) + q_B\,\widehat{\sigma}(h)$.

Proposition 2.1. *Under Assumptions 2.1–2.3 and 2.5, conditional on $\mathcal{F}_{\mathrm{tr}}$*

$$\liminf_{|I_{\mathrm{cal}}|\to\infty} \mathbb{P}\{\, X_\star(h) \le \widehat{U}_B(h) \;\; \forall h = 1,\dots,H \,\} \;\ge\; 1-\alpha.$$

Ties in block-maxima are broken conservatively toward larger values of q_B. In finite-samples, we keep the last incomplete block; empirically, the coverage is flat in $B \in [10, 30]$, matching the expected insensitivity of maxima based ranks. We set B=20 to mitigate serial dependence across blocks while keeping a sufficiently fine quantile grid; sensitivity is flat over $B \in \{10, 15, 20, 30\}$.

With $S_{\pm}(x) = \max_h |x(h) - \widehat{\mu}(h)|/\widehat{\sigma}(h)$ and $q_{\pm}$ its calibration quantile, the band $[\widehat{\mu}(h) \pm q_{\pm}\widehat{\sigma}(h)]$ achieves $\ge 1-\alpha$ uniform coverage [6].

2.5. Extensions: Dependence, shift, and stability

Jackknife+ or split averaging reduces the splits' randomness while retaining the guarantees [12].

Let Z_i be training-only state features (e.g., the recent RV level/slope and realized quarticity). Fit a logistic density ratio model on I_{tr} to estimate $\widehat{w}(Z) \propto \pi_{\mathrm{te}}(Z)/\pi_{\mathrm{cal}}(Z)$, freeze $\widehat{w}$, and use the smallest q such that $\sum_{i:\,S_i\le q} \widehat{w}(Z_i) \;\ge\; (1-\alpha)\sum_i \widehat{w}(Z_i)$; clip $\widehat{w} \in [0.1, 10]$ for stability. Approximate validity requires standard overlap (positivity) and bounded-weight conditions [13].

For monitoring, we compute rolling coverages over windows of size w (typically, $w = 100$) on the test stream; this does not alter the guarantees but diagnoses local miscalibration and regime shifts.

HAR-type summaries [3] can aid grouping $G(\cdot)$ or diagnostics; inference remains model-agnostic.

2.6. Algorithm, complexity, and diagnostics

Implementation mirrors [5]: Compute forward curves, fit $(\widehat{\mu}, \widehat{\sigma})$ on the training [7–10], calibrate the scores by (2.4), take the order statistic q, and form (2.6).

(1) Build $X_i(h)$ for all $i \in I_{\text{tr}} \cup I_{\text{cal}}$ and $h = 1{:}H$ using prefix sums.
(2) Fit $\widehat{\mu}$ by isotonic regression using the pool-adjacent-violators (PAV) algorithm on $\overline{X}(h)$; compute residuals $R_i(h)$ and robust scales $\widehat{\sigma}(h)$ with a small positive floor.
(3) Define training-only groups G and/or weights $\widehat{w}(Z)$; freeze them.
(4) Compute scores S_i on I_{cal} (or block max scores M_b), take the (weighted) $(1-\alpha)$ order statistic q (or q_B).
(5) Output $\widehat{U}(h) = \widehat{\mu}(h) + q\,\widehat{\sigma}(h)$ (with q equal to the chosen calibration quantile, e.g., q or q_B) and report the diagnostics.

The forward curve's construction is $O(nH)$, isotonic PAV is $O(H)$, calibration scoring is $O(mH)$—all linear in H [7]. The memory is $O((|I_{\text{tr}}|+|I_{\text{cal}}|)\,H)$ with a streaming option that keeps only the running means and per-horizon robust scale summaries. We report 95% Wilson confidence intervals for coverage in the main tables. Where interval estimates for width are needed, a simple option is a moving–block bootstrap on the test stream.

3. Empirical analysis

We study eight liquid assets spanning cryptocurrencies, broad equities, precious metals, energy, and interest-rate exposure: BTC-USD, ETH-USD, SPY, QQQ, GLD, GDX, USO, and TLT. Daily adjusted closes from 2019 to 2024 are obtained from Yahoo Finance [18] and converted to close-to-close log returns $r_t = \log P_t^{\text{adj}} - \log P_{t-1}^{\text{adj}}$. For each date t, we construct FRV curves $X_t(h) = \left(\sum_{j=1}^{h} r_{t+j}^2\right)^{1/2}$, the square root of forward realized-variance, and evaluate uniform one-sided split-conformal bands $\widehat{U}(h) = \widehat{\mu}(h) + q\,\widehat{\sigma}(h)$ under strict chronology, with block-maxima calibration (baseline B=20; see Proposition 2.1); see (2.5) and (2.6).

3.1. Data characteristics and stylized facts

Table 1 summarizes distributional diagnostics (Jarque–Bera) and short-memory tests (Ljung–Box on returns and squared returns). Returns are negatively skewed (especially for crypto and USO), strongly leptokurtic (most pronounced for BTC/ETH/USO), and exhibit marked volatility clustering in squares, motivating a monotone baseline in h, robust per-horizon scales, and an upper one-sided uniform objective.

Table 1. Descriptive statistics of daily log returns.

		Distribution							JB		Ljung–Box $Q(10)$			
Symbol	**Class**	**N**	**Mean**	**SD**	**Min**	**Max**	**Skew**	**Kurt**	**Stat**	p	**Levels**	p	**Squares**	p
BTC-USD	Crypto	2190	0.0015	0.0343	−0.4647	0.1718	−1.1025	21.25	30841.8	0.0000	21.0	0.0209	48.2	0.0000
ETH-USD	Crypto	2190	0.0014	0.0436	−0.5507	0.2307	−1.1301	18.34	21931.4	0.0000	37.2	0.0001	83.2	0.0000
SPY	Equity	1507	0.0006	0.0125	−0.1159	0.0867	−0.8430	16.19	11104.1	0.0000	222.8	0.0000	2080.2	0.0000
QQQ	Equity	1507	0.0008	0.0154	−0.1276	0.0813	−0.5752	9.49	2727.9	0.0000	138.3	0.0000	1265.6	0.0000
GDX	Commodity	1507	0.0004	0.0241	−0.2591	0.1686	−0.6170	16.19	11019.4	0.0000	12.9	0.2307	987.6	0.0000
GLD	Commodity	1507	0.0005	0.0094	−0.0552	0.0474	−0.3450	5.70	486.7	0.0000	12.0	0.2879	168.6	0.0000
USO	Commodity	1507	0.0000	0.0270	−0.2919	0.1542	−2.0945	25.41	32640.7	0.0000	28.3	0.0016	378.0	0.0000
TLT	Treasury	1507	−0.0001	0.0108	−0.0690	0.0725	0.0944	7.21	1117.2	0.0000	49.7	0.0000	1353.8	0.0000

Figure 1 visualizes the distributional shape (histogram with Gaussian overlay and quantile–quantile (Q–Q) plots), highlighting heavy tails and left skew in crypto and USO.

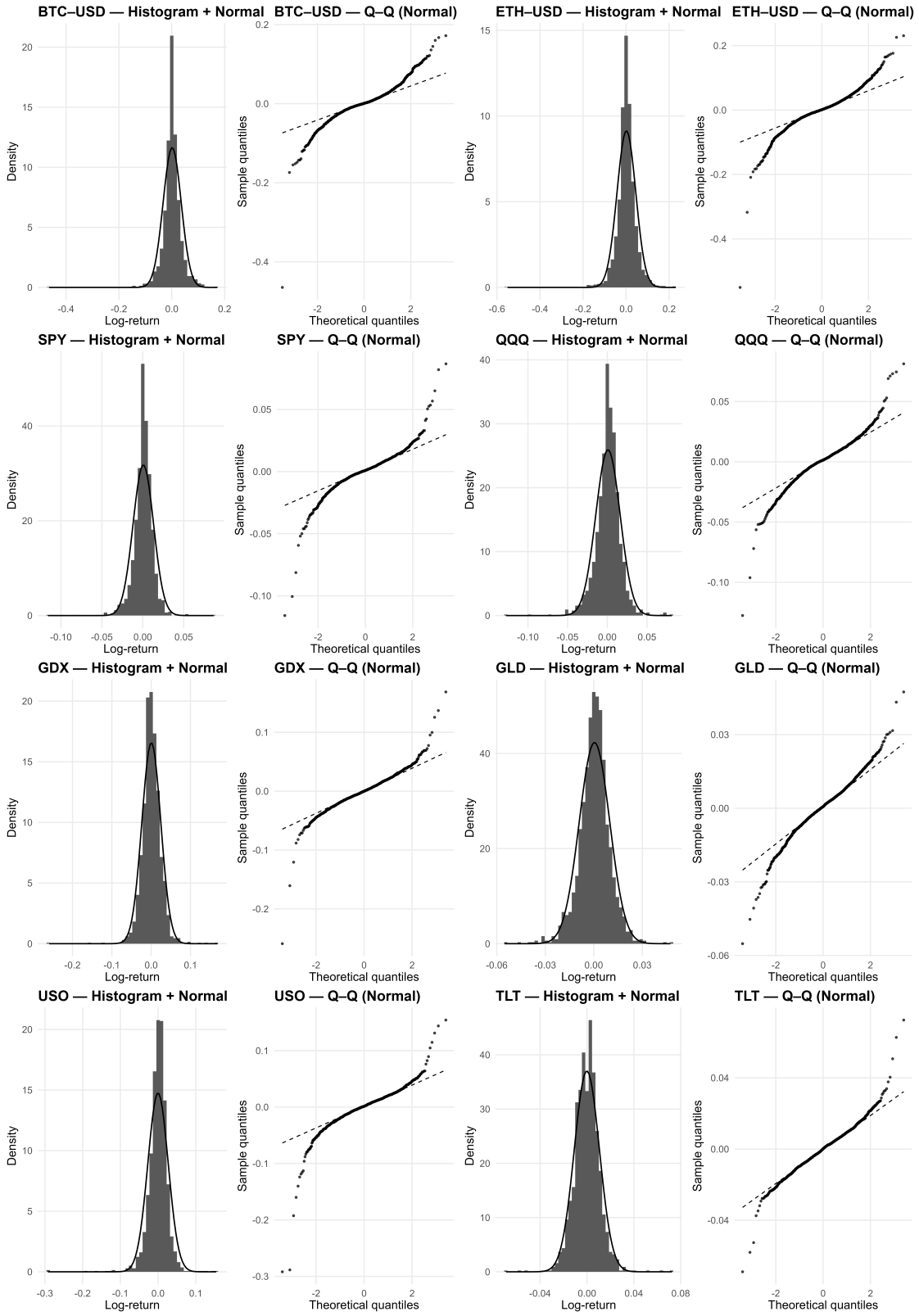


Figure 1. Distributional diagnostics by asset: Histograms with Gaussian overlays and Q–Q panels.

Figure 2 shows autocorrelation functions (ACFs) for levels and squares: Return autocorrelation is modest (GLD/GDX is weakest; SPY/QQQ/TLT and crypto are stronger at short lags), while squared returns are highly persistent, consistent with volatility clustering.

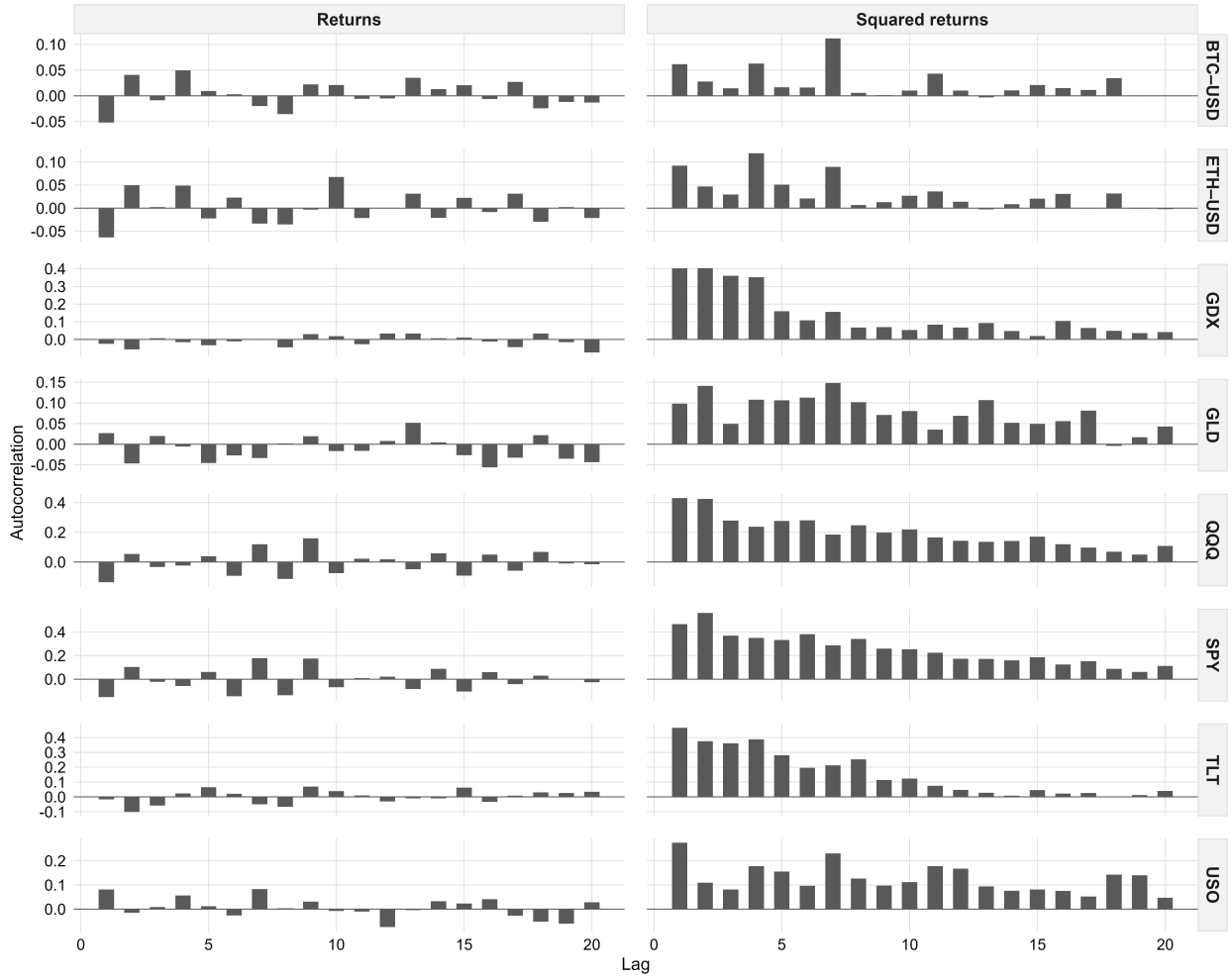
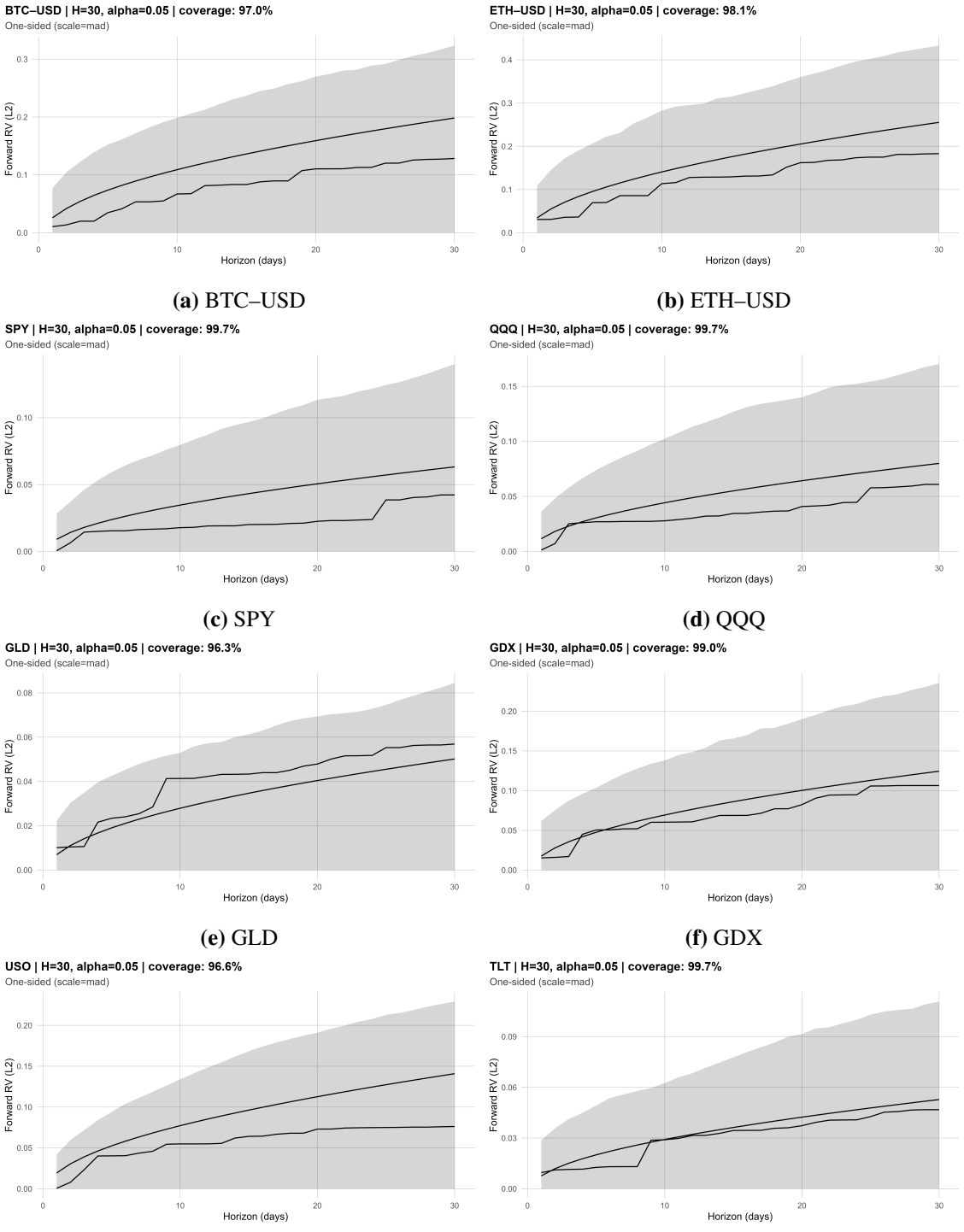


Figure 2. Autocorrelation diagnostics: ACF of returns (left) and squared returns (right).

Heavy tails and clustering jointly endorse (a) the monotone baseline $\widehat{\mu}$, (b) robust per-horizon scales $\widehat{\sigma}$, and (c) an upper one-sided uniform objective.

3.2. Primary empirical specification (upper isotonic regression with scaled conformal score, UIR–SC): H=30, α=0.05

Under the UIR–SC design (monotone $\widehat{\mu}$, robust $\widehat{\sigma}$, sup-score S), out-of-sample uniform coverage is at/above 95% for all assets, near ceiling for SPY/QQQ/TLT; the widths adapt strongly to tail risk. Figure 3 displays the last test FRV envelopes by asset, and Table 2 summarizes the coverage (with Wilson CIs), mean one-sided width, and the calibrated q. The tightest bands appear in GLD/TLT; BTC/ETH/USO remain wider yet conservative. For an audit table with calibrated thresholds and two-sided equivalents, see Appendix A, Table A1; a policy map summarizing coverage width trade-offs is provided in Appendix A.2, Figure A1.



(a) BTC–USD **(b)** ETH–USD

(c) SPY **(d)** QQQ

(e) GLD **(f)** GDX

(g) USO **(h)** TLT

Figure 3. Uniform one-sided FRV bands for the last test curve (H=30, α=0.05).

Table 2. UIR–SC baseline (H=30, α=0.05, split 0.6/0.2/0.2, robust scale = MAD).

Symbol	Coverage (%)	CI low	CI high	Mean width	q
BTC-USD	96.99	94.92	99.23	0.23	2.03
ETH-USD	98.15	96.39	99.06	0.31	2.39
SPY	99.66	98.11	99.94	0.09	2.31
QQQ	99.66	98.11	99.94	0.12	2.33
GLD	96.28	93.47	97.91	0.06	2.62
GDX	98.99	97.06	99.65	0.16	2.93
USO	96.62	93.89	98.15	0.16	1.39
TLT	99.66	98.11	99.94	0.08	3.37

3.3. Robustness across design choices

Table 3 shows a monotonic increase in coverage as α decreases, as implied by order statistic calibration. Figure 4 plots the two-sided equivalent width $2q$ vs. α with the steepest slopes in BTC/ETH/USO, consistent with thicker tails.

Table 4 and Figure 5 indicate near-linearity in $\sqrt{H}$ for mean one-sided width, reflecting the L^2 aggregation geometry; deviations are asset-specific but modest.

Table 3. Coverage by miscoverage level α.

Symbol	α = 0.10 (%)	α = 0.05 (%)	α = 0.025 (%)
BTC-USD	91.20	96.99	99.77
ETH-USD	93.52	98.15	99.54
SPY	98.99	99.66	100.00
QQQ	99.32	99.66	100.00
GLD	91.55	96.28	97.97
GDX	97.30	98.99	99.66
USO	92.91	96.62	99.32
TLT	95.61	99.66	100.00

Table 4. Mean one-sided width by horizon H.

Symbol	H=15	H=30	H=45
BTC-USD	0.17	0.23	0.28
ETH-USD	0.21	0.31	0.37
SPY	0.07	0.09	0.12
QQQ	0.09	0.12	0.14
GLD	0.05	0.06	0.07
GDX	0.12	0.16	0.19
USO	0.11	0.16	0.21
TLT	0.06	0.08	0.09

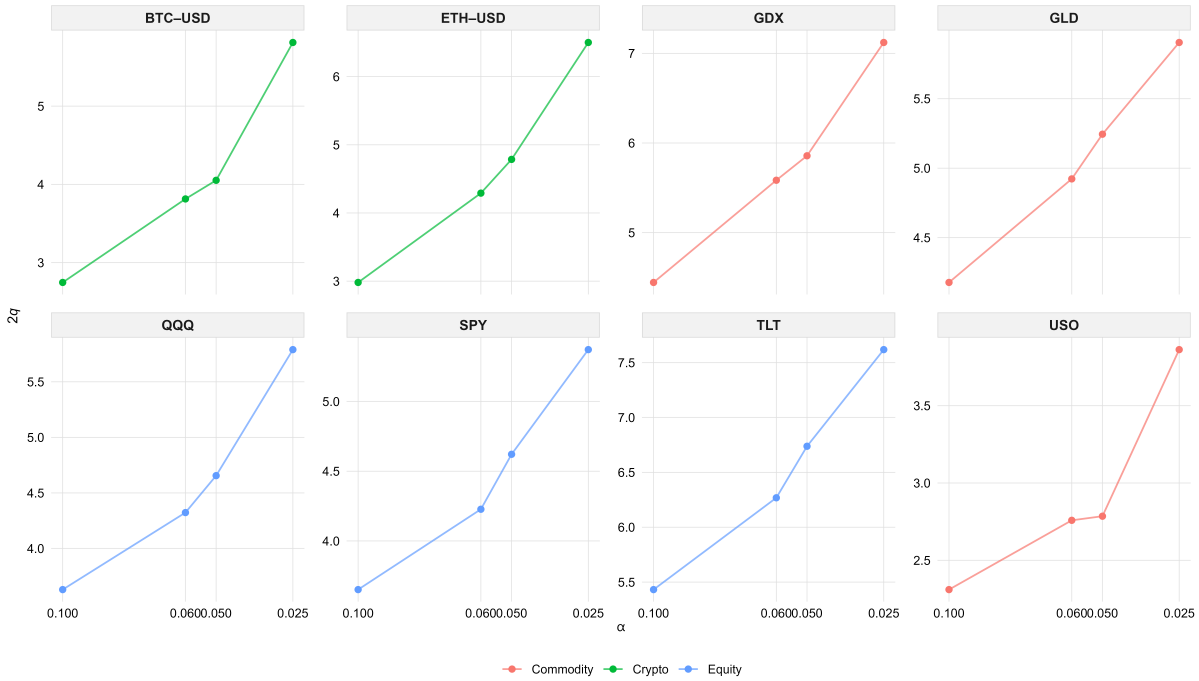


Figure 4. Alpha sensitivity: Two-sided equivalent width $2q$ vs. α.

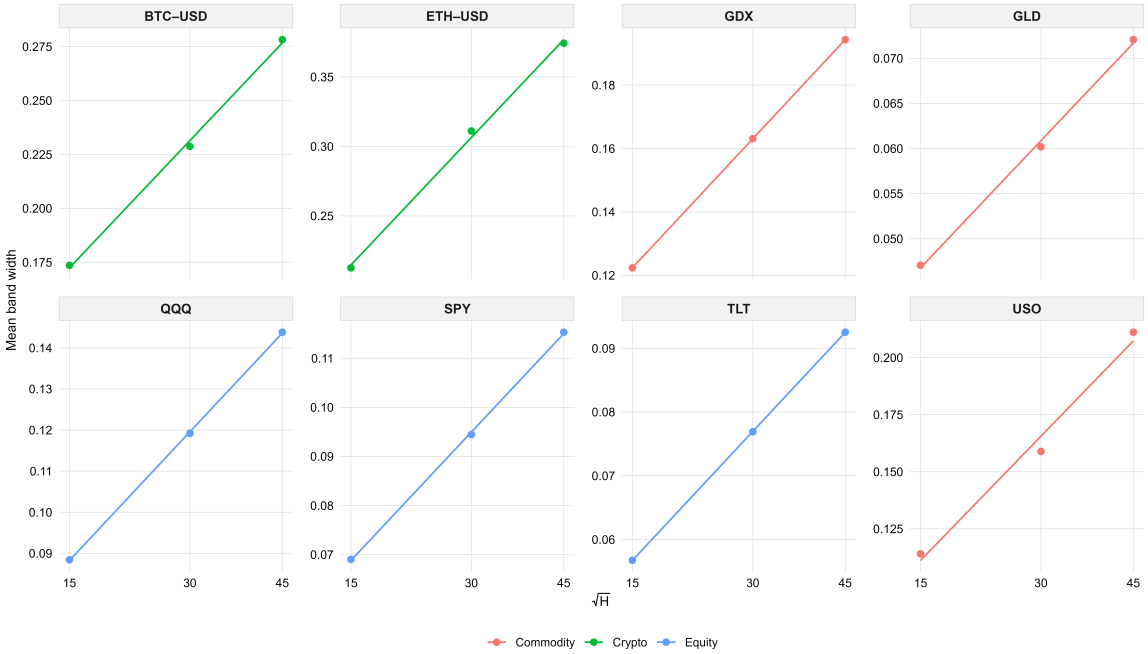


Figure 5. Horizon sensitivity: Mean one-sided width vs. $\sqrt{H}$.

Table 5 and Figure 6 show that larger training shares sharpen $(\widehat{\mu}, \widehat{\sigma})$ and shrink q at a fixed α; equities and treasuries benefit most, with a mild counterexample in USO, reflecting regime shifts.

Table 5. Two-sided equivalent width $2q$ by sample split (training/calibration/testing; dimensionless).

Symbol	0.6 / 0.2 / 0.2	0.7 / 0.15 / 0.15
BTC-USD	4.05	2.80
ETH-USD	4.79	1.82
SPY	4.62	1.58
QQQ	4.66	1.77
GLD	5.24	4.73
GDX	5.86	3.34
USO	2.79	2.99
TLT	6.74	5.18

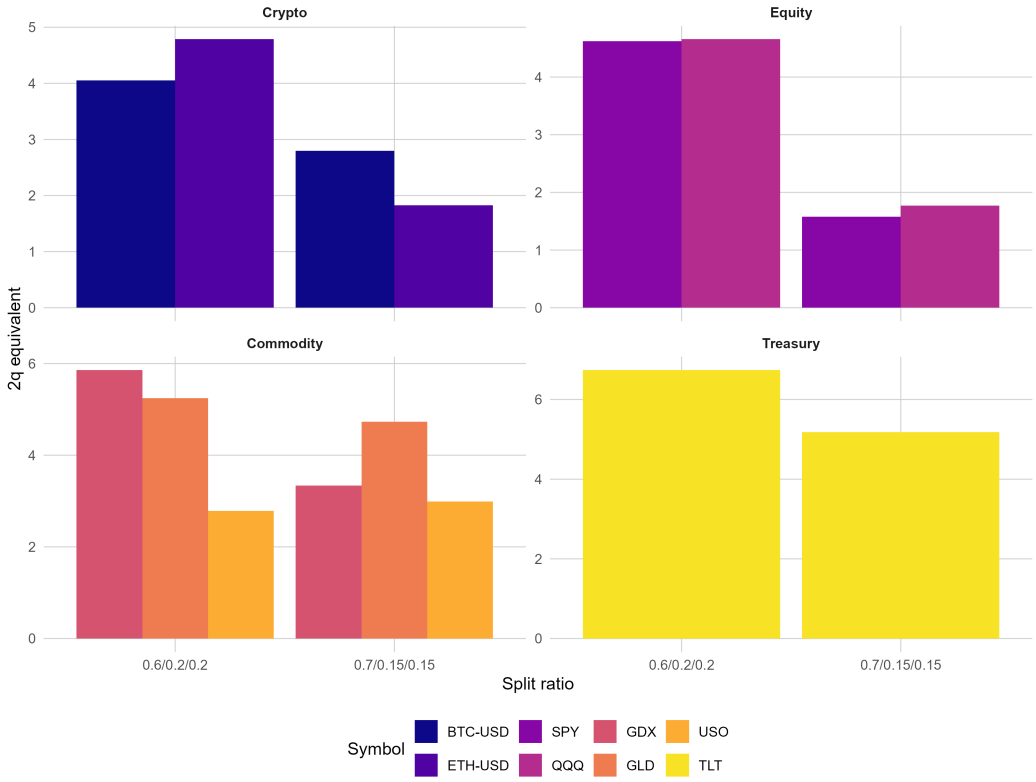


Figure 6. Sensitivity to training–calibration–testing split: Two-sided equivalent width $2q$ across-assets.

Table 6 and Figure 7 jointly show that Huber's s is typically more conservative than MAD (larger $2q$) when the baseline dispersion is small (SPY, QQQ, TLT), reflecting its greater sensitivity to heavier tails.

Table 6. Two-sided equivalent width $2q$ by robust scale estimator.

Symbol	MAD	Huber's s
BTC-USD	4.05	5.02
ETH-USD	4.79	5.33
SPY	4.62	5.86
QQQ	4.66	6.04
GLD	5.24	5.79
GDX	5.86	6.66
USO	2.79	3.79
TLT	6.74	7.22

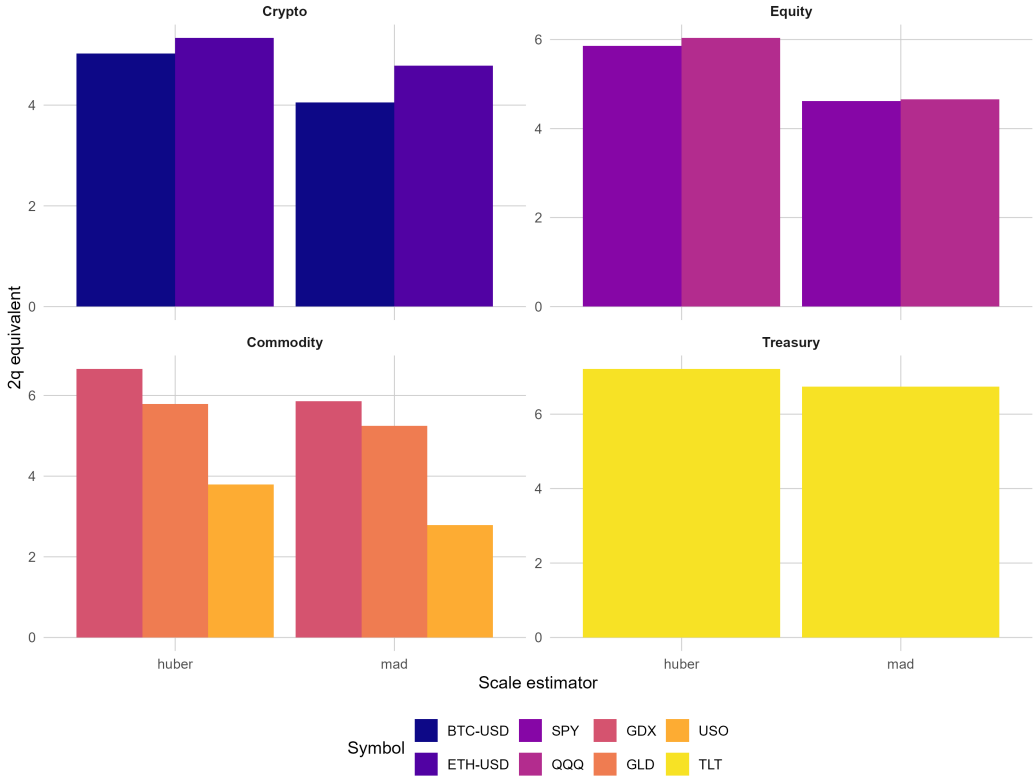


Figure 7. Robust scale comparison (MAD vs. Huber's s): Effective width $2q$.

Figure 8 documents flat realized coverage over $B \in \{10, 15, 20, 30\}$, indicating the insensitivity of rank dependence in this range. Appendix A.2, Figure A1 provides a compact policy map, juxtaposing coverage versus the effective width ($2q$) across α and block choices; it confirms flat B sensitivity and monotonic gains as α decreases, with stronger trade-offs for crypto and USO. Figure 9 shows that union of scores and jackknife+ retain nominal coverage while mildly tightening $2q$, especially for SPY/QQQ/TLT.

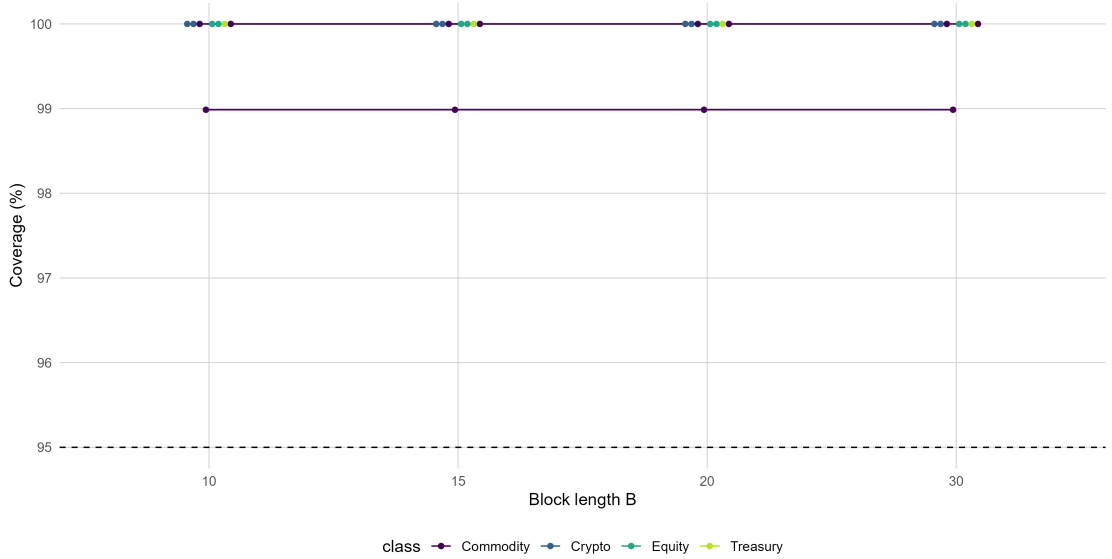


Figure 8. Block length robustness: Realized coverage vs. block size $B \in \{10, 15, 20, 30\}$; coverage is essentially flat across B.

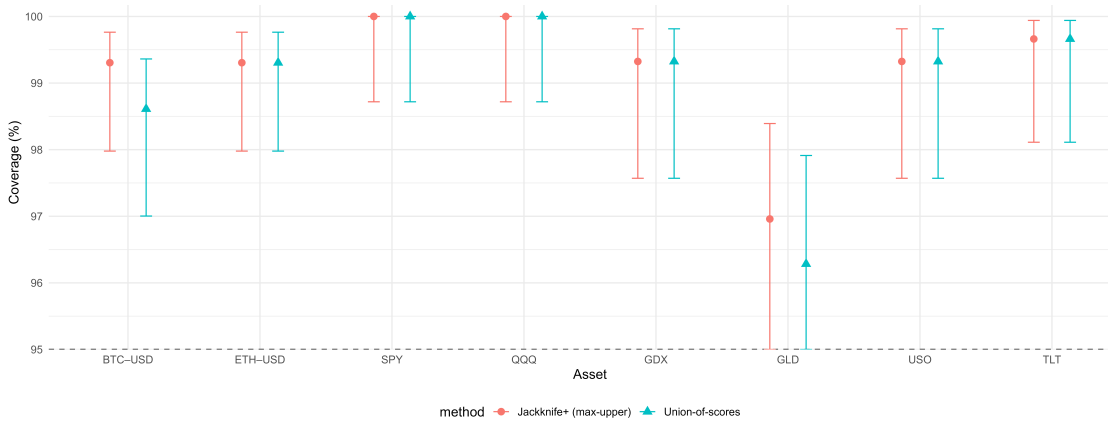


Figure 9. Cross conformal aggregation (union of scores; jackknife+): Coverage with 95% CIs.

3.4. Aggregation and conditional validity

Table 7 reports the test block coverage under union and jackknife+; both stabilize high rank order statistics without sacrificing validity. Figure 10 displays the Mondrian grouping: group-wise q_g rises across low/medium/high volatility terciles, as expected under conditional regime risk.

Table 7. Cross conformal coverage (union; jackknife+), test block.

Symbol	Union cov (%)	Jackknife+ cov (%)
BTC-USD	98.61	99.31
ETH-USD	99.31	99.31
SPY	100.00	100.00
QQQ	100.00	100.00
GLD	96.28	96.96
GDX	99.32	99.32
USO	99.32	99.32
TLT	99.66	99.66

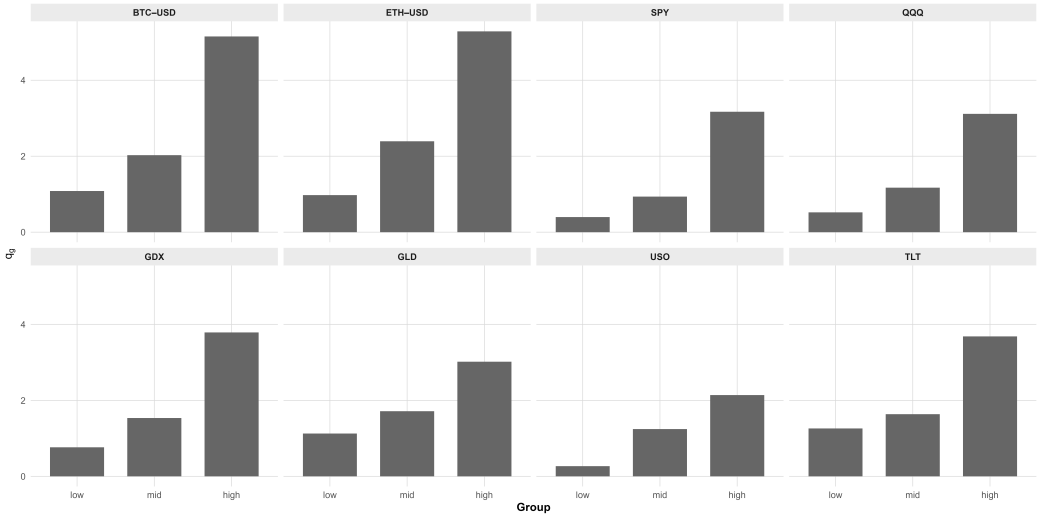


Figure 10. Mondrian grouping: group-wise q_g across low/medium/high volatility regimes.

3.5. Benchmarks and unconditional performance

Table 8 summarizes unconditional parametric benchmarks from the heterogeneous autoregressive realized volatility (HAR-RV) and GARCH(1,1) models at H=30, α=0.05. HAR-RV tends to undercover broad equities (SPY, QQQ), whereas GARCH(1,1) attains near ceiling coverage there with similar or slightly smaller widths. For crypto and commodities, both specifications are conservative, with GARCH(1,1) generally yielding narrower bands. Figure 11 visualizes the coverage width trade-off.

Table 8. Parametric benchmarks at H=30 and α=0.05.

Symbol	Cov HAR (%)	Mean W HAR	2q HAR	Cov GARCH (%)	Mean W GARCH	2q GARCH
BTC-USD	98.61	0.4391	29.2276	96.53	0.2742	4.0723
ETH-USD	98.61	0.5019	25.5504	96.06	0.3356	3.9574
SPY	90.20	0.0835	16.2227	100.00	0.0889	4.3636
QQQ	91.22	0.1146	16.0980	100.00	0.1138	4.1291
GDX	95.27	0.1929	15.7066	98.65	0.1777	4.4267
GLD	94.59	0.0799	13.6832	95.95	0.0721	4.3308
USO	92.91	0.1608	14.9052	98.65	0.1536	3.2725
TLT	93.92	0.0996	13.9013	99.32	0.0789	3.8360

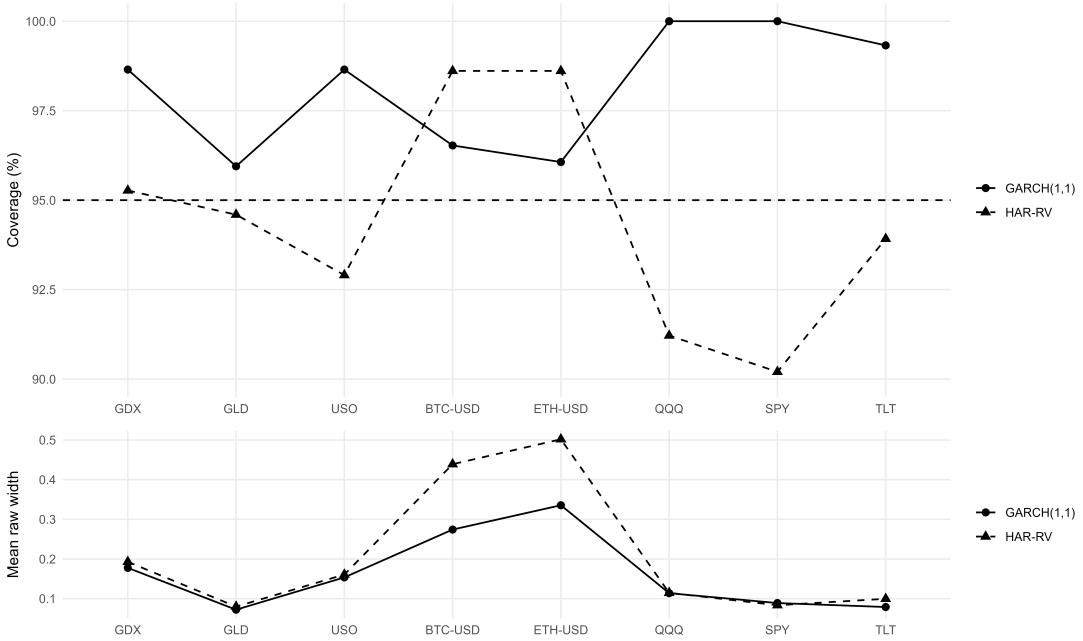


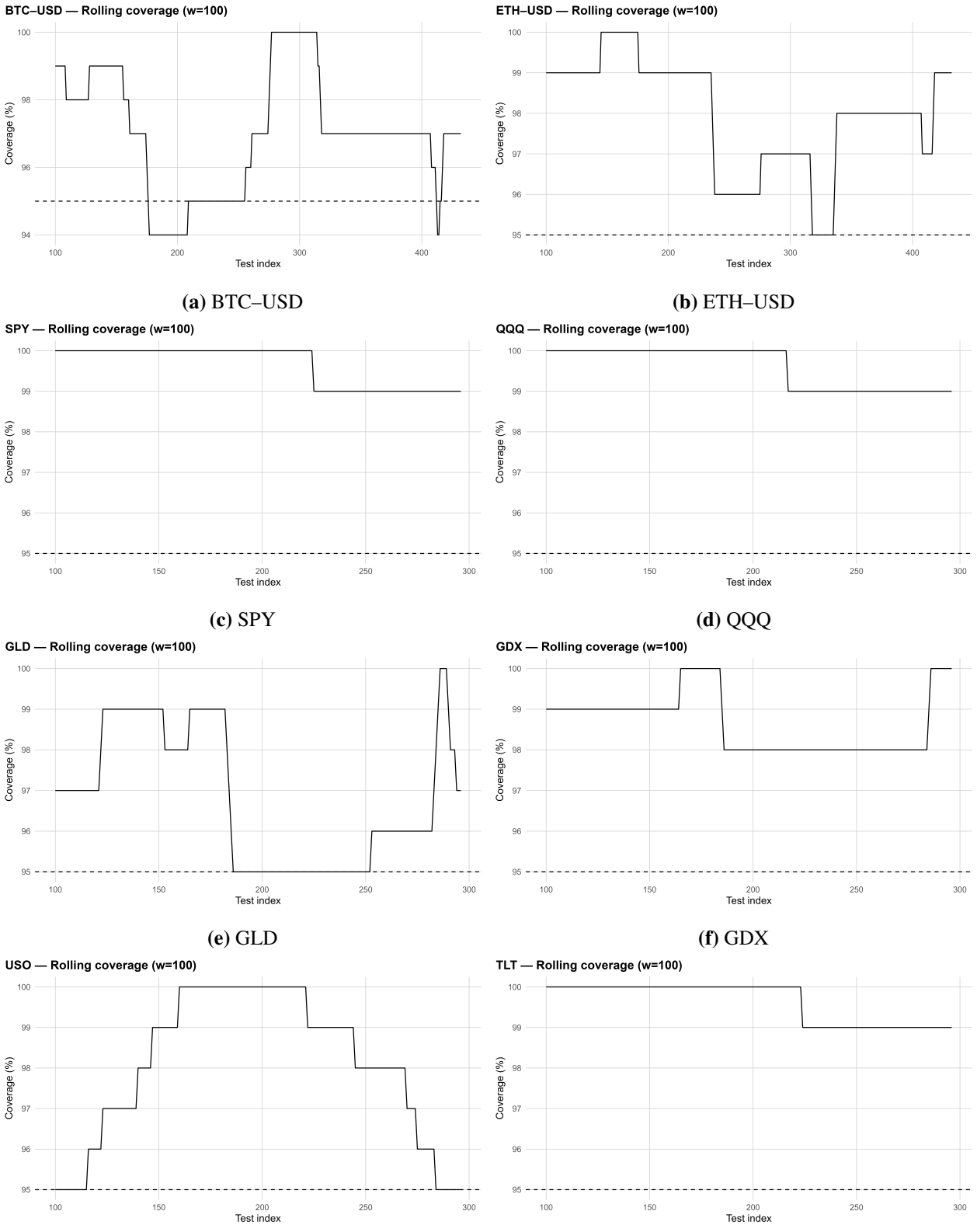
Figure 11. Parametric benchmarks at $H = 30$ and $\alpha = 0.05$: Top: Unconditional coverage (dashed line = nominal 95%); bottom: Mean raw width. Markers: Circles = GARCH(1,1), triangles = HAR-RV.

3.6. Dynamic reliability diagnostics

Figure 12 shows rolling coverage (W=100). SPY/QQQ/TLT are near degenerate at one; BTC/ETH/USO exhibit rare, clustered misses during volatility bursts (consistent with marginal guarantees); GLD dips mid-sample toward nominal levels and recovers thereafter.

3.7. Practical implications and notes about the methods

Doubling H increases the one-sided band width by approximately $\sqrt{2}$; achieving envelopes that are about 50% wider requires H to be roughly 2.25 times larger. For BTC-USD, ETH-USD, and USO, moving from α=0.05 to 0.025 yields meaningful safety gains, whereas for SPY, QQQ, and TLT, α=0.05 already delivers near saturated reliability. We accompany coverage estimates with Wilson confidence intervals; block-maxima scoring handles weak dependence asymptotically (Proposition 2.1); union of scores/jackknife+ and Mondrian conditioning preserve finite-sample distribution-free guarantees while improving stability and conditional reliability.



(a) BTC–USD **(b)** ETH–USD

(c) SPY **(d)** QQQ

(e) GLD **(f)** GDX

(g) USO **(h)** TLT

Figure 12. Rolling coverage (W=100) by asset. Dashed line: Nominal $1 - \alpha = 95\%$.

4. Simulation

We evaluate the finite-sample behavior of the uniform upper band in (2.6), using the block-maxima calibrated quantile q_B of Proposition 2.1, under two standard daily return data generating processes (DGPs):

(i) An SV-AR(1) model with log volatility (stochastic volatility with autoregressive AR(1) log variance)

$$h_t = \mu + \phi\,(h_{t-1} - \mu) + \sigma_\eta \eta_t, \qquad r_t = \exp(h_t/2)\epsilon_t,$$

(ii) A log–HAR–RV where $RV_t = \exp(v_t)$ and

$$v_t = c_0 + \beta_d \log(RV_{t-1}) + \beta_w \log \overline{RV}_{t-1:t-5} + \beta_m \log \overline{RV}_{t-1:t-22} + \sigma_u u_t,$$

and the return is defined as $r_t = \sqrt{RV_t}\,\epsilon_t$ for HAR. Here, $(\epsilon_t, \eta_t, u_t) \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0,1)$. To place both DGPs on a common scale, we target the same unconditional daily standard deviation σ_0=0.015 for r_t. For HAR, this yields

$$c_0 = (1 - \beta_d - \beta_w - \beta_m)\Big\{\log(\sigma_0^2) - \tfrac{1}{2}\sigma_u^2\Big\},$$

which follows the common quasi-autoregressive of order one (AR(1)) approximation for $v_t = \log RV_t$ with a lognormal correction $-\frac{1}{2}\sigma_u^2$; we verified numerically that it yields $\mathbb{E}(r_t^2) \approx \sigma_0^2$. For SV, stationarity with $\mathrm{Var}(h_t) = \sigma_\eta^2/(1-\phi^2)$ and $\mathbb{E}(r_t^2) = \mathbb{E}(\exp(h_t)) = \sigma_0^2$ gives

$$\mu \;=\; \log(\sigma_0^2) \;-\; \tfrac{1}{2}\,\frac{\sigma_\eta^2}{1-\phi^2}.$$

Unless stated otherwise, we set ϕ=0.98 and σ_η=0.20 for the SV model, and $(\beta_d, \beta_w, \beta_m)$=(0.55, 0.30, 0.10) with σ_u=0.20 for the HAR specification. We simulate N=6500 daily returns, discard the first 500, construct FRV curves with horizon H=30, and use a chronological 0.6/0.2/0.2 training/calibration/testing split. Calibration employs non-overlapping block-maxima scores with block length B=20 (ties broken conservatively); results are stable for $B \in [10, 30]$ up to Monte Carlo noise. In this section, the reported q and $2q$ therefore correspond to the block-maxima quantile q_B. Table 9 summarizes the baseline simulation results: coverage is conservatively above the nominal 95% for both DGPs, and the SV-AR(1) bands are wider, reflecting heavier effective tails and stronger persistence at a fixed unconditional scale. Coverage is computed as the fraction of test curves uniformly below $\widehat{U}(h)$ for all $h \le H$. Widths are in FRV units, while $2q$ is a dimensionless two-sided equivalent. Mean one-sided width grows approximately linearly in $\sqrt{H}$, consistent with the L^2 aggregation geometry of FRV.

Table 9. Simulation baseline (H=30, α=0.05; block length B=20).

Model	Coverage (%)	Mean width (one-sided)	q	$2q$
SV-AR(1)	99.7	0.222	9.08	18.20
HAR-RV	98.5	0.109	4.51	9.01

Table 10 summarizes the level sensitivity of the bands. Three patterns emerge. First, conservative uniform validity is maintained across levels: Both DGPs exceed the nominal 95% at α=0.05, with

wider bands under SV-AR(1), again reflecting persistence and heavier effective tails. Second, level monotonicity is clearly visible: Lowering α increases coverage and $2q$ as implied by order statistic calibration [5]. Third, the design is robust: The results are insensitive to block lengths $B \in [10, 30]$, with coverage and $2q$ remaining essentially unchanged across this range.

Table 10. Level sensitivity: Coverage and $2q$ as α varies (H=30, B=20).

	Coverage (%)			**2q**	
Model	**α=0.10**	**α=0.05**	**α=0.025**	**α=0.10**	**α=0.05**
SV-AR(1)	98.9	99.7	100.0	14.60	18.20
HAR-RV	97.1	98.5	99.0	8.28	9.01

5. Conclusions and recommendations

We introduce uniform, one-sided conformal bands for FRV curves that provide pathwise protection over a fixed horizon. The construction is model-agnostic: A monotone (isotonic) baseline across horizons and robust, training-only per-horizon scales yield a scaled sup-norm score whose order statistic on a chronological calibration block delivers a uniform upper envelope. On eight liquid assets (2019–2024), the baseline specification achieves conservative uniform coverage while adapting the width to tail risk—widest for crypto and oil, tightest for broad equities and treasuries. A stable empirical regularity emerges: The mean one-sided width increases nearly linearly with $\sqrt{H}$, turning horizon selection into a transparent safety–tightness trade-off. Controlled simulations under SV-AR(1) and log-HAR-RV at a common unconditional scale corroborate these findings: Both exceed nominal coverage, with wider bands under SV reflecting persistence and heavier effective tails.

(i) For relatively thin-tailed series, $\alpha \approx 0.05$ with $H \in [20, 40]$ is a dependable default; heavier tails warrant either a smaller α or a shorter H. (ii) One-sided width scales approximately as $\sqrt{H}$: Doubling H inflates the width by $\approx \sqrt{2}$; achieving $\sim 50\%$ wider envelopes requires H to be roughly 2.25 times larger. (iii) Larger training shares sharpen $(\widehat{\mu}, \widehat{\sigma})$ and typically reduce the calibrated q at a fixed α, provided that the calibration block still contains tens to low hundreds of curves; split aggregation (union; jackknife+) mitigates split randomness without sacrificing validity. (iv) MAD scaling yields tight, stable envelopes, while Huber's s is more conservative and helpful when quiet horizons produce very small $\widehat{\sigma}(h)$. (v) Block-maxima calibration with moderate block sizes ($B \in [10, 30]$, default B=20) effectively reduces rank dependence and is empirically insensitive within this range. In deployment, tracking hit/miss sequences, cumulative coverage, and rolling coverage offers low friction monitoring and triggers for recalibration. Relative to parametric sets (e.g., HAR/GARCH) that can be narrow in calm regimes yet have low coverage in turbulence, the proposed bands maintain distribution-free validity across regimes while remaining width competitive in quieter markets. Guarantees are marginal and rely on exchangeability across calibration and test blocks; pronounced regime shifts can cause localized undercoverage even when averages meet the target. The calibration grid is discrete—thresholds move in steps of $1/(m+1)$—so mild overcoverage is typical for small m. Our empirical work uses daily close-to-close inputs with cumulative L^2 aggregation; intraday measures could further tighten short horizon dispersion. The main analysis adopts a single chronological split (partly offset by split aggregation) and treats assets independently; joint multi-asset/multi-horizon envelopes are left open and may be more challenging under rough/fractional volatility and long memory [19–21].

Several extensions preserve the distribution-free spirit while improving reliability under shift: Mondrian (group conditional) and importance weighted calibration, split aggregation as a default stability layer, intraday realized measures for short horizons, and non-asymptotic theory under explicit dependence. Multivariate, portfolio-level uniform control is a natural next step. Beyond finance, model-agnostic pathwise safe envelopes can support operational risk management in energy and computing, where distribution-free protection against extreme excursions is valuable [22–24]. Overall, the method is simple, fast, and operationally transparent: Its near $\sqrt{H}$ scaling, tunable conservatism via α, and compatibility with grouping and split aggregation make uniform, one-sided FRV bands a practical tool for pathwise volatility control in production settings.

Use of Generative-AI tools declaration

The author used a generative AI tool only for language polishing.

Conflict of interest

The author declares no conflict of interest in this paper.

References

1. O. E. Barndorff-Nielsen, N. Shephard, Econometric analysis of realized volatility and its use in estimating stochastic volatility models, *J. R. Stat. Soc. B*, **64** (2002), 253–280. https://doi.org/10.1111/1467-9868.00336
2. T. G. Andersen, T. Bollerslev, F. X. Diebold, P. Labys, Modeling and forecasting realized volatility, *Econometrica*, **71** (2003), 579–625. https://doi.org/10.1111/1468-0262.00418
3. F. Corsi, A simple approximate long-memory model of realized volatility, *J. Financ. Economet.*, **7** (2009), 174–196. https://doi.org/10.1093/jjfinec/nbp001
4. V. Vovk, A. Gammerman, G. Shafer, *Algorithmic learning in a random world*, Boston, MA: Springer, 2005. https://doi.org/10.1007/b106715
5. J. Lei, M. G'Sell, A. Rinaldo, R. J. Tibshirani, L. Wasserman, Distribution-free predictive inference for regression, *J. Am. Stat. Assoc.*, **113** (2018), 1094–1111. https://doi.org/10.1080/01621459.2017.1307116
6. A. N. Angelopoulos, S. Bates, Conformal prediction: A gentle introduction, *Found. Trends Mach. Le.*, **16** (2023), 494–591. https://doi.org/10.1561/2200000101
7. R. E. Barlow, D. J. Bartholomew, J. M. Bremner, H. D. Brunk, *Statistical inference under order restrictions: The theory and application of isotonic regression*, London and New York: John Wiley & Sons, 1972.
8. T. Robertson, F. T. Wright, R. L. Dykstra, *Order restricted statistical inference*, Chichester: John Wiley & Sons, 1988.
9. P. J. Huber, Robust estimation of a location parameter, In: *Breakthroughs in statistics*, New York, NY: Springer, 1992. https://doi.org/10.1007/978-1-4612-4380-9_35

10. P. J. Rousseeuw, C. Croux, Alternatives to the median absolute deviation, *J. Am. Stat. Assoc.*, **88** (1993), 1273–1283. https://doi.org/10.1080/01621459.1993.10476408

11. R. A. Maronna, R. D. Martin, V. J. Yohai, M. Salibián-Barrera, *Robust statistics: Theory and methods (with R)*, 2Eds, Chichester: John Wiley & Sons, 2019.

12. R. F. Barber, E. J. Candès, A. Ramdas, R. J. Tibshirani, Predictive inference with the jackknife+, *Ann. Statist.*, **49** (2021), 486–507. https://doi.org/10.1214/20-AOS1965

13. R. J. Tibshirani, R. F. Barber, E. J. Candès, A. Ramdas, Conformal prediction under covariate shift, 2020, arXiv: 1904.06019. https://doi.org/10.48550/arXiv.1904.06019

14. Y. Romano, E. Patterson, E. J. Candès, Conformalized quantile regression, 2019, arXiv: 1905.03222. https://doi.org/10.48550/arXiv.1905.03222

15. J. O. Ramsay, B. W. Silverman, *Functional data analysis*, New York, NY: Springer. https://doi.org/10.1007/b98888

16. T. Hsing, R. Eubank, *Theoretical foundations of functional data analysis, with an introduction to linear operators*, John Wiley & Sons, 2015. https://doi.org/10.1002/9781118762547

17. C. Xu, Y. Xie, Conformal prediction for time series, *IEEE T. Pattern Anal.*, **45** (2023), 11575–11587. https://doi.org/10.1109/TPAMI.2023.3272339

18. Yahoo Finance, Historical daily adjusted prices for BTC-USD, ETH-USD, SPY, QQQ, GLD, GDX, USO, and TLT (2019–2024), 2025. Available from: https://finance.yahoo.com.

19. W. Xiao, J. Yu, Asymptotic theory for estimating drift parameters in the fractional Vasicek model, *Economet. Theor.*, **35** (2019), 198–231. https://doi.org/10.1017/S0266466618000051

20. W. Xiao, J. Yu, Asymptotic theory for rough fractional Vasicek models, *Econ. Lett.*, **177** (2019), 26–29. https://doi.org/10.1016/j.econlet.2019.01.020

21. X. Wang, W. Xiao, J. Yu, Modeling and forecasting realized volatility with the fractional Ornstein–Uhlenbeck process, *J. Econometrics*, **232** (2023), 389–415. https://doi.org/10.1016/j.jeconom.2021.08.001

22. Z. Du, H. Yin, X. Zhang, H. Hu, T. Liu, M. Hou, Decarbonisation of data centre networks through computing power migration, In: *2025 IEEE 5th International Conference on Computer Communication and Artificial Intelligence (CCAI)*, 2025, 871–876. https://doi.org/10.1109/CCAI65422.2025.11189418

23. S. Giannelos, I. Konstantelos, X. Zhang, G. Strbac, A stochastic optimization model for network expansion planning under exogenous and endogenous uncertainty, *Electr. Pow. Syst. Res.*, **248** (2025), 111894. https://doi.org/10.1016/j.epsr.2025.111894

24. Z. Dong, X. Zhang, L. Zhang, S. Giannelos, G. Strbac, Flexibility enhancement of urban energy systems through coordinated space heating aggregation of numerous buildings, *Appl. Energ.*, **374** (2024), 123971. https://doi.org/10.1016/j.apenergy.2024.123971

A. Appendix

A.1. Supplementary table

Table A1. Uniform one-sided FRV bands: Audit table for the held-out test block (H=30, α=0.05). Two-sided equivalent equals $2q$.

Symbol	Coverage (%)	q	$2q$
BTC-USD	98.61	2.429	4.857
ETH-USD	99.31	2.722	5.444
SPY	99.66	2.311	4.622
QQQ	99.66	2.328	4.656
GDX	98.99	2.929	5.858
GLD	96.28	2.622	5.245
USO	96.96	1.476	2.952
TLT	99.66	3.369	6.739

A.2. Supplementary figure

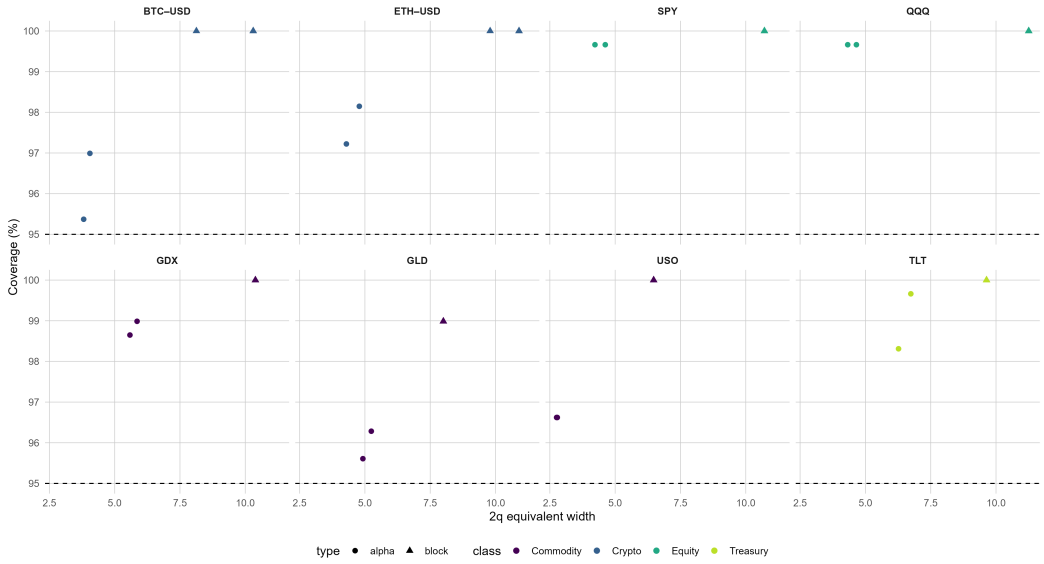


Figure A1. Coverage width trade-off under α and a block length of B.

A.3. Proof sketches and technical details

Let $c>0$. Both the isotonic baseline and the robust per-horizon scales are scale equivariant, so $(\widehat{\mu}, \widehat{\sigma})$ computed from $\{cX_i\}$ satisfy

$$(\widehat{\mu}, \widehat{\sigma}) \;\mapsto\; (c\widehat{\mu}, c\widehat{\sigma}),$$

and the residuals transform as $R_i(h) = X_i(h) - \widehat{\mu}(h) \mapsto cR_i(h)$. Since the score

$$S(x) \;=\; \max_{1\le h\le H} \frac{(x(h) - \widehat{\mu}(h))_+}{\widehat{\sigma}(h)}$$

is homogeneous of degree zero, we have $S(cX_i) = S(X_i)$ for all i. The calibration order statistic $q = S_{(k)}$ is therefore invariant, and the band

$$\widehat{U}(h) = \widehat{\mu}(h) + q\,\widehat{\sigma}(h)$$

scales linearly, $\widehat{U}(h) \mapsto c\widehat{U}(h)$, when the input curves are multiplied by c.

Under Assumptions 2.1–2.4, the calibration and test scores are exchangeable conditional on the training σ-field $\mathcal{F}_{\text{tr}}$. With $m = |I_{\text{cal}}|$ and

$$k = \lceil (m+1)(1-\alpha) \rceil, \qquad q = S_{(k)},$$

exchangeability implies the usual order statistic bound

$$\mathbb{P}(S(X_\star) \le q \mid \mathcal{F}_{\text{tr}}) \;\ge\; \frac{k}{m+1} \;\ge\; 1-\alpha.$$

By Lemma 2.2, $S(x) \le q$ is equivalent to

$$x(h) \;\le\; \widehat{\mu}(h) + q\,\widehat{\sigma}(h) \qquad \text{for all } h \le H.$$

Applying this to $X_\star$ yields

$$\mathbb{P}\{X_\star(h) \le \widehat{U}(h) \text{ for all } h \le H\} \;=\; \mathbb{P}(S(X_\star) \le q) \;\ge\; 1-\alpha,$$

which is the stated uniform (pathwise) coverage guarantee.

AIMS Press

© 2025 the Author(s), licensee AIMS Press. This is an open access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0)