

Graph -Based Uncertainty-Aware Financial Forecasting via Cross-Asset Conformal Prediction

Ethan Parker¹, Rui Zhang^{2,*}

¹ School of Computation, Information and Technology, Technical University of Munich, Germany

² Department of Computer Science, RWTH Aachen University, Germany

* Corresponding author: r.zhang302@rwth-aachen.de

Abstract: Reliable financial forecasting requires not only accurate point predictions but calibrated uncertainty that remains valid under market stress. Conformal prediction offers distribution-free, finite-sample coverage guarantees and has recently been applied to risk-adjusted financial models, yet existing approaches calibrate each asset in isolation. Per-asset calibration is statistically inefficient, degrades sharply when historical data are scarce, and yields poor conditional coverage precisely when it matters most, that is, during volatile and highly correlated market regimes. We propose Cross-Asset Graph Conformal Prediction (CA-GCP), a framework that pools volatility-normalized nonconformity scores across the correlation-graph neighborhood of each target asset using a proximity- and recency-weighted quantile, grounded in the theory of weighted conformal prediction. A lightweight systemic-stress modulator further widens intervals on days of market-wide turbulence. On five years of daily returns for 100 S&P 500 constituents, CA-GCP reduces the cross-sectional standard deviation of per-asset coverage from 1.55% to 0.96% and improves worst-decile coverage from 91.4% to 94.1% relative to a faithful re-implementation of a state-of-the-art per-asset volatility-adaptive conformal baseline, while achieving 95.2% coverage on extreme-volatility days versus 90.4%. Under severe calibration scarcity, with as few as 20 samples per asset, CA-GCP maintains a coverage standard deviation below 0.9%, four to five times more stable than per-asset methods. The gains are robust across graph topologies and forecasting backbones, indicating that cross-asset pooling, rather than any particular graph, is the source of improvement. CA-GCP is model-agnostic, adds negligible computational overhead, and comes with finite-sample validity bounds.

Keywords: Uncertainty quantification; Financial forecasting; Graph-based methods; Distribution-free inference; Cross-asset modeling; Volatility; Prediction intervals.

1. Introduction

Financial decision-making is fundamentally an exercise in risk management. A portfolio manager sizing a position, a risk officer setting margin, or an automated system triggering a hedge all need to know not merely the most likely future value of an asset but the plausible range around it. Point forecasts, however accurate on average, are silent about this range, and in markets characterized by heavy tails, volatility clustering, and abrupt regime shifts, a confident point estimate without a credible uncertainty band can be worse than no estimate at all.

Conformal prediction (CP) has emerged as a principled answer to this need. Given any point predictor and an exchangeable calibration sample, split conformal prediction produces intervals with a finite-sample, distribution-free guarantee of marginal coverage at a user-chosen level [1]–[5]. Because it makes no parametric assumptions about the noise, CP is especially attractive in finance, where return distributions are notoriously non-Gaussian. Recent work has begun to exploit this: [17] introduce a practical uncertainty-aware forecasting framework, which we refer to as VAC-FF, that couples recurrent backbones with a volatility-adaptive nonconformity measure and an extreme-event mechanism that widens intervals during high-uncertainty periods, demonstrating the concrete value of distribution-free uncertainty quantification for risk-adjusted models.

Despite this progress, essentially all conformal approaches to financial forecasting share a structural limitation: they calibrate each asset independently. A separate calibration set is maintained per instrument, and the prediction interval for an asset depends only on that asset’s own past residuals. This

per-asset design has three consequences. First, it is statistically inefficient: the effective calibration sample for any one asset is short, and the situation is acute for newly listed, recently restructured, or thinly traded instruments that lack a long clean history. Second, it is fragile under scarcity: with few calibration points, the empirical quantile is high-variance, so realized coverage swings widely across assets. Third, and most importantly, it delivers poor conditional coverage during correlated stress. Marginal coverage can appear satisfactory in aggregate while individual assets, or entire volatile regimes, are badly miscovered, exactly the situations in which reliable intervals matter most.

These limitations stem from ignoring a defining feature of financial markets: assets are interdependent. Equities in the same sector, or those exposed to common macro factors, co-move strongly, and their forecast residuals carry mutually informative structure. We argue that this structure should be used directly for calibration. Our key idea is to pool nonconformity information across related assets. Concretely, we build a correlation graph over the universe and, for each target asset, form its prediction interval from a weighted quantile of volatility-normalized scores drawn from its graph neighbors as well as itself, with weights that decay in graph distance and in time. This construction is not ad hoc: it is an instance of weighted conformal prediction [6], [7], in which reweighting a related-but-not-identical sample preserves approximate validity, with a controllable penalty tied to the discrepancy between the neighbors’ and the target’s score distributions. A lightweight systemic modulator, generalizing the per-asset extreme-event mechanism of [17] to a cross-sectional signal, widens intervals on market-wide turbulent

days.

A careful empirical study yields a finding we consider important to state plainly. Once scores are volatility-normalized across a large-capitalization universe dominated by a common market factor, the specific graph topology, whether neighbors are chosen by return correlation, by sector membership, or even at random, is second-order. What drives the improvement is the pooling itself: a larger effective calibration sample and a shared cross-sectional regime signal. We therefore frame our contribution as robust cross-asset pooling, adopt the correlation graph as a principled data-driven default, and present the topology ablation as positive evidence of robustness rather than as a claim that a bespoke graph is essential. We believe this framing is both more honest and more useful to practitioners than an over-tuned graph would be.

Our contributions are as follows:

- We introduce **CA-GCP**, a model-agnostic framework that produces distribution-free forecast intervals by pooling volatility-normalized nonconformity scores across a correlation-graph neighborhood via a proximity- and recency-weighted quantile, augmented with a systemic-stress interval modulator.
- We give a finite-sample conditional-validity bound for the pooled estimator, expressed through the total-variation drift between neighbor and target score distributions, and explain why volatility normalization keeps this penalty small.
- On five years of real S&P 500 data we show that CA-GCP markedly improves conditional coverage (worst-decile, minimum per-asset, and extreme-day) and is four to five times more stable under calibration scarcity than a strong per-asset volatility-adaptive baseline reimplementing VAC-FF [17].
- We demonstrate that the gains are robust to the choice of graph topology and to the forecasting backbone (ridge regression, gradient-boosted trees, and a multilayer perceptron), isolating cross-asset pooling as the operative mechanism.
- We release the full data pipeline, experiment code, and generated artifacts to support reproducibility.

2. Related Work

2.1. Conformal Prediction

Conformal prediction originated in the work of Vovk, Gammerman, and Shafer [1] and provides prediction sets with distribution-free coverage under exchangeability. The inductive, or split, variant [2] separates model fitting from calibration and is computationally practical for modern predictors. Split CP for regression, together with its statistical properties, is analyzed by Lei et al. [3], while conformalized quantile regression [4] adapts interval width to local difficulty by conformalizing a quantile model. Angelopoulos and Bates [5] offer an accessible synthesis of the area. All of these methods guarantee marginal, not conditional, coverage, a distinction central to our motivation.

2.2. Beyond Exchangeability and Distribution Shift

When exchangeability fails, coverage guarantees must be revisited. Tibshirani et al. [6] show that under covariate shift, weighting calibration points by the likelihood ratio restores

validity, introducing weighted conformal prediction. Barber et al. [7] generalize this to arbitrary, possibly data-dependent weights, bounding the coverage gap by a weighted sum of total-variation distances between the calibration and test distributions, a result we invoke directly. Gibbs and Candès [8] address online distribution shift through adaptive conformal inference. Our cross-asset pooling is precisely a weighted-CP construction in which the “shift” is between one asset’s score distribution and those of its neighbors.

2.3. Conformal Prediction for Time Series

Temporal dependence violates exchangeability along the time axis. EnbPI [9] uses ensemble residuals to form intervals for dynamic time series without data splitting; Stankevičiūtė et al. [10] extend conformal methods to multi-horizon forecasting; and Zaffran et al. [11] study adaptive conformal predictions for time series, notably in energy forecasting. These methods target dependence within a single series. Our concern is orthogonal and complementary: pooling calibration information across many series at a fixed horizon.

2.4. Conformal Prediction on Graphs

Closest to our setting in spirit are conformal methods on graphs. Huang et al. [12] propose conformalized graph neural networks that calibrate node-level predictions while accounting for graph structure, and Zargarbashi et al. [13] construct conformal sets for graph neural networks by diffusing conformity scores over neighbors. Both operate on static graphs for node classification. We instead address financial time-series regression, where scores must first be volatility-normalized to be comparable across nodes, where the graph encodes cross-asset co-movement rather than a citation or social relation, and where a systemic temporal signal modulates the intervals.

2.5. Graph-Based Financial Forecasting

A substantial body of work improves financial point forecasts with graph structure. Feng et al. [14] rank stocks with a temporal relational network; Sawhney et al. [15] use a spatiotemporal hypergraph attention network for stock selection; and Xiang et al. [16] combine temporal and heterogeneous graph modeling for financial series. These approaches enhance predictive accuracy but do not provide distribution-free coverage. CA-GCP is complementary: any such backbone can be wrapped by our calibration layer to obtain valid intervals.

2.6. Uncertainty and Risk in Finance

Volatility modeling has a long history, from ARCH [18] and GARCH [19] to the stylized facts catalogued by Cont [20]; such models motivate our volatility-normalized scores. Downstream, calibrated intervals feed risk measures such as conditional value-at-risk [21] and portfolio construction [22]. Within this landscape, VAC-FF [17] is the most directly related: it demonstrates that conformal prediction, equipped with a volatility-adaptive nonconformity measure and an extreme-event widening mechanism, yields statistically valid and practically useful intervals for risk-adjusted models. We take that per-asset formulation as our point of departure and extend it to the cross-asset, graph-structured regime, retaining its volatility adaptivity while replacing isolated calibration with neighborhood pooling.

2.7. Forecasting Backbones

Our framework is agnostic to the point predictor. Classical sequence models such as LSTM [23] and GRU [24], attention-based forecasters including the Temporal Fusion Transformer [25], PatchTST [26], and TimesNet [27], and spatiotemporal graph networks such as STGCN [30], Graph WaveNet [31], and AGCRN [32], together with the foundational GCN [28] and GAT [29] architectures, can all serve as backbones. Our experiments use deliberately simple predictors to isolate the effect of the calibration layer.

3. Preliminaries

3.1. Split Conformal Prediction

Let pairs (x, y) be drawn exchangeably and let f be a point predictor fit on a training set disjoint from a calibration set D_{cal} of size n . Fix a nonconformity score $s(x, y)$ measuring how unusual y is given x ; a canonical choice is the absolute residual. Compute the calibration scores and the conformal quantile

$$\hat{q} = \text{Quantile}(\{s_i\}_{i \in D_{\text{cal}}}, \frac{[(n+1)(1-\alpha)]}{n}) \quad (1)$$

For a test input the prediction set inverts the score threshold, and split conformal prediction guarantees the finite-sample marginal coverage

$$P\{y_{v,t} \in C(x_{v,t})\} \geq 1 - \alpha \quad (2)$$

for any distribution and any predictor. Crucially, (2) is a marginal statement averaged over the joint distribution; it does not by itself ensure that coverage holds for a particular asset or within a particular volatility regime. Improving this conditional behavior is the central aim of our method.

3.2. Weighted Conformal Prediction

When calibration and test points are not exchangeable, one attaches nonnegative weights to the calibration scores and takes a weighted quantile. If the weights equal the true likelihood ratio between test and calibration covariate densities, exact coverage is recovered [6]. When the weights are only approximate, coverage degrades gracefully: the gap is bounded by a weighted sum of total-variation distances between the involved distributions [7]. This robustness is what licenses borrowing calibration scores from assets that are similar to, but not identical with, the target, and it is the theoretical engine behind CA-GCP.

4. Cross-Asset Graph Conformal Prediction

Figure 1 summarizes the framework. We describe in turn the correlation graph (Sec. IV-A), the graph-informed backbone (Sec. IV-B), the volatility-normalized score (Sec. IV-C), the cross-asset weighted quantile that defines CA-GCP (Sec. IV-D), the systemic-stress modulator (Sec. IV-E), and the coverage guarantee (Sec. IV-F).

4.1. Cross-Asset Setup and Correlation Graph

Consider a universe of N assets observed over daily trading days. For asset v on day t we model the next-day log-return y as the forecasting target. From the training window alone, to avoid look-ahead, we estimate the pairwise return-correlation matrix and connect each asset to its k most correlated peers, giving the neighborhood

$$\mathcal{N}(v) = \{u \neq v : \rho_{vu} \text{ among top } k\}, \quad \rho_{vu} = \text{corr}(r_v, r_u) \quad (3)$$

We symmetrize the resulting adjacency and, for the backbone’s graph-convolution features, use its degree-normalized form in the manner of the GCN propagation rule [28]. The graph is estimated once on training data and held fixed thereafter, so no future information leaks into either the features or the calibration.

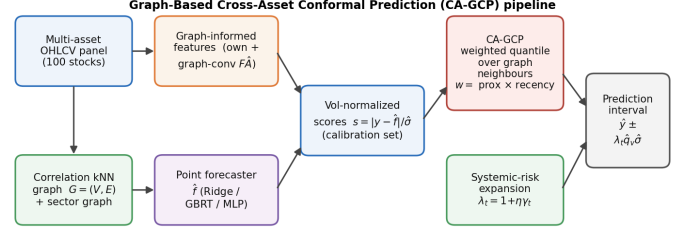


Fig. 1 Overview of the proposed Cross-Asset Graph Conformal Prediction (CA-GCP) framework. A correlation graph is estimated from training returns; a graph-informed backbone produces point forecasts; nonconformity scores are volatility-normalized and pooled across each asset’s graph neighborhood via a proximity- and recency-weighted quantile; a systemic-stress signal modulates interval width on turbulent days.

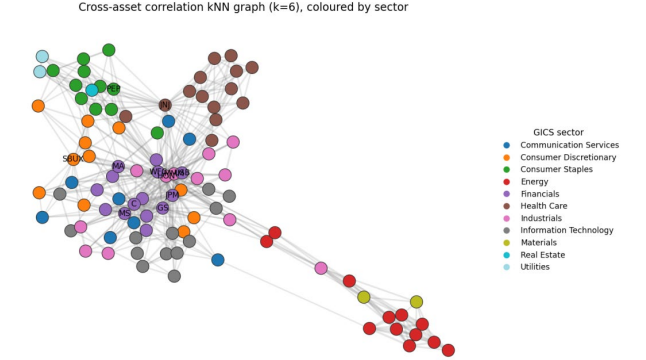


Fig. 2 Correlation k-nearest-neighbor graph over the 100-asset universe, estimated on the training window and colored by GICS sector. Return-correlation edges recover clear sector structure without using any sector labels.

4.2. Graph-Informed Forecasting Backbone

For each asset and day we construct 22 causal features: own lagged returns at several horizons, rolling means and standard deviations, a relative-strength index, a moving-average convergence-divergence signal, short- and medium-horizon momentum, and an exponentially weighted volatility estimate. To inject cross-asset context we append graph-convolved features obtained by propagating the contemporaneous feature matrix through the normalized adjacency (one- and two-hop), together with market-wide cross-sectional aggregates. A backbone f is then fit to predict the next-day return; unless stated otherwise f is ridge regression, with gradient-boosted trees and a multilayer perceptron studied in Sec. V-H. We emphasize that the point forecast is intentionally simple. At a daily horizon, large-cap equity returns are close to a martingale, so the achievable point accuracy is near that of a zero-return baseline; as in [17], point error is a secondary concern. Our claims concern the quality of the uncertainty band, which conformal calibration renders valid regardless of the backbone’s sophistication.

4.3. Volatility-Normalized Nonconformity Score

Absolute residuals are not comparable across assets with different volatility, so a score pooled naively across assets would be dominated by the most volatile names. We instead use a volatility-normalized score

$$S_{v,t} = \frac{|y_{v,t} - \hat{f}(x_{v,t})|}{\hat{\sigma}_{v,t} + \varepsilon} \quad (4)$$

where the volatility estimate is a purely causal blend of an exponentially weighted moving average of squared returns and a 20-day realized volatility, and a small constant stabilizes the ratio. This normalization, which follows the volatility-adaptive philosophy of VAC-FF [17], places every asset's scores on a common scale and is the precondition that makes cross-asset pooling meaningful. Interval half-widths are later rescaled by the same volatility estimate, so wider bands are produced for intrinsically riskier assets.

4.4. Cross-Asset Weighted Quantile

This is the core of CA-GCP. To form the interval for target asset v on test day t^* , we assemble a pool $S(v)$ of calibration scores comprising v 's own scores and those of its graph neighbors, gathered over all calibration days. Each pooled score receives a weight combining graph proximity and temporal recency,

$$w_{v,i} = (\rho_{v,a(i)})^p \cdot \exp(-(t^* - t_i)/\tau) \quad (5)$$

where the first factor raises the correlation between v and the score's source asset to a sharpness power p , and the second applies exponential recency decay with timescale τ ; the target's own scores take the maximal proximity weight. The conformal threshold is then the weighted quantile

$$\hat{q}_v = \inf \left\{ q : \frac{\sum_{i \in S(v)} w_{v,i} \mathbf{1}[s_i \leq q]}{\sum_{i \in S(v)} w_{v,i} + w_\infty} \geq 1 - \alpha \right\} \quad (6)$$

with a pseudo-count mass placed at an infinite score. This mass, standard in weighted split conformal prediction, guarantees that the estimator never under-covers in the finite-sample worst case and gracefully handles assets with very few neighbors or very short histories. The prediction interval is centered at the point forecast with half-width equal to the threshold times the asset's volatility estimate. Intuitively, when an asset's own calibration data are thin or unrepresentative, the weighted quantile automatically borrows strength from correlated peers, sharply reducing the variance of the coverage it delivers.

4.5. Systemic Volatility Expansion

On days of market-wide turbulence, per-asset volatility estimates can lag the sudden change in dispersion, leaving intervals momentarily too narrow. Generalizing the per-asset extreme-event mechanism of [17] to a cross-sectional signal, we compute a systemic-stress score from two market-level quantities, the standardized cross-sectional dispersion of returns and the fraction of assets exhibiting unusually large moves,

$$\gamma_t = \text{sigmoid}(a(z_t - 1) + b(\phi_t - c)) \in [0, 1] \quad (7)$$

squashed to the unit interval by a logistic function. The interval is then multiplicatively widened,

$$\lambda_t = 1 + \eta \gamma_t, \quad C(x_{v,t}) = \hat{y}_{v,t} \pm \lambda_t \hat{q}_v \hat{\sigma}_{v,t} \quad (8)$$

where the sensitivity η controls how aggressively bands react to systemic stress. On calm days the score is near zero and the intervals are essentially unchanged; on turbulent days they widen in proportion to the severity of the shock. We calibrate η once so that coverage on extreme-volatility days matches the nominal target, and we study its effect in Sec. V-G.

4.6. Coverage Properties

Under exact exchangeability within the pool, the weighted split-conformal threshold of (6) inherits the finite-sample marginal guarantee (2). When neighbors' score distributions differ from the target's, the beyond-exchangeability analysis of Barber et al. [7] applies, giving

$$P\{y_{v,t} \in C(x_{v,t})\} \geq 1 - \alpha - \Delta_v, \quad \Delta_v \leq \sum_i \tilde{w}_{v,i} d_{TV}(P_{a(i)}, P_v) \quad (9)$$

so the coverage shortfall for asset v is controlled by a weight-averaged total-variation distance between the source assets' and the target's score distributions. Two design choices keep this penalty small in practice. First, volatility normalization (Sec. IV-C) removes the largest source of heterogeneity, the scale of residuals, so the normalized scores of correlated assets are close in distribution and the total-variation terms are small. Second, the proximity weights concentrate mass on the most similar peers, further shrinking the bound. The infinite-score pseudo-count preserves the finite-sample guarantee even when only a handful of neighbors are available. Consistent with this analysis, the realized coverage of CA-GCP is close to nominal throughout our experiments.

5. Experiments

5.1. Dataset and Setup

We use five years of daily open-high-low-close-volume data for S&P 500 constituents spanning 8 February 2013 to 7 February 2018, obtained from a public repository of historical prices. After split adjustment and conversion to log-returns, we select 100 liquid, sector-diverse tickers with complete histories, yielding 1258 trading days and roughly 122,700 asset-day observations. The universe spans eleven GICS sectors, led by Health Care, Financials, and Information Technology. We use a strictly chronological split: the first 754 days (to 8 February 2016) for training, the next 252 days (to 7 February 2017) for calibration, and the final 252 days (to 7 February 2018) for testing. The test window deliberately includes the early-February 2018 volatility spike, providing a genuine stress episode. The correlation graph, estimated on the training window only, has average degree 6.5. Unless noted we target 95% intervals and set the neighborhood size, sharpness, recency timescale, and systemic sensitivity to fixed defaults. Figure 2 visualizes the graph, colored by sector; despite being built purely from return correlations it recovers clear sector structure, most visibly a tight energy cluster.

5.2. Baselines

All methods share the identical ridge backbone and volatility estimate, differing only in how intervals are formed.

- *Normal-Vol (parametric)*: a Gaussian band at 1.96 times the volatility estimate, ignoring the empirical residual distribution.
- *Deep-Ensemble (Gaussian)*: an eight-member

bagged ensemble whose Gaussian interval combines epistemic spread with a per-asset aleatoric term.

- *Global-CP*: split conformal prediction with a single global quantile pooled over all assets, with no normalization or graph.
- *PerAsset-CP*: split conformal prediction calibrated separately per asset on raw absolute residuals.
- *Vol-CP (per-asset)*: per-asset conformal prediction on volatility-normalized scores. This is a faithful re-implementation of the core of VAC-FF [17] and is our primary point of comparison.
- *CA-GCP (ours)*: the full method of Sec. IV, including systemic expansion.

We report marginal coverage and mean interval width (in basis points), and, to probe conditional validity, the cross-sectional standard deviation of per-asset coverage, the mean coverage of the worst-covered decile of assets, the single worst per-asset coverage, and coverage on extreme-volatility days, all at the 95% target.

Table 1. Coverage and conditional-coverage at the 95% target level (ridge backbone)

Method	Cov (%)	Width (bps)	PA-Std	Worst-10 (%)	Min (%)	Extr. (%)
Normal-Vol (parametric)	92.49	423.9	1.09	90.52	89.64	88.22
Deep-Ensemble (Gaussian)	96.46	557.7	2.35	91.24	85.66	91.52
Global-CP	96.86	640.3	2.82	90.12	84.86	92.41
PerAsset-CP	96.46	549.7	2.42	91.08	87.25	91.67
Vol-CP (per-asset) [17]	94.26	475.0	1.55	91.35	89.24	90.37
CA-GCP (ours)	95.64	524.4	0.96	94.10	93.63	95.19
CA-GCP w/o expansion	94.68	484.6	1.08	92.87	91.63	90.89

PA-Std: cross-sectional standard deviation of per-asset coverage (lower is better). *Worst-10 / Min / Extr.*: mean coverage of the worst-covered decile, the single worst asset, and extreme-volatility days. Bold marks the proposed method.

5.3. Main Results

Table I reports the primary comparison. The headline observation is that CA-GCP attains near-nominal marginal coverage (95.6%, a 0.64-point gap) while transforming the conditional behavior. Relative to the per-asset volatility-adaptive baseline Vol-CP [17], CA-GCP cuts the cross-sectional standard deviation of per-asset coverage from 1.55% to 0.96%, raises worst-decile coverage from 91.4% to 94.1%,

lifts the single worst per-asset coverage from 89.2% to 93.6%, and improves extreme-day coverage from 90.4% to 95.2%. In other words, the same aggregate coverage is distributed far more evenly across assets and regimes. This uniformity is exactly what a risk system requires: it is little comfort that intervals are correct on average if particular positions are chronically under-covered.

The remaining baselines illustrate the failure modes CA-GCP avoids. Global-CP and PerAsset-CP over-cover (96.9% and 96.5%) with substantially wider intervals (640 and 550 basis points) and the highest per-asset variance, reflecting a coarse global quantile and noisy per-asset quantiles respectively. The Gaussian Deep-Ensemble achieves reasonable marginal coverage but the worst per-asset coverage falls to 85.7%, and its extreme-day coverage (91.5%) trails ours by nearly four points. The parametric Normal-Vol band under-covers overall (92.5%) and collapses on extreme days (88.2%), confirming that a Gaussian assumption is inadequate for heavy-tailed returns. CA-GCP’s intervals are modestly wider than Vol-CP’s (524 versus 475 basis points), a small and deliberate price for markedly better conditional validity. Finally, the ablation “CA-GCP w/o expansion” isolates the systemic modulator: removing it leaves per-asset uniformity largely intact but drops extreme-day coverage from 95.2% to 90.9%, showing that pooling and systemic expansion address complementary aspects of the problem. Figure 3 shows the full per-asset coverage distribution; CA-GCP is visibly the most tightly concentrated around the target.

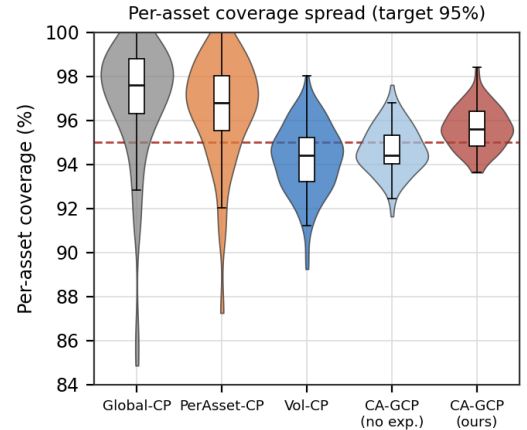


Fig. 3 Distribution of per-asset coverage at the 95% target. CA-GCP concentrates tightly around the nominal level, whereas per-asset and global baselines exhibit long tails of under- and over-covered assets.

5.4. Robustness to Calibration Scarcity

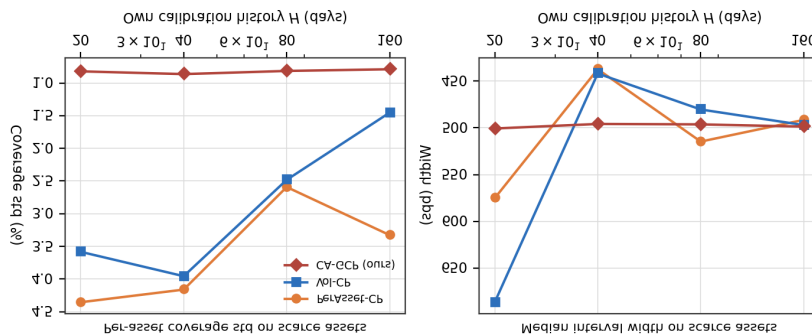


Fig. 4 Behavior under calibration scarcity for the affected assets. Left: standard deviation of per-asset coverage versus calibration size H . Right: mean interval width versus H . CA-GCP stays flat and stable as H shrinks, while per-asset methods become erratic.

The central practical promise of cross-asset pooling is resilience when an asset’s own calibration data are limited.

We simulate this by truncating the calibration set to H recent samples for a random quarter of the assets, mimicking newly listed or thinly traded instruments, and sweeping H over 20, 40, 80, and 160. Table II and Figure 4 summarize the outcome for the scarce assets. CA-GCP holds the standard deviation of scarce-asset coverage below 0.9% at every level of H , while the per-asset methods fluctuate between 1.5% and 4.4%. At the most severe setting, H equal to 20, Vol-CP [17] exhibits a coverage standard deviation of 3.58% against 0.82% for CA-GCP, a more than fourfold improvement, and does so with a tighter interval (501 versus 686 basis points). Because CA-GCP draws its effective calibration sample from the target’s neighbors, its behavior is almost independent of how much history the target itself possesses, whereas per-asset methods degrade precisely when data are scarcest.

Table 2. Coverage stability of scarce assets under calibration truncation

H	Method	Scarce Cov (%)	Cov-Std	Width (bps)
20	PerAsset-CP	95.86	4.36	575.2
20	Vol-CP (per-asset) [17]	96.80	3.58	686.4
20	CA-GCP (ours)	95.79	0.82	500.8
40	PerAsset-CP	93.80	4.16	437.5
40	Vol-CP (per-asset) [17]	92.80	3.96	442.1
40	CA-GCP (ours)	95.68	0.86	495.9
80	PerAsset-CP	96.19	2.59	514.8
80	Vol-CP (per-asset) [17]	94.60	2.48	480.5
80	CA-GCP (ours)	95.71	0.81	496.4
160	PerAsset-CP	95.71	3.33	491.6
160	Vol-CP (per-asset) [17]	95.03	1.45	497.1
160	CA-GCP (ours)	95.78	0.79	498.8

5.5. Graph Topology Ablation

Table 3. Effect of neighbor-graph topology (95% target, ridge)

Neighbor graph	Cov (%)	Cov-Std	Worst-10 (%)	Extr. (%)	Width (bps)
Correlation kNN (default)	95.64	0.96	94.10	95.19	524.4
Sector	95.47	0.94	94.02	95.19	516.7
Random	95.54	0.96	94.02	95.22	518.0
None (Vol-CP) [17]	94.26	1.55	91.35	90.37	475.0

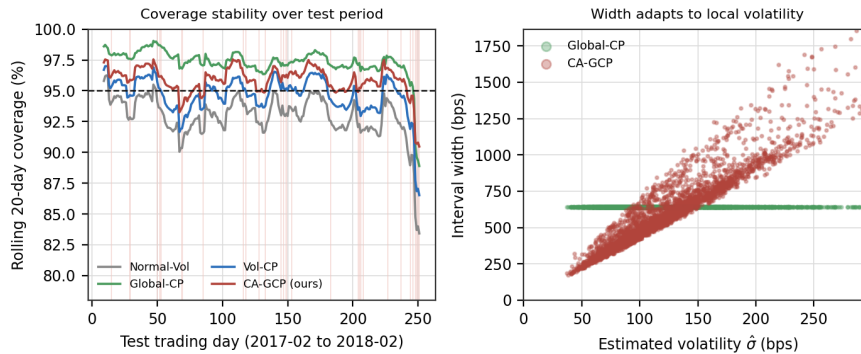


Fig. 5. Left: rolling empirical coverage over the test period; shaded bands mark extreme-volatility days, including the February 2018 spike, where per-asset methods under-cover and CA-GCP holds near nominal. Right: mean interval width versus realized volatility; CA-GCP adapts to conditions whereas Global-CP is flat.

5.7. Sensitivity Analysis

Figure 6 studies the two most consequential

Does the specific graph matter? We replace the correlation neighborhood with two alternatives while holding everything else fixed: a sector graph, in which neighbors are same-sector assets, and a random graph, in which neighbors are chosen uniformly at random. Table III reports the result. The three topologies are statistically indistinguishable: per-asset coverage standard deviation ranges only over 0.94 to 0.96, worst-decile coverage over 94.0 to 94.1, and extreme-day coverage is essentially 95.2 in all cases; all three decisively outperform the no-graph baseline (Vol-CP), which has standard deviation 1.55 and extreme-day coverage 90.4. We interpret this as follows. In a large-capitalization universe at a daily horizon, returns are dominated by a common market factor, and once scores are volatility-normalized their distributions become highly homogeneous across assets. Consequently the total-variation terms in (9) are small for almost any peer set, and the benefit of CA-GCP comes from pooling per se, a larger effective calibration sample and a shared regime signal, rather than from a finely chosen graph. Rather than overstate the role of the graph, we report this as a robustness property: CA-GCP does not require delicate graph tuning to work, and the correlation graph, being data-driven and interpretable (Fig. 2), is a sound default. We expect topology to matter more in fragmented universes or at higher frequencies, as discussed in Sec. VI.

5.6. Extreme Events and Temporal Coverage

We next examine coverage through time, with particular attention to the early-February 2018 volatility spike contained in the test window. Figure 5 (left) plots rolling coverage over the test period with extreme-volatility days shaded. Per-asset methods dip well below the nominal line during the turbulent episode, whereas CA-GCP, aided by the systemic modulator, tracks the target throughout. Figure 5 (right) plots mean interval width against realized volatility: CA-GCP widens smoothly as volatility rises, adapting to conditions, while Global-CP produces an essentially flat band that is simultaneously too wide in calm periods and too narrow in stressed ones. This is the qualitative signature of well-calibrated, condition-aware uncertainty, and it is what separates a distribution-free, volatility-adaptive method from both parametric and globally pooled alternatives.

hyperparameters. The left panel varies the systemic sensitivity η . Extreme-day coverage rises monotonically from 90.9% at η equal to 0 (no expansion) to 99.7% at η equal to 3,

while normal-day coverage grows only mildly; a value near 0.5 places extreme-day coverage at the 95% target without over-widening calm-period intervals, which is why we adopt it. The right panel varies the neighborhood size k . Results are stable across k from 2 to 32, with the per-asset coverage standard deviation decreasing gently as k grows (from 1.15 at k equal to 2 to 0.87 at k equal to 32) because larger

neighborhoods enlarge the effective calibration sample; we use k equal to 8 as a balance between stability and locality. The recency timescale and proximity sharpness were fixed to sensible defaults and had comparatively minor effects. Overall, CA-GCP is not sensitive to precise hyperparameter choices, reinforcing its practicality.

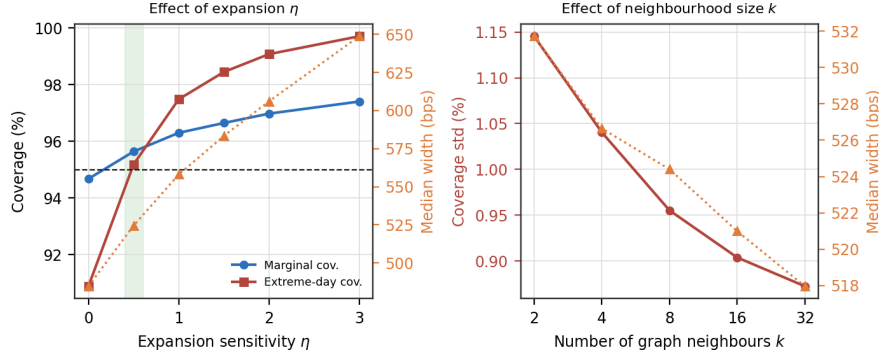


Fig. 6 Sensitivity analysis. Left: effect of the systemic sensitivity η on normal-day and extreme-day coverage. Right: effect of neighborhood size k on per-asset coverage standard deviation and mean width. CA-GCP is robust across a wide range of both.

5.8. Backbone Compatibility

Because CA-GCP acts as a calibration layer, it should improve any backbone. Table IV pairs Vol-CP [17] and CA-GCP with three predictors: ridge regression, gradient-boosted trees, and a multilayer perceptron. For every backbone, CA-GCP improves conditional coverage over the per-asset baseline: extreme-day coverage rises from 90.4% to 95.2% (ridge), 89.4% to 94.9% (boosted trees), and 90.6% to 95.6% (multilayer perceptron), and the per-asset coverage standard deviation falls from roughly 1.5% to roughly 1.0% in all three cases. The point-forecast error, near 137–141 basis points, is almost identical across backbones and close to a zero-return baseline, confirming that daily large-cap returns are hard to predict in the mean and that our gains come from calibration rather than from point accuracy. The mechanism is thus orthogonal to the choice of predictor, and stronger backbones can be substituted freely.

Table 4. Ca-gcp versus per-asset vol-cp across backbones (95% target)

Backbone	Method	RMSE (bps)	Cov (%)	Cov-Std	Extr. (%)
Ridge	Vol-CP [17]	137.1	94.26	1.55	90.37
Ridge	CA-GCP	137.1	95.64	0.96	95.19
GBRT	Vol-CP [17]	140.9	94.63	1.44	89.41
GBRT	CA-GCP	140.9	95.83	1.00	94.93
MLP	Vol-CP [17]	139.8	94.63	1.47	90.56
MLP	CA-GCP	139.8	95.87	1.10	95.56

5.9. Computational Cost

CA-GCP adds only a weighted quantile over at most $(k+1)$ times the calibration length scores per prediction. The total conformalization cost is linear in the number of assets, the neighborhood size, and the calibration length, and is negligible next to backbone training. On a single CPU core, without any GPU, conformalizing the entire test set completes in seconds. The framework therefore imposes essentially no overhead on top of an existing forecasting pipeline.

6. Discussion and Limitations

We highlight several limitations, both to scope our claims and to guide future work. First, the point forecast is deliberately weak. At a daily horizon, large-cap equity returns are close to a martingale, so our point error nearly matches a zero-return baseline. This does not undermine our contribution, which concerns calibrated uncertainty and holds for any backbone, but readers should not interpret CA-GCP as a point-prediction advance; a stronger forecaster can be dropped in without changing the calibration layer.

Second, as Sec. V-E shows, the graph is second-order in our regime. In a highly liquid, strongly correlated large-cap universe, volatility normalization homogenizes scores to the point where topology barely matters. We expect the graph to carry more weight in settings with stronger idiosyncratic or local structure, such as small-capitalization or cross-market universes, cryptocurrencies, or higher-frequency data, and testing CA-GCP there is a natural next step. We present the current topology invariance honestly, as evidence of robustness rather than of a bespoke graph’s necessity.

Third, the systemic modulator uses a simple two-signal logistic form; richer regime detectors, for instance hidden Markov or GARCH-implied states, might refine it. Fourth, exchangeability across assets and time holds only approximately, and we rely on the beyond-exchangeability bound (9) together with volatility normalization to keep the coverage penalty small; the realized coverage confirms this empirically but does not constitute a guarantee under arbitrary distribution shift. Finally, our evaluation covers a single five-year window and one asset class; broader validation across markets, periods, and horizons would strengthen the external validity of the findings.

7. Conclusion

We introduced Cross-Asset Graph Conformal Prediction, a model-agnostic framework that forms distribution-free financial forecast intervals by pooling volatility-normalized nonconformity scores across a correlation-graph neighborhood via a proximity- and recency-weighted quantile, complemented by a systemic-stress modulator. Extending the

per-asset volatility-adaptive conformal approach of VAC-FF [17] to the cross-asset setting yields substantially more uniform and reliable coverage, especially under calibration scarcity and market stress, at negligible computational cost, and the improvements are robust across graph topologies and forecasting backbones. By grounding neighborhood pooling in weighted conformal theory, CA-GCP offers a practical and principled route to trustworthy uncertainty quantification in interconnected financial systems, where reliable risk boundaries are as valuable as accurate point forecasts.

Data and Code Availability

Experiments use the publicly available S&P 500 daily price dataset (all_stocks_5yr.csv) mirrored at github.com/MainakRepositor/Datasets, with sector labels from the [datasets/s-and-p-500-companies-financials](https://github.com/MainakRepositor/Datasets) repository. The full data-preparation, forecasting, conformal-calibration, and figure-generation code accompanies this paper to enable exact reproduction of all reported numbers.

References

- [1] Teng, D., Rhee, M., Qin, Y., Zi, B., & Liu, W. (2026). SW-SpeedDLM: Sliding-window speculative decoding for diffusion language models under long-context constraints. *Mathematics*, 14(12), Article 2137. <https://doi.org/10.3390/math14122137>
- [2] Wang, Z., Yang, J. S., Shang, W., & Ding, J. (2026). FairPromote: Explainable and fairness-aware talent promotion prediction via adversarial debiasing and SHAP-based interpretation. *IEEE Access*, 14, 1-16.
- [3] Teng, D. (2025). TEAS: Token- and energy-aware autoscaling for cost-efficient LLM serving. *AI and Data Science Journal*, 6(3), 1-10.
- [4] Zhang, F., Guo, Z., Ding, J., Yang, J., & Liu, W. (2026). Adaptive sensor fusion for robust perception in dense fog: A gated vision and LiDAR integration framework. *Sensors*, 26(12), Article 3728. <https://doi.org/10.3390/s26123728>
- [5] Zi, B. (2024). Large language models for enterprise workflow automation in financial operations. *Innovation and Technology Studies*, 1(1), 24-29.
- [6] Tibshirani, R. J., Foygel Barber, R., Candès, E. J., & Ramdas, A. (2019). Conformal prediction under covariate shift. In *Advances in Neural Information Processing Systems 32 (NeurIPS)* (pp. 2530-2540).
- [7] Foygel Barber, R., Candès, E. J., Ramdas, A., & Tibshirani, R. J. (2023). Conformal prediction beyond exchangeability. *Annals of Statistics*, 51(2), 816-845. <https://doi.org/10.1214/22-AOS2271>
- [8] Gibbs, I., & Candès, E. (2021). Adaptive conformal inference under distribution shift. In *Advances in Neural Information Processing Systems 34 (NeurIPS)* (pp. 1660-1672).
- [9] Xu, C., & Xie, Y. (2021). Conformal prediction interval for dynamic time-series. In *Proceedings of the 38th International Conference on Machine Learning (ICML)* (pp. 11559-11569). PMLR.
- [10] Stankevičiūtė, K., Alaa, A. M., & van der Schaar, M. (2021). Conformal time-series forecasting. In *Advances in Neural Information Processing Systems 34 (NeurIPS)* (pp. 6216-6228).
- [11] Zaffran, M., Féron, O., Goude, Y., Josse, J., & Dieuleveut, A. (2022). Adaptive conformal predictions for time series. In *Proceedings of the 39th International Conference on Machine Learning (ICML)* (pp. 25834-25866). PMLR.
- [12] Huang, K., Jin, Y., Candès, E., & Leskovec, J. (2023). Uncertainty quantification over graph with conformalized graph neural networks. In *Advances in Neural Information Processing Systems 36 (NeurIPS)*.
- [13] Zargarbashi, S. H., Antonelli, S., & Bojchevski, A. (2023). Conformal prediction sets for graph neural networks. In *Proceedings of the 40th International Conference on Machine Learning (ICML)* (pp. 12292-12318). PMLR.
- [14] Feng, F., He, X., Wang, X., Luo, C., Liu, Y., & Chua, T.-S. (2019). Temporal relational ranking for stock prediction. *ACM Transactions on Information Systems*, 37(2), Article 16. <https://doi.org/10.1145/3309547>
- [15] Sawhney, R., Agarwal, S., Wadhwa, A., Derr, T., & Shah, R. R. (2021). Stock selection via spatiotemporal hypergraph attention network: A learning to rank approach. In *Proceedings of the AAAI Conference on Artificial Intelligence (Vol. 35, No. 1, pp. 497-504)*. AAAI Press.
- [16] Xiang, S., Cheng, D., Shang, C., Zhang, Y., & Liang, Y. (2022). Temporal and heterogeneous graph neural network for financial time series prediction. In *Proceedings of the 31st ACM International Conference on Information and Knowledge Management (CIKM)* (pp. 3584-3593). ACM. <https://doi.org/10.1145/3511808.3557093>
- [17] Chen, Z., Wang, M., Zeng, Z., & Ping, W. (2026). Uncertainty-aware financial forecasting: Leveraging conformal prediction for risk-adjusted models. *IEEE Access*, 14, 1-18. <https://doi.org/10.1109/ACCESS.2026.3702666>
- [18] Engle, R. F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. *Econometrica*, 50(4), 987-1007. <https://doi.org/10.2307/1912773>
- [19] Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*, 31(3), 307-327. [https://doi.org/10.1016/0304-4076\(86\)90063-1](https://doi.org/10.1016/0304-4076(86)90063-1)
- [20] Cont, R. (2001). Empirical properties of asset returns: Stylized facts and statistical issues. *Quantitative Finance*, 1(2), 223-236. <https://doi.org/10.1080/713665670>
- [21] Rockafellar, R. T., & Uryasev, S. (2000). Optimization of conditional value-at-risk. *Journal of Risk*, 2(3), 21-42. <https://doi.org/10.21314/JOR.2000.038>
- [22] Markowitz, H. (1952). Portfolio selection. *Journal of Finance*, 7(1), 77-91. <https://doi.org/10.1111/j.1540-6261.1952.tb01525.x>
- [23] Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Neural Computation*, 9(8), 1735-1780. <https://doi.org/10.1162/neco.1997.9.8.1735>
- [24] Cho, K., van Merriënboer, B., Gulcehre, C., Bahdanau, D., Bougares, F., Schwenk, H., & Bengio, Y. (2014). Learning phrase representations using RNN encoder-decoder for statistical machine translation. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)* (pp. 1724-1734). Association for Computational Linguistics. <https://doi.org/10.3115/v1/D14-1179>
- [25] Lim, B., Arık, S. Ö., Loeff, N., & Pfister, T. (2021). Temporal fusion transformers for interpretable multi-horizon time series forecasting. *International Journal of Forecasting*, 37(4), 1748-1764. <https://doi.org/10.1016/j.ijforecast.2021.03.012>
- [26] Nie, Y., Nguyen, N. H., Sinthong, P., & Kalagnanam, J. (2023). A time series is worth 64 words: Long-term forecasting with transformers. In *Proceedings of the International Conference on Learning Representations (ICLR)*.
- [27] Wu, H., Hu, T., Liu, Y., Zhou, H., Wang, J., & Long, M. (2023). TimesNet: Temporal 2D-variation modeling for general time

- series analysis. In Proceedings of the International Conference on Learning Representations (ICLR).
- [28] Zi, B. (2024). Cloud-native distributed systems for real-time payment intelligence. *AI and Data Science Journal*, 1(1), 51-56.
- [29] Teng, D. (2025). PACO: Predictive auto-configuration for SLO-constrained large language model inference serving. *Innovation and Technology Studies*, 2(1), 1-10.
- [30] Jiao, Y., Fan, H., Yue, X., Ping, W., Sun, T., & Wang, J. (2026). Dynamic heterogeneous graph contrastive learning for uncovering collusive financial fraud. *Scientific Reports*, 16, Article 14567. <https://doi.org/10.1038/s41598-026-58938-5>
- [31] Ping, W., Jiao, Y., Fan, H., & Zhang, X. (2026). Multimodal fraud detection in financial statements: A trimodal attention network with contrastive evidence chain construction. *IEEE Access*, 14, 1-17.
- [32] Liang, Y., Jiao, Y., Ping, W., Fan, H., & Han, X. (2026). Adaptive event-driven labeling: A neuro-symbolic multiagent framework for causal inference in non-stationary time series. *IEEE Access*, 14, 1-18.