Risk Management (2026) 28:59
https://doi.org/10.1057/s41283-026-00243-6

RESEARCH

Reliable value at risk estimation with conformal prediction

Milo Ivancevic[1,2] · Khuong An Nguyen[2] · Zhiyuan Luo[2]

Received: 13 May 2026 / Accepted: 17 August 2026

Abstract

Value-at-Risk (VaR), the most widely used measure of market risk, is typically evaluated through backtesting of point forecasts. Such procedures, however, say little about the uncertainty of the estimated quantile. Existing interval methods are each tied to a specific model class and fail when its underlying assumptions are violated. We propose Quantile Dynamically-Tuned Adaptive Conformal Inference (QDtACI), a model-agnostic conformal calibration layer that constructs finite-sample intervals around *any* VaR forecast, using only the return series and the forecast itself. QDtACI adapts dynamically-tuned adaptive conformal inference to the quantile setting through two components: a pinball-loss nonconformity score aligned with the quantile objective and an asymmetric interval construction, and a multi-speed expert-aggregation mechanism driven by a smoothed violation error and a composite loss on coverage, width, and stability. On synthetic GARCH data, where the true VaR is observable, QDtACI attains near-nominal coverage of the true VaR when the underlying forecast is well-specified, and its coverage degrades in a controlled way as the forecast is misspecified. Against the Delta method, a bootstrap, and the DtACI baseline, it achieves coverage closer to nominal at comparable or better interval quality (Winkler score). Applied to a portfolio of 24 fixed-income assets (2016–2024) with VaR forecasts from CAViaR, DCC-GARCH, and copula models, the intervals are stable in calm periods and widen sharply during stress, including the COVID-19 shock and the 2022–2023 monetary tightening. Because true coverage cannot be measured on real data, we further provide a return-only diagnostic that indicates when interval calibration can be trusted as a proxy for coverage of the true VaR. QDtACI thus offers a single, broadly applicable procedure for uncertainty quantification in VaR, whose reliability tracks the quality of the underlying forecast.

Keywords Value at risk · Conformal prediction · GARCH · Copulas · CAViaR

Extended author information available on the last page of the article

Published online: 02 September 2026

Introduction

Value-at-Risk (VaR) is defined as a quantifiable threshold, indicating the worst expected loss under normal market conditions at a given confidence level (Danielsson 2011). It occupies a central place in financial risk management, particularly in the management of market risk. It is largely endorsed as a standard in reporting practice by authorities for several reasons, among others as it underpins regulatory capital requirements under the Basel framework, informs internal risk limits, and serves as a primary input to margining and collateral decisions (Basel Committee on Banking Supervision 2016; Chen 2014). Because these uses attach real economic consequences to the reported number, e.g. capital charges, position limits, and margin calls all scale with it, the reliability of a VaR estimate is not merely a statistical technicality and academic concern, but a practical necessity.

Yet, VaR is difficult to estimate reliably. It generally involves predicting an unknown parameter of the return distribution, which presents a unique statistical challenge. In the real-world, calculating VaR at extreme confidence levels and on different sample sizes, different volatility environments, and modelling assumptions could introduce unavoidable errors, while small changes in specification could shift the estimated quantile greatly (Christoffersen and Gonçalves 2005). For this reason, relying solely on point estimates could be misleading, as it doesn't convey in any way any of these estimation uncertainties. One figure representing VaR, thus, could falsely project the impression of precision, especially during periods of market stress, precisely when accurate risk assessment matters most, and when point forecasts are most prone to fail. Concequently, a growing literature argues that point VaR should be accompanied by an interval that quantifies the uncertainty around it (He et al. 2020; Beutner et al. 2024). However, constructing such intervals remains a persistent challenge. Providing reliable predictive confidence for an extreme, model-dependent quantile is difficult, and existing approaches are each constrained in ways that limit their applicability.

Existing approaches to interval estimation for VaR fall into a few broad families, each analytically appealing but constrained in ways that limit general use. The classical *Delta method* approximates the variance of the VaR estimator by linearising it as a smooth function of the estimated model parameters (Duan 1994). Under standard conditions and assuming asymptotic normality of the parameter estimators, the Delta method yields analytical expressions for the variance of VaR and thus enables the construction of asymptotic confidence intervals (Casella and Berger 2002; van der Vaart 1998). However, because the Delta method relies on first-order approximations and on the correct specification of the underlying parametric model, its finite-sample performance can be poor. More precisely, for extreme quantiles the VaR is highly nonlinear and sensitive to tail behavior, which can result in substantial undercoverage and unreliable confidence intervals (McNeil et al. 2015; Pritsker 2006). Closely related *asymptotic analytical* intervals, based on the limiting normal distribution of the VaR estimator, share this fragility: their variance depends on the return density evaluated at the VaR threshold (Francq and Zakoian 2004; McNeil et al. 2015). Estimating that density at extreme quantiles is itself error-prone, so even small errors translate into large distortions at high confidence levels or under heavy tails.

A second family relies on *bootstrap resampling*. They seek to approximate the sampling distribution of the VaR estimator through repeated simulation. Bootstrap approaches are attractive because they are relatively easy to implement and do not require explicit analytical derivation of the estimator's variance. In the context of VaR, both parametric bootstraps (resampling from an estimated model) and nonparametric bootstraps (resampling directly from the empirical distribution) have been proposed (Christoffersen and Gonçalves 2004; McNeil et al. 2015). Their validity, however, depends on faithfully reproducing the features of financial data, such as conditional heteroskedasticity, heavy tails, and serial dependence. Where these are not preserved, coverage can distort badly, especially in small samples or during stress episodes (Lahiri 2003; Pritsker 2006). Comparative studies point out this weakness: Goh and Pooi (2013) find a hypothesis-testing approach more reliable than normal-approximation or bootstrap intervals, yet it too rests on distributional assumptions that may fail in extreme downturns. The simulation-based methods of Wang (2003) improve accuracy at a computational cost that limits real-time use. *Bayesian* credible intervals (Contreras and Bhattacharjee 2004) incorporate parameter uncertainty directly through the posterior, but require prior assumptions about tail behavior that are difficult to justify when such information is scarce. More recent work continues to advance on this topic, while remaining subject to these constraints: He et al. (2020) employ a residual bootstrap under fixed-design assumptions, and Beutner et al. (2024) use metamodeling that may struggle to represent the tail.

Across these families of methods, a common pattern emerges: no single approach is both broadly applicable and robust. Each is tied to a particular model class or set of assumptions. A parametric form is a prerequisite for the Delta and asymptotic methods, a faithful resampling scheme for the bootstrap, and an informative prior for the Bayesian approach. Each of these procedures therefore becomes unreliable precisely when those assumptions fail, most acutely during periods of heightened volatility when risk assessment matters most. In short, these methods are *model-specific*: an interval procedure derived for one VaR forecasting model does not, in general, transfer to another. This is a significant practical limitation, because VaR is produced in practice by a wide range of forecasting methods, parametric volatility models, quantile regressions, historical simulation, copula constructions, and increasingly machine-learning models, for many of which no analytical interval method exists at all.

This motivates a different approach: rather than deriving a separate interval procedure for each VaR model, we seek a single calibration layer that can be wrapped around *any* VaR forecast, using only the return series and the forecast itself. Conformal prediction offers a natural foundation for this (Vovk et al. 2005). It is a distribution-free framework for uncertainty quantification that makes minimal assumptions and provides finite-sample coverage guarantees under exchangeability, treating the underlying forecast model as a black box. These properties, model-agnosticism and finite-sample validity, make it well suited to the VaR setting, where the diversity of forecasting models and the unreliability of parametric interval methods under stress are precisely the difficulties at issue.

Based on this, we propose *Quantile-DtACI* (QDtACI), a conformal calibration layer for Value-at-Risk that can be applied to any VaR point forecast, requiring only the realized returns and the forecast series. QDtACI adapts the dynamically-tuned adaptive conformal inference of Gibbs and Candès (2024) to the specific demands of quantile-based risk measurement. Our contributions are as follows.

First, we develop a conformal calibration procedure aligned with the quantile-estimation objective. Standard conformal methods target the conditional mean and use symmetric nonconformity scores; for VaR a one-sided tail quantile, this is inappropriate. QDtACI instead uses the pinball (quantile) loss as its nonconformity score, which is elicitable for the target quantile, and constructs the interval asymmetrically around the VaR forecast so that the two bounds respond independently to recent calibration behavior.

Second, we introduce a multi-speed adaptive mechanism suited to the sparse-violation VaR setting. Instead of the binary pinball-gradient update of the base method, which registers only whether a breach occurred, not its persistence, experts adapt through an exponentially smoothed violation error, yielding a continuous and more stable adaptation signal. Experts are then combined through a composite loss that jointly penalizes miscoverage, excessive interval width, and instability relative to the current consensus, reflecting the practical requirement that risk intervals be accurate, sharp, and stable rather than merely accurate.

Third, and most distinctively, QDtACI is *model-agnostic*: because it operates only on returns and the forecast, it applies uniformly across VaR models, including forecasts such as historical simulation, quantile regressions, copula constructions, and the growing class of machine-learning models, for which no analytical interval method exists. Since the method applies across such heterogeneous forecasts, the relevant question is not whether a model-specific interval outperforms another on a single forecast, but when a single calibration layer delivers reliable coverage across many, and under what conditions it does so. On synthetic data, where the true VaR is observable, we show that QDtACI attains near-nominal coverage of the true VaR when the underlying forecast is well-specified, and we quantify how coverage degrades as the forecast is progressively misspecified. Because coverage of the true VaR cannot be measured on real data, we complement the intervals with a return-only diagnostic that indicates when the calibration can be trusted.

Altogether, these elements make QDtACI a single, model-agnostic procedure for uncertainty quantification in VaR that is effective under changing market conditions. The remainder of the paper is organized as follows. "Conformal prediction" section introduces the conformal-prediction background and develops the QDtACI methodology. Empirical results" section reports the empirical results on both synthetic and real-world data, and Discussion and conclusion" section concludes.

Conformal prediction

Conformal prediction provides a general framework for valid uncertainty quantification in machine learning. Instead of outputting just a point estimate, it produces a prediction set (or interval) that is guaranteed to contain the true outcome with a user-specified probability.

The general framework of conformal prediction can be found in Vovk et al. (2005) and is briefly summarized below. Let $(X_1, Y_1), \ldots, (X_n, Y_n) \in \mathbb{R}^d \times \mathbb{R}$ represent a set of observed training data where X_i is a vector of features and Y_i is the corresponding label. For example, X_i may represent the asset information and Y_i is its return. Let (X_{n+1}, Y_{n+1}) be a new test point where only X_{n+1} is observed. To construct a prediction set, conformal inference begins by considering candidate values y representing possible realizations of Y_{n+1}.

For each candidate y, a nonconformity score function

$$S : (\mathbb{R}^d \times \mathbb{R})^n \times \mathbb{R}^d \times \mathbb{R} \to \mathbb{R}$$

is used to evaluate how well the imputed test point (X_{n+1}, y) aligns with the previously observed data $(X_1, Y_1), \ldots, (X_n, Y_n)$. The smaller the score, the more "conformal" the imputed test point (X_{n+1}, y) is to the observed training set. Nonconformity scores are defined based on the underlying machine learning algorithm used. For regression problems, this typically involves first estimating a regression model based on the available training data. The fitted model is used for all $n + 1$ data points and measures how well the prediction of y conforms with the model of X_{n+1}.

We define the nonconformity score $S(\cdot)$ as the absolute residual:

$$S\left((X_j, Y_j)_{1 \le j \le n}, (X_{n+1}, y)\right) := \left|y - \hat{\mu}(X_{n+1})\right|,$$

where $\hat{\mu}$ is an estimator of $\mathbb{E}[Y \mid X]$ based on the augmented training data $(X_1, Y_1), \ldots, (X_n, Y_n), (X_{n+1}, y)$.

Alternatively, it can be defined in terms of estimated probabilities:

$$S\left((X_j, Y_j)_{1 \le j \le n}, (X_{n+1}, y)\right) := 1 - \hat{\pi}(y \mid X_{n+1}),$$

where $\hat{\pi}(y \mid X_{n+1})$ represents a fitted estimate of $P(Y_{n+1} = y \mid X_{n+1})$.

Such nonconformity scores could be denoted by S_i^y, where:

$$S_i^y := S((X_j, Y_j)_{1 \le j \le n, j \ne i}, (X_{n+1}, y), (X_i, Y_i))$$

denotes the nonconformity score of the i-th training example when the candidate value for X_{n+1} is y, and:

$$S_{n+1}^y := S((X_j, Y_j)_{1 \le j \le n}, (X_{n+1}, y))$$

denote the nonconformity score of the new test point with the plausible label y.

In the final step of conformal prediction, the candidate value y is added to the prediction set if the test score is small relative to the training scores. For any $\tau \in \mathbb{R}$ and distribution D, let $\mathrm{Quantile}(\tau, D)$ denote the τ-th quantile of D, with the convention that

$$\mathrm{Quantile}(\tau, D) = \infty \quad \text{if } \tau \ge 1; \quad \text{and } \mathrm{Quantile}(\tau, D) = -\infty \quad \text{if } \tau \le 0.$$

Consequently, the conformal prediction set for a given coverage level $1-\alpha$ is formally defined as:

$$\hat{C}_\alpha(X_{n+1}) = \left\{ y : S^y_{n+1} \le \text{Quantile}\left(1-\alpha, \frac{1}{n+1}\sum_{i=1}^{n+1}\delta_{S^y_i}\right)\right\},$$

where $\delta_{S^y_i}$ denotes a Dirac measure at the value S^y_i. Regardless of the choice of nonconformity score, the resulting prediction set satisfies the following coverage guarantee for any model:

$$\mathbb{P}(Y_{n+1} \in \hat{C}_\alpha(X_{n+1})) \ge 1-\alpha.$$

This validity is automatically guaranteed under the assumption that the data are exchangeable; however, different choices of nonconformity scores may yield non-equivalent prediction sets. The usefulness of the resulting prediction set is therefore largely determined by the choice of nonconformity score.

For computational efficiency, inductive conformal prediction (also known as split conformal prediction) is often used, whereby the training data are divided into a proper training set and a calibration set (*Cal*). The proper training set is used to estimate the regression model, while the calibration set is used to compute nonconformity scores and construct prediction sets.

Adaptive conformal inference (ACI)—conformal prediction for time series

In general, the only necessary requirement for the nonconformity score function $S(\cdot)$ to ensure theoretical validity is exchangeability, that is, that it remains unchanged under permutations of its first n arguments. However, in time series settings, where the $(X_1, Y_1), \ldots, (X_n, Y_n), (X_{n+1}, Y_{n+1})$ exhibit temporal dependence, this assumption is often violated. Consequently, standard conformal prediction methods may lose their validity, and it becomes necessary to design procedures that do not rely on permutation invariance.

To address this issue, Adaptive Conformal Inference (ACI) proposed by Gibbs and Candès (2021) replaces the fixed miscoverage level α with a time-varying sequence α_t and updates it dynamically in an online setting. The method works as follows:

At time step t, ACI constructs the predictive set:

$$\hat{C}_{\alpha_t}(X_t) = \left[\hat{\mu}(X_t) \pm Q_{1-\alpha_t}(S_{Cal_t})\right],$$

where:

- $\hat{\mu}(X_t)$ is the predicted mean response for X_t,
- $Q_{1-\alpha_t}(S_{Cal_t})$ is the $(1-\alpha_t)$-th quantile of the residual scores computed from the calibration set Cal_t,
- The interval is centred on $\hat{\mu}(X_t)$, and its width depends on the quantile value.

Second, in order to improve adaptation, the miscoverage level α_t is updated recursively using:

$$\alpha_{t+1} = \alpha_t + \gamma\left(\alpha - \mathbb{I}\{Y_t \notin \hat{C}_{\alpha_t}(X_t)\}\right),$$

where:

- $\gamma \geq 0$ is the learning rate, which generally controls the speed of adaptation.
- $\mathbb{I}\{Y_t \notin \hat{C}_{\alpha_t}(X_t)\}$ is an indicator function that:
  - equals 1 if Y_t is not covered by the interval $\hat{C}_{\alpha_t}(X_t)$.
  - equals 0 if Y_t is covered.

If the prediction interval fails to cover Y_t, then $\alpha_{t+1} \leq \alpha_t$, which increases the interval size. On the other hand, if the prediction interval successfully covers y_t, then $\alpha_{t+1} \geq \alpha_t$, which shrinks the interval size. This adaptive mechanism ensures that the model dynamically adjusts to shifts in the data distribution.

This general methodology is later augmented by several authors, including Zaffran et al. (2022), but the most prominent improvement was proposed by Gibbs and Candès (2024), which will be explained in detail in the next section and will serve as the baseline for the methodology used in this paper.

Dynamically-tuned ACI (DtACI)

To enhance Adaptive Conformal Inference (ACI), Gibbs and Candès (2024) introduced Dynamically-tuned Adaptive Conformal Inference (DtACI). This method introduces a self-adjusting mechanism for the step-size parameter γ. DtACI selects the optimal step-size by leveraging an online expert aggregation approach. DtACI tends to be dynamic, as it continuously updates expert weights using an exponential re-weighting mechanism, thus allowing it to adapt quickly to distribution shifts and abrupt changes in the data distribution, which is often observed in the case of financial data.

This method builds upon previous work on ACI (Gibbs and Candès 2021; Gradu et al. 2023; Vovk 1990), while removing the need for manual tuning of γ, making it highly adaptive to non-stationary data distributions. Rather than selecting a single step size, DtACI maintains a collection of experts, each corresponding to a distinct step size, and dynamically reweights them based on their empirical performance. Based on Gibbs and Candès (2024), the DtACI framework is described as follows.

Let α_t denote the adaptive miscoverage level at time *t*. For a given observation Y_t , we use

$$\beta_t := \inf\left\{\beta : Y_t \in \hat{C}_\beta(X_t)\right\},$$

to represent the smallest conformity threshold such that the conformal prediction set contains Y_t.

The pinball loss function is defined as $\ell(\beta_t, \alpha_t)$

$$\ell(\beta_t, \alpha_t) := \alpha(\beta_t - \alpha_t) - \min\bigl(0, \beta_t - \alpha_t\bigr),$$

and is convex in α_t. Its subgradient with respect to α_t is given by

$$\nabla_{\alpha_t} \ell(\beta_t, \alpha_t) = \mathrm{err}_t - \alpha,$$

where $\mathrm{err}_t := \mathbb{I}\{Y_t \notin \hat{C}_{\alpha_t}(X_t)\}$ indicates a coverage error at time *t*.

Under this formulation, the ACI update

$$\alpha_{t+1} = \alpha_t + \gamma(\alpha - \mathrm{err}_t)$$

can be interpreted as a subgradient descent step with the learning rate γ:

$$\alpha_{t+1} = \alpha_t - \gamma \nabla_{\alpha_t} \ell(\beta_t, \alpha_t).$$

As noted, DtACI runs *K* experts in parallel, each with a *fixed* step size γ_k. Each expert *k* maintains its own miscoverage sequence $\{\alpha_t^k\}_{t \geq 1}$, updated according to the ACI recursion.

The weight of expert *k* is updated using exponential reweighting based on the pinball loss:

$$\tilde{w}_t^k = w_t^k \cdot \exp\bigl(-\eta\, \ell(\beta_t, \alpha_t^k)\bigr)\,,$$

where $\eta > 0$ is a learning rate that controls sensitivity to recent performance.

To prevent expert weights from vanishing, a smoothing step is applied:

$$w_{t+1}^k = (1 - \sigma)\tilde{w}_t^k + \frac{\sigma}{K} \sum_{j=1}^{K} \tilde{w}_t^j,$$

where $\sigma \in (0, 1)$ controls the degree of weight sharing.

The normalized expert probabilities are thus defined as

$$p_t^k = \frac{w_t^k}{\sum_{j=1}^{K} w_t^j}.$$

The final adaptive miscoverage level used to construct the conformal prediction set is obtained via the weighted aggregation

$$\alpha_t = \sum_{k=1}^{K} p_t^k \alpha_t^k.$$

This aggregation induces an adaptive effective step size, enabling DtACI to rapidly respond to distributional shifts while preserving the coverage guarantees inherited from ACI and online learning (Gibbs and Candès 2024).

In summary, DtACI employs an online expert aggregation approach to dynamically adjust the step size γ. It employs an exponential weighting mechanism that continuously adjusts the importance of experts based on the pinball loss function. Such continuous adjustment makes DtACI very flexible and responsive to recent performance. Its adaptiveness means that it can perform well under both smooth and sudden distribution shifts, due to the rapid reallocation of importance to the best-performing experts. As such, it is the prime candidate for testing this framework in financial data settings, especially during periods of sudden and abrupt market changes.

Conformal inference for VaR—developing DtACI into QDtACI

Quantile-based nonconformity score for VaR-centered intervals

Three distinct levels appear throughout this section, and it is important to keep them apart. First, the *VaR level*$a = 0.05$, which is fixed and represents the target violation rate of the underlying risk model. Second, the *interval confidence level*β (e.g. 90 , 95, 99%), which governs how wide a conformal interval is placed around the VaR forecast. Third, the *adaptive level*α_t, which is updated online at each time step. It is important to note that β does not alter the quantile being estimated, e.g. a 99% conformal interval is a wider bracket around the *same*5% VaR, not an estimate of the 1% VaR.

A note on convention. In "Adaptive conformal inference (ACI)—conformal prediction for time series" section and following, α_t denotes a *miscoverage* level, as defined in Gibbs and Candès (2021, 2024). For the VaR application we instead work with α_t as a *confidence* level, so that larger α_t corresponds to a wider interval and the conformal quantile is read at position α_t rather than $1 - \alpha_t$. Either convention may be used: the confidence level is simply one minus the miscoverage level, and switching between them reverses the sign of the adaptive update. We adopt the confidence form because the interval is constructed asymmetrically around a one-sided risk forecast, and because the adaptation then reads naturally: a violation raises α_t and widens the interval.

In the proposed framework, predictive intervals are constructed around the model's forecasted Value-at-Risk (VaR) level q_t, corresponding to the target quantile level $a \in (0, 1)$, fixed at $a = 0.05$ throughout. To ensure that the conformal calibration is consistent with the quantile estimation objective, we employ the *quantile loss* (also known as the *pinball loss*) as the nonconformity score.

Definition Let Y_t denote the realized return at time *t*, and q_t the predicted quantile (VaR) associated with level *a*. The nonconformity score is defined as

$$S_a(Y_t, q_t) = (a - \mathbb{I}\{Y_t \leq q_t\})\,(Y_t - q_t),$$

where $\mathbb{I}\{\cdot\}$ is the indicator function that equals 1 if the event is true and 0 otherwise. This formulation yields positive scores proportional to the magnitude of misestimation relative to the target quantile.

Properties The quantile score function is asymmetric with respect to the forecast error $(Y_t - q_t)$, assigning a higher penalty to deviations below the predicted quantile when $a < 0.5$ (which is always true for VaR, usually calculated for $a = 0.05$ or $a = 0.01$). It satisfies the following key properties:

- *Quantile consistency:* Minimizing $\mathbb{E}[S_a(Y_t, q_t)]$ with respect to q_t yields the true conditional *a*-quantile of the return distribution.
- *Directional interpretability:* The sign of $(a - \mathbb{I}\{Y_t \leq q_t\})$ indicates the direction of model miscalibration: if $Y_t < q_t$, the VaR is underestimated (violation occurs), while $Y_t > q_t$ corresponds to conservatism.
- *Non-negativity:* The score is always non-negative, equal to zero only when $Y_t = q_t$, and can be aggregated over a rolling calibration window to form an empirical distribution of quantile misestimations.

This set of properties ensures that the nonconformity score used in the study is optimized precisely for the quantile estimation task, in contrast to symmetric scores that could otherwise be used (e.g. absolute residual) as a measure of misalignment with VaR.

Empirical distribution for calibration. For a calibration window $\mathcal{I}_t = \{t - W, \ldots, t - 1\}$, the empirical distribution of nonconformity scores is

$$\mathcal{S}_t = \{S_a(Y_\tau, q_\tau) : \tau \in \mathcal{I}_t\},$$

from which conformal quantiles are estimated to adjust the lower and upper widths of the predictive interval around q_t. This construction ensures that both sides of the VaR-centered interval adapt coherently to the realized quantile deviations while preserving consistency with the desired confidence level β.

Asymmetric interval construction. The lower and upper edges of the predictive interval are obtained from two conformal quantile positions rather than a single symmetric one, allowing the two sides to respond independently to recent calibration behavior. Let $p = \lfloor \bar{\alpha}_t\, |\mathcal{S}_t| \rfloor$ denote the base quantile position in the score distribution, where $\bar{\alpha}_t$ is the aggregated adaptive confidence level of "Multi-speed adaptive calibration via expert averaging" section, and let $\hat{v}_t$ be the realized violation rate of the VaR forecast over a short recent window (the most recent $\min(50,\, W/5)$ observations). Define a directional adjustment:

$$(\delta_t^{\mathrm{lo}}, \delta_t^{\mathrm{up}}) = \begin{cases} \left(1,\ 1 + \dfrac{a - \hat{v}_t}{a}\right), & \hat{v}_t < a, \\ \left(1 + \dfrac{\hat{v}_t - a}{a},\ 1\right), & \hat{v}_t \geq a, \end{cases}$$

with each factor clipped to [0.5, 2.0], where a is the VaR level. The lower and upper critical values are then read at the adjusted positions $\lfloor p\,\delta_t^{\mathrm{lo}} \rfloor$ and $\lfloor p\,\delta_t^{\mathrm{up}} \rfloor$, giving the interval $[\,q_t - c_t^{\mathrm{lo}},\ q_t + c_t^{\mathrm{up}}\,]$. Intuitively, when recent violations exceed the VaR level the lower edge is widened (the side facing the losses the model is underestimating), and when they fall short of it the upper edge is widened instead; under on-target behavior both factors equal one and the construction reduces to the symmetric case. To prevent the upper edge from inflating without bound during transient calm spells, it is additionally capped at $q_t + \lambda_{\mathrm{ub}}\,|q_t|$ with $\lambda_{\mathrm{ub}} = 0.5$.

Rank-based score stabilization

As noted earlier, the pinball score $S_a(Y_t, q_t)$ is non-negative for all realized returns Y_t: on non-violation days it equals $a \cdot (Y_t - q_t) \geq 0$, and on violation days $(1-a)(q_t - Y_t) \geq 0$, equal to zero only when $Y_t = q_t$. A correctly specified VaR model therefore generates a calibration score distribution concentrated near zero but with strictly positive values. This raises a natural question: whether reshaping that distribution, e.g. pulling the dense lower mass upward and compressing the sparse upper tail, would improve the discrimination of the conformal quantile across confidence levels. We therefore include an optional monotonic rank-based transformation in the formulation, and test it empirically.

Definition Let $\mathcal{S}_t = \{S_a(Y_\tau, q_\tau) : \tau \in \mathcal{I}_t\}$ denote the calibration window of length $W = |\mathcal{I}_t|$, with order statistics $s_{(1)} \leq \cdots \leq s_{(W)}$, and let $S_{\min} = s_{(1)}$, $S_{\max} = s_{(W)}$. For each $s_{(i)}$, define its normalized rank

$$r_i = \frac{i-1}{W-1}, \qquad r_i \in [0,1].$$

For $\kappa > 0$, define the transformation $\phi_\kappa : [0,1] \to [0,1]$ by

$$\phi_\kappa(r) = \frac{1}{2}\left(1 + \frac{\tanh\big(\kappa(r - \frac{1}{2})\big)}{\tanh(\kappa/2)}\right),$$

and the stabilized score as

$$s_{(i)}^{(\mathrm{stab})} = S_{\min} + \phi_\kappa(r_i)\,(S_{\max} - S_{\min}).$$

The function ϕ_κ is smooth and strictly increasing for all $\kappa > 0$, with $\phi_\kappa(0) = 0$ and $\phi_\kappa(1) = 1$. The rank ordering of scores is preserved by construction: the observation at rank $\lceil aW \rceil$ before the transformation remains at that rank afterwards, and only its numeric value changes, predictably via ϕ_κ. The identity transformation is recovered at $\kappa = 0$. The formal argument is provided in "Appendix 1" and "Appendix 2".

Parameter selection. The curvature parameter κ controls the strength of redistribution: as $\kappa \to 0$ the transformation approaches the identity, while larger values produce a stronger S-shaped correction. Because the transformation modifies the score distribution used for calibration, it carries no theoretical guarantee that nominal coverage is preserved, and κ therefore requires empirical validation. We conduct this validation in the synthetic setting, where the true data-generating VaR is observable, evaluating $\kappa \in \{0, 1, 2, 3\}$; the results are reported in "The score transformation" section.

Multi-speed adaptive calibration via expert averaging

QDtACI adopts the expert-aggregation structure of DtACI (Gibbs and Candès 2024), maintaining a collection of calibration experts that evolve in parallel at different adaptation speeds, with each expert tracking deviations from its target breach rate using a distinct learning rate. It departs from DtACI in two respects. First, expert states are updated via an exponentially smoothed violation error rather than a direct pinball loss gradient. The raw pinball gradient switches between only two values and is insensitive to the magnitude of violations; thus a small breach and a catastrophic breach have exactly the same impact. The exponential moving average instead accumulates recent violation history into a continuous signal, providing smoother and more stable adaptation in the VaR setting where violations are rare by construction. Second, expert weights are determined by a composite loss that jointly penalizes miscoverage, interval width, and deviation from the ensemble, rather than pinball loss alone, reflecting the practical requirement in risk management that intervals should be accurate but also stable and have reasonable width. We refer to the resulting procedure as *Quantile-DtACI (QDtACI)* to emphasise its specialisation to quantile-based risk forecasting.

Expert structure. Assume that there are N experts. Let $i \in \{1, \ldots, N\}$ index a collection of calibration experts and let t denote time. Each expert maintains:

- an adaptive confidence level $\alpha_{i,t}$,
- an exponentially smoothed calibration error $\epsilon_{i,t}$.

Experts differ only through their learning rates $\gamma_i > 0$. Small values of γ_i produce slowly varying and stable calibration paths, whereas larger values allow rapid adjustment to abrupt regime changes.

Internal adaptation. At time t, each expert i records whether the previous return breached its own lower interval bound,

$$v_{i,t} = \mathbb{I}\{Y_t < L_{i,t-1}\},$$

where $L_{i,t-1}$ is the lower edge of expert i's interval at the previous step. Each expert updates its smoothed calibration error according to

$$\epsilon_{i,t} = \lambda\, \epsilon_{i,t-1} + (1-\lambda)(v_{i,t} - a^*), \qquad \lambda \in [0,1),$$

where λ controls the memory decay of past deviations and a^* is the ACI target breach rate. Unlike the VaR level *a*, which is fixed, the target a^* varies with the interval confidence level: it is calibrated on synthetic data, where the true VaR is observable, as the breach rate that yields nominal coverage, and declines roughly linearly with the confidence level (from about 0.04 at 90% to about 0.025 at 99%), remaining below the VaR level *a* throughout. The adaptive level is then updated via

$$\alpha_{i,t+1} = \Pi_{[0,1]}\big(\alpha_{i,t} + \gamma_i \epsilon_{i,t}\big),$$

where $\Pi_{[0,1]}$ denotes projection onto the unit interval. Under the confidence convention adopted in this section, this update *increases*$\alpha_{i,t}$ when observed breaches exceed the target level, thus widening the interval, and decreases it otherwise, so that a breach acts to widen the band.

Performance-based weighting. To combine experts, we assign each expert *i* an instantaneous performance loss $L_{i,t}$. The loss is a bounded convex combination of three components: (1) a calibration term (penalizing violation rates that deviate from the target), (2) a width penalty (discouraging excessively wide intervals), and (3) a stability penalty (discouraging excessive deviation of the expert from the ensemble):

$$L_{i,t} = \omega_{\text{cov}}\,(\widehat{p}_{i,t} - a^*)^2 + \omega_{\text{wid}}\,\text{clip}\left(\frac{\text{Width}_{i,t}}{w_0} - 1,\ 0,\ 1\right) + \omega_\alpha\,(\alpha_{i,t} - \bar{\alpha}_{t-1})^2,$$

where:

- $\widehat{p}_{i,t}$ is a rolling violation rate for expert *i*,
- $\text{Width}_{i,t}$ is the width of the prediction interval produced by expert *i*,
- w_0 is a reference width, taken as the nominal conformal interval width over the current calibration window,
- $\bar{\alpha}_{t-1}$ is the previously aggregated level.

The loss coefficients $\omega_{\text{cov}}, \omega_{\text{wid}}$ and ω_α satisfy

$$\omega_{\text{cov}} + \omega_{\text{wid}} + \omega_\alpha = 1, \qquad \omega_{\text{cov}}, \omega_{\text{wid}}, \omega_\alpha \ge 0.$$

Time-decayed expert weights. To emphasize recent performance while retaining historical information, we define a time-decayed cumulative loss:

$$\bar{L}_{i,t} = \sum_{s=1}^{t} \delta^{\,t-s} L_{i,s}, \qquad \delta \in (0,1).$$

Expert weights are then computed using an exponential mapping:

$$\tilde{w}_{i,t} = \frac{\exp(-\eta\,\bar{L}_{i,t})}{\sum_{j=1}^{N}\exp(-\eta\,\bar{L}_{j,t})}, \qquad \eta > 0.$$

To prevent weight collapse and ensure stability, we mix expert weights with the uniform distribution,

$$w_{i,t} = (1-\rho)\tilde{w}_{i,t} + \rho\frac{1}{N}, \qquad \rho \in (0,1),$$

so that $\sum_{i=1}^{N} w_{i,t} = 1$.

Aggregated adaptive level. The effective adaptive confidence level used for interval construction is

$$\bar{\alpha}_t = \sum_{i=1}^{N} w_{i,t}\,\alpha_{i,t}.$$

Thus, each expert's level $\alpha_{i,t}$ evolves through its own ACI update above; the weights $w_{i,t}$ govern only how the experts are aggregated.

Interpretation. This multi-speed structure enables simultaneous responsiveness to short-term volatility spikes and long-term distribution drift. Fast experts react quickly to sudden regime shifts, while slow experts provide stability during calm periods. The performance-based weighting dynamically reallocates importance toward the most reliable adaptation speeds while maintaining regularization through uniform mixing.

Implementation details

In all experiments, QDtACI is implemented with $N = 40$ experts whose learning rates γ_i are spaced logarithmically over the interval $[10^{-4}, 10^{-1}]$, providing coverage of both slow and fast adaptation speeds. The EMA memory parameter is set to $\lambda = 0.80$, corresponding to a half-life of approximately three trading days, and the cumulative-loss decay to $\delta = 0.95$. Expert weights are formed with softmax temperature $\eta = 0.5$ and uniform-mixing weight $\rho = 0.1$. The composite loss coefficients are fixed as $\omega_{\text{cov}} = 0.90$, $\omega_{\text{wid}} = 0.05$, and $\omega_{\alpha} = 0.05$ across all experiments, reflecting the primary objective of achieving correct coverage while applying mild regularization on interval width and expert stability. The calibration window is set to $W = 250$ trading days, and the initial adaptive confidence level is set to $\bar{\alpha}_0 = 0.95$ for all experts. Conformal quantiles use the finite-sample $(n+1)$ correction, and the score distribution's upper support is stabilised at its 99% quantile to limit the influence of extreme observations. Algorithm 1 summarises the full QDtACI procedure.

Require: Returns $\{Y_t\}$, VaR forecasts $\{q_t\}$, VaR level a, target breach rate a^*, window W, experts N with rates $\{\gamma_i\}_{i=1}^{N}$, parameters $\lambda, \delta, \eta, \rho$, loss weights $(\omega_{\text{cov}}, \omega_{\text{wid}}, \omega_\alpha)$, transformation κ

Initialise $\alpha_{i,0} \leftarrow \bar{\alpha}_0$, $\epsilon_{i,0} \leftarrow 0$, $\bar{L}_{i,0} \leftarrow 0$, $w_{i,0} \leftarrow 1/N$ for all i

for $t = W, W+1, \ldots$ **do**

$\mathcal{S}_t \leftarrow \{\, S_a(Y_\tau, q_\tau) : \tau \in \{t-W, \ldots, t-1\} \,\}$ ▷ pinball scores over the window

apply rank transformation ϕ_κ to $\mathcal{S}_t$ ▷ identity at $\kappa = 0$

Expert update (for each i):

$v_{i,t} \leftarrow \mathbb{I}\{Y_t < L_{i,t-1}\}$ ▷ own lower-bound breach

$\epsilon_{i,t} \leftarrow \lambda\, \epsilon_{i,t-1} + (1-\lambda)(v_{i,t} - a^*)$ ▷ smoothed breach error

$\alpha_{i,t} \leftarrow \Pi_{[0,1]}\big(\alpha_{i,t-1} + \gamma_i\, \epsilon_{i,t}\big)$ ▷ ACI update (breach widens)

Expert weighting:

$L_{i,t} \leftarrow \omega_{\text{cov}}(\widehat{p}_{i,t} - a^*)^2 + \omega_{\text{wid}}\, \text{clip}\big(\frac{\text{Width}_{i,t}}{w_0} - 1, 0, 1\big) + \omega_\alpha(\alpha_{i,t} - \bar{\alpha}_{t-1})^2$

$\bar{L}_{i,t} \leftarrow \delta\, \bar{L}_{i,t-1} + L_{i,t}$ ▷ decayed cumulative loss

$\tilde{w}_{i,t} \leftarrow \dfrac{\exp(-\eta\, \bar{L}_{i,t})}{\sum_j \exp(-\eta\, \bar{L}_{j,t})}, \quad w_{i,t} \leftarrow (1-\rho)\tilde{w}_{i,t} + \rho/N$

Aggregate and construct interval:

$\bar{\alpha}_t \leftarrow \sum_{i=1}^{N} w_{i,t}\, \alpha_{i,t}$

$p \leftarrow \lfloor \bar{\alpha}_t\, |\mathcal{S}_t| \rfloor$; compute $\hat{v}_t$ over the last $\min(50, W/5)$ days

set $(\delta_t^{\text{lo}}, \delta_t^{\text{up}})$ from $\hat{v}_t$ vs. a; read $c_t^{\text{lo}}, c_t^{\text{up}}$ at $\lfloor p\, \delta_t^{\text{lo}} \rfloor, \lfloor p\, \delta_t^{\text{up}} \rfloor$

$L_t \leftarrow q_t - c_t^{\text{lo}}$; $U_t \leftarrow \min\big(q_t + c_t^{\text{up}},\ q_t + \lambda_{\text{ub}}|q_t|\big)$

output interval $[L_t, U_t]$

end for

Algorithm 1 Quantile-DtACI (QDtACI)

Empirical results

In this section we evaluate QDtACI methodology in two complementary settings. First, we use synthetic data generated from a known process, where the true VaR is observable and can serve as a benchmark, which lets us measure directly how well the intervals cover the quantity of interest, which is impossible with real data. Second, we apply the method to a real bond portfolio, where the true VaR is not known and the intervals are instead assessed through their observable behavior.

The analysis proceeds in five parts. "Synthetic experimental design" section describes the data-generating process and the point-VaR specifications that QDtACI is applied to. "Benchmark against parametric interval methods" section benchmarks QDtACI against parametric interval methods on well-specified forecasts. "The score transformation" section analyses the score transformation and justifies the identity setting adopted throughout. "Forecast quality and coverage of the true VaR" section establishes how coverage of the true VaR depends on the quality of the underlying point forecast, and connects this to a return-only diagnostic that can be applied in a real-world setting, when the true VaR is not observable. "Application to a real bond portfolio" section presents the real-data application.

We score intervals against the true VaR where it is observable (the synthetic data-generating process), using the prediction interval coverage probability (PICP), the Winkler score, and mean interval width. Lower Winkler is better, while PICP should approach the nominal level.

Synthetic experimental design

Synthetic return series are generated using a GARCH(1,1) process. A total of 3100 observations are simulated to accommodate subsequent calculations and the loss of observations arising from rolling-window estimation. Returns are generated according to:

$$r_t = \sigma_t \varepsilon_t,$$
$$\sigma_t^2 = \omega + \alpha r_{t-1}^2 + \beta \sigma_{t-1}^2,$$

where $\varepsilon_t \sim \mathcal{N}(0,1)$ and $\omega > 0$.

The GARCH parameters are set to $\alpha = 0.10$ and $\beta = 0.85$, ensuring covariance stationarity and a high degree of volatility persistence consistent with empirical financial return series. Because the data are generated from a known process, the true VaR can be computed directly from the data-generating parameters and serves as the evaluation benchmark throughout this section.

We construct QDtACI around three point-VaR specifications chosen to span a range of forecast quality relative to this true VaR. Two are *well-specified*: a GARCH(1, 1) model estimated on the simulated returns, and a CAViaR model (Engle and Manganelli 2004). The GARCH specification is estimated by searching orders p, q up to 4; the GARCH(1, 1) is recovered as preferred, which is precisely the data-generating process. Also, CAViaR is well suited to this setting because it models the conditional quantile directly, without a parametric distributional assumption, and targets the VaR level of a single return series. Together, these two represent close to the best achievable univariate forecasts for this process.[1] The third specification is *deliberately crude*: a historical-simulation VaR computed on a rolling 250-day window, which responds slowly to volatility changes and is known to perform poorly during periods of elevated volatility and regime shifts, owing to its implicit assumption that returns are exchangeable over the estimation window (Pritsker 2006; McNeil et al. 2015) (Fig. 1).

This spread of specifications is intentional, and its purpose is not to show that QDtACI attains valid coverage regardless of the centre - it does not, as "Forecast quality and coverage of the true VaR" section makes clear. Rather, because the true VaR is observable here, the well-specified and crude centres together let us *characterize* how interval calibration relates to forecast quality: the two model-based VaRs

[1] The copula-based VaR used in the real-data application ("Application to a real bond portfolio" section) is inherently a multi-asset construction and cannot be evaluated on this univariate synthetic series; it is therefore absent from the synthetic experiments.

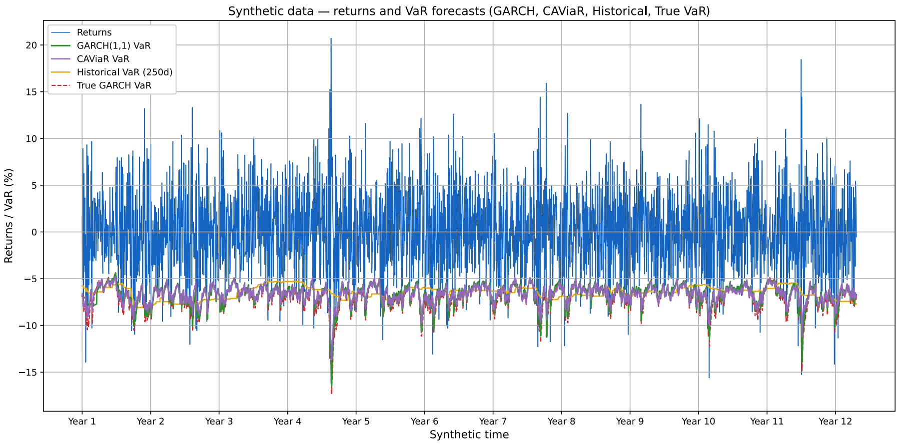

Fig. 1 The three point-VaR specifications on the synthetic data, together with the true VaR (dashed) and realized returns. The GARCH(1, 1) and CAViaR forecasts track the true VaR closely, whereas the historical-simulation VaR lags it, responding slowly to changes in volatility. This spread of forecast quality is what allows the coverage–quality relationship to be characterized in "Forecast quality and coverage of the true VaR" section

(GARCH(1,1) and CAViaR) establish the behavior when the forecast is close to the truth, and the historical VaR exposes what happens when it is not. Since the true VaR is never observable in practice and some misspecification is unavoidable, characterizing this dependence and providing a means to detect the regime from returns alone ("Forecast quality and coverage of the true VaR" section) is of direct practical value. Each specification is evaluated over its own usable window after estimation and conformal burn-in.

Benchmark against parametric interval methods

We compare QDtACI with two parametric interval constructions, the Delta method and a parametric (block) bootstrap, on two well-specified methods, a GARCH(1, 1) VaR and a CAViaR VaR. Predictions for both of these methods pass standard point-VaR backtests ("Forecast quality and coverage of the true VaR" section). As a further comparison, we report DtACI, Gibbs and Candès (2024) adaptive conformal method modified to the VaR setting. QDtACI departs from this canonical method in three aspects: (i) experts are scored by a composite loss combining a rolling coverage error, an interval-width efficiency term, and the stability of the adaptive level (weights 0.90/0.05/0.05), rather than by a coverage objective alone; (ii) the interval is constructed asymmetrically for the one-sided VaR; and (iii) an optional rank transformation shapes the conformity scores, set to the identity throughout ("The score transformation" section), so both methods here operate on the raw scores. All methods are scored against the true VaR over a common evaluation window. Results are presented in Table 1.

Coverage. At the 95% level QDtACI maintains near-nominal coverage of the true VaR. It attains 0.961 (for GARCH) and 0.936 (for CAViaR), close to nomi-

Table 1 Interval methods scored against the true VaR, at the 95% and 99% levels

Centre	Method	PICP	Winkler	Width	n
95% level					
GARCH(1, 1)	Delta	0.895	0.032	0.018	2600
	Bootstrap	0.580	0.042	0.007	2600
	QDtACI	0.961	0.030	0.028	2600
	DtACI (baseline)	1.000	0.087	0.087	2600
CAViaR	Delta	0.853	0.029	0.013	2850
	Bootstrap	0.858	0.039	0.018	2850
	QDtACI	0.936	0.034	0.027	2850
	DtACI (baseline)	0.998	0.084	0.084	2850
99% level					
GARCH(1, 1)	Delta	0.961	0.058	0.024	2600
	Bootstrap	0.722	0.104	0.009	2600
	QDtACI	0.992	0.038	0.036	2600
	DtACI (baseline)	1.000	0.089	0.089	2600
CAViaR	Delta	0.940	0.049	0.017	2850
	Bootstrap	0.938	0.070	0.024	2850
	QDtACI	0.976	0.047	0.035	2850
	DtACI (baseline)	1.000	0.091	0.091	2850

PICP is coverage of the true VaR (nominal in parentheses); Winkler is a proper interval score (lower is better); width is the mean interval width. Delta and bootstrap are parametric interval methods; DtACI is the adaptive conformal baseline

nal, based on PICP. The Delta method undercovers substantially, 0.895 and 0.853 at the 95% level, for these two models respectively. Similarly, the bootstrap, restricted to parameter uncertainty, undercovers more severely (0.580 on the well-specified GARCH). At the 99% level situation is similar, QDtACI has near-nominal coverage (0.992 on GARCH and 0.976 on CAViaR), whereas the parametric methods fall short (0.961 and 0.940 for Delta, with lower coverage for the bootstrap, notably 0.722 on GARCH). QDtACI thus attains near-nominal coverage in the tail, where the parametric methods undercover more substantially.

Interval quality (Winkler). On the Winkler score, QDtACI is competitive with or superior to the parametric methods. At 95% it is ahead of Delta on GARCH (0.030 versus 0.032) and modestly behind on CAViaR (0.034 versus 0.029), while ahead of the bootstrap on both VaR predictions. At the 99% level QDtACI has the lowest Winkler score of all methods on both VaR predictions (0.038 and 0.047, versus 0.058 and 0.049 for Delta and 0.104 and 0.070 for the bootstrap). The QDtACI advantage at 99% reflects how the Winkler score works: it penalizes realisations outside the interval more heavily at higher confidence levels. The parametric methods undercover the tail and incur these penalties, while QDtACI covers it and avoids them. The parametric methods' modest Winkler edge is now confined to the single 95% CAViaR case; at 99% it is gone entirely, with valid coverage and interval quality aligning rather than trading off.

Comparison with the Gibbs and Candès (2024) baseline. The DtACI comparison isolates the value of QDtACI's adaptations. DtACI over-covers by defaulting to a wide band rather than by calibrating to the target; its width ratio to QDtACI is nearly constant across VaR predictions and confidence levels, indicating a fixed inflation rather than adaptation. This is the direct consequence of the composite loss: because DtACI's objective contains no width-efficiency term, nothing restrains its width once coverage is met, whereas QDtACI's efficiency term holds the interval to roughly one-third of DtACI's width, at comparable coverage. The added adaptation to the method, therefore, earns its place.

The score transformation

The method includes a rank transformation of the conformity scores parameterised by κ, with $\kappa = 0$ recovering the raw conformal scores (the identity). Table 2 reports coverage, the Winkler score and mean width across the 90, 95 and 99% levels, for $\kappa = 0, 1, 2, 3$, for the two models.

Increasing κ lowers coverage of the true VaR at every confidence level, on both centres, while the mean interval width remains essentially unchanged: the adaptive controller re-targets the breach rate regardless of κ, so the transformation reshapes the scores without altering the band's size. The Winkler score is not improved in return; it rises monotonically with κ, and $\kappa = 0$ attains the lowest score in every cell of the table. In other words, $\kappa > 0$ costs coverage and worsens the proper score at no saving in width, on both centres and across all three levels. We therefore set $\kappa = 0$: the transformation is retained in the formulation but set to the identity, so the deployed method uses the raw conformal scores. The historical VaR, whose behavior differs qualitatively, is examined in "Forecast quality and coverage of the true VaR" section.

Forecast quality and coverage of the true VaR

As noted, the adaptive layer of QDtACI controls an *observable* quantity, the rate at which returns breach the lower bound, whereas the property of interest, coverage of the *latent* true VaR, is observable only on synthetic data. The two do not necessarily coincide. This section shows when they do, using a controlled experiment, and provides a return-only diagnostic that indicates the regime on real data.

The decoupling. Figure 2 contrasts a well-specified VaR method (GARCH) with a crude one (a 250-day historical VaR) over the sample path. In both, the rolling lower-

Table 2 Sensitivity to the transformation parameter κ, for the two well-specified methods, at the 90, 95 and 99% levels

κ	PICP			Winkler			Width		
	90%	95%	99%	90%	95%	99%	90%	95%	99%
GARCH(1, 1)									
0	0.912	0.961	0.992	0.0251	0.0301	0.0380	0.0221	0.0280	0.0356
1	0.902	0.958	0.992	0.0256	0.0305	0.0385	0.0222	0.0281	0.0357
2	0.868	0.939	0.987	0.0268	0.0316	0.0404	0.0221	0.0277	0.0359
3	0.813	0.898	0.978	0.0288	0.0341	0.0441	0.0221	0.0276	0.0363
CAViaR									
0	0.888	0.936	0.976	0.0279	0.0341	0.0465	0.0216	0.0275	0.0354
1	0.882	0.933	0.975	0.0281	0.0344	0.0472	0.0219	0.0275	0.0356
2	0.862	0.921	0.972	0.0293	0.0357	0.0488	0.0221	0.0280	0.0359
3	0.817	0.891	0.968	0.0314	0.0378	0.0518	0.0218	0.0276	0.0367

Increasing κ lowers coverage at every level at essentially unchanged width, while the Winkler score rises. The identity transformation, $\kappa = 0$, is adopted throughout

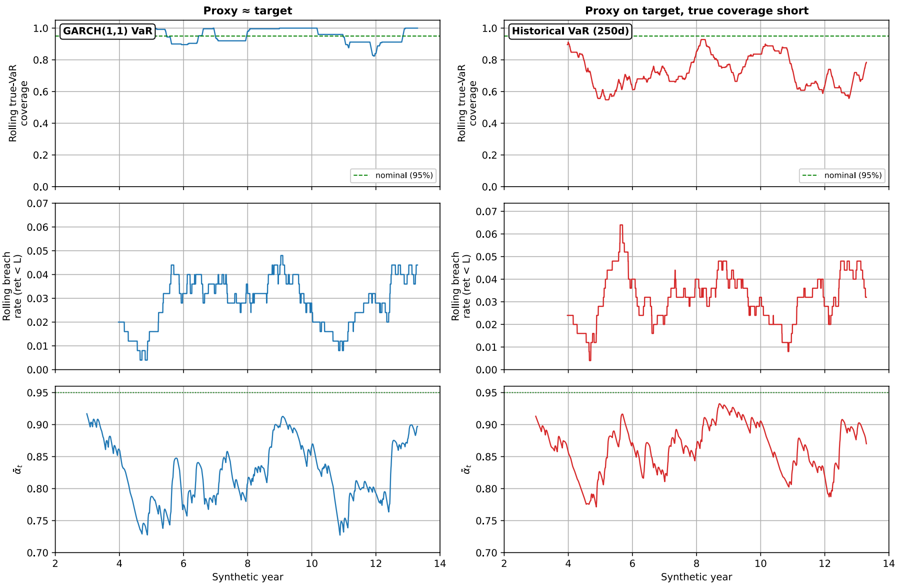

Fig. 2 Adaptive calibration controls the lower-breach rate, not coverage of the true VaR. Left: a well-specified centre, for which the breach rate and true coverage coincide. Right: a crude centre, for which the breach rate behaves comparably yet true coverage (top) falls well below nominal. The confidence level $\bar{\alpha}_t$ (bottom) show almost consistent higher level for historical VaR. Thus, the coverage divergence originates in the centre, not in the adaptation

breach rate varies over time within a comparable range, and the adaptive confidence levels follow broadly similar paths, though the historical centre's level sits somewhat higher. Yet the rolling coverage of the true VaR tracks nominal for the GARCH centre and falls persistently short for the historical one. In short, the observable quantity (the breach rate) is controlled comparably across both regimes, while the latent target (true coverage) diverges, because the historical centre is misplaced and the calibration layer cannot observe that misplacement.

A controlled degradation experiment. To isolate the role of the centre, we construct the QDtACI intervals around a degrading point forecast, replacing the centre at time *t* with the true VaR delayed by *k* days, that is, delaying whole point prediction of true VaR by *k* days. At $k = 0$ the band is built around the exact true VaR; larger *k* reproduces the slow reaction of a trailing estimator. Every lag is scored against the same true VaR over an identical window. Figure 3 reports the result. Coverage of the true VaR falls monotonically with the lag, from essentially complete at $k = 0$ to 0.59 at $k = 10$ and 0.54 by $k = 30$. Crucially, the realized breach rate controlled by the QDtACI methodology remains mostly flat across all lags: the breach rate at $k = 30$ is no higher than at $k = 0$. The coverage loss therefore does not reflect a failure of the calibration layer to do its job: it continues to control the breach rate at every lag. What it cannot do is correct the *location* of the interval, which is inherited from the point forecast. Coverage of the true VaR requires the band to be both correctly controlled and correctly placed, and so degrades exactly as the centre degrades. The three benchmark VaR forecasts are presented at this graph as horizontal lines: the

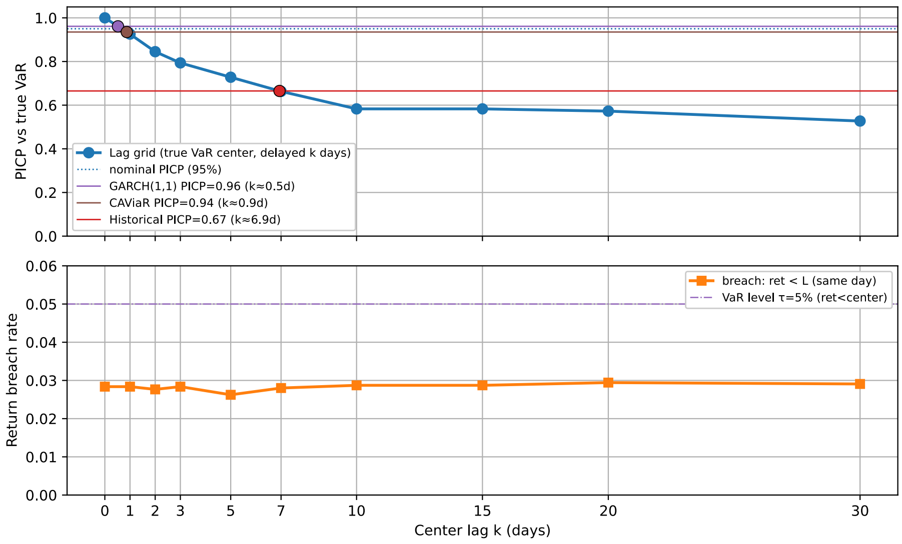

Fig. 3 Coverage of the true VaR degrades monotonically as the centre is delayed by k days (top), while the realized breach rate (bottom) controlled by the adaptive conformal predictor remains largely flat. Thus, the coverage loss is due to mis-location of a correctly sized band, not due to failed mechanism. Horizontal lines refer to the achieved PICP at a 95% confidence level of three real VaR point predictions

Table 3 Backtests on each VaR forecast (p values; pass if $p > 0.05$), with the true-VaR coverage of the corresponding QDtACI interval at the 95% level

Centre	Rate (%)	Kupiec	Christ	DQ	Verdict	True-VaR PICP
GARCH(1, 1)	5.1	0.830	0.150	1.000	pass	0.961
CAViaR	5.1	0.764	0.291	1.000	pass	0.936
Historical	6.1	0.007	0.006	0.791	fail	0.665

Forecasts that pass the backtests attain near-nominal coverage; the forecasts that fail do not

well-specified GARCH (0.96) and CAViaR (0.94) are thus placed between zero and one effective lag, and the historical VaR at 0.67 PICP, at almost 7d lag.

A return-only diagnostic. On real data the true VaR is latent, so coverage cannot be measured directly. The relationship above, however, is governed by the quality of the forecast, and forecasts quality from returns alone is assessable through standard point-VaR backtests. Table 3 reports the Kupiec, Christoffersen and dynamic-quantile (DQ) tests on each forecast, beside the true-VaR coverage measured on the synthetic data.

The two centres that pass the backtests are exactly those that deliver near-nominal true-VaR coverage; the centre that fails is the one that undercovers. The historical VaR fails the unconditional-coverage (Kupiec) and conditional-coverage (Christoffersen) tests. Failing any of the standard tests flags a forecast whose intervals should be read as breach-calibrated bands rather than as true-VaR coverage. This provides the practitioner with a return-only means of identifying the regime, subject to the usual limitations of backtesting, including finite-sample power and susceptibility to in-sample overfitting of the forecast to the test.

Interpretation. What the calibration layer controls is the breach rate and, through it, the interval width, doing so robustly regardless of forecast quality. What it cannot supply is a correctly located centre. Coverage of the true VaR therefore coincides with the controlled breach rate when the centre is well-specified and separates from it, gracefully and monotonically, as the centre deteriorates. This is the desirable failure mode for a risk interval: rather than becoming confidently narrow and wrong, the method retains a sensibly sized, breach-calibrated band and loses true-VaR coverage only in proportion to the mislocation of the forecast it wraps.

Application to a real bond portfolio

Data. The data consist of daily returns on a portfolio of 24 senior, plain-vanilla fixed-income instruments sourced from Bloomberg, covering October 2014 to October 2024.[2] All bonds are issued by financial institutions in the United States, United Kingdom, France and Italy, with maturities exceeding ten years and denominated in USD, GBP, EUR or JPY. The long duration of the portfolio implies high sensitivity to interest-rate movements, which is particularly relevant to the post-2022 period. An initial window is reserved for estimating the VaR models and for the conformal calibration burn-in.

The sample spans several distinct market regimes: a relatively *calm* period (2016–early 2020) of moderate, stable volatility; the acute *COVID-19 shock* (February–May 2020), one of the sharpest volatility spikes on record; a *post-COVID recovery* through 2021 as policy support stabilised markets; the *2022–2023 monetary tightening*, the most aggressive rate-hiking cycle in decades, compounded by geopolitical stress; and a *recent normalization* as volatility subsided.

Point-VaR forecasts. Three VaR models serve as inputs (forecast centres) to QDtACI: a CAViaR model (Engle and Manganelli 2004), a DCC-GARCH model, and a copula model, with estimation detailed in "Appendix 3", "Appendix 4" and "Appendix 5". QDtACI is applied to each of these under the configuration fixed on the synthetic data ($\kappa = 0$; "The score transformation" section). Table 4 reports the point-VaR backtests together with the resulting interval-width summary. All three forecasts target the nominal 5% level closely and are not rejected by the Kupiec test. The Christoffersen independence test is likewise not rejected for any of the three, though the copula model's p value (0.055) sits close to the conventional threshold, suggesting potential mild residual clustering of violations. The fact that violations may still cluster, even for well-specified forecasts is precisely the motivation for quantifying uncertainty around the point forecast rather than relying on it alone. Figure 4 shows the three forecasts against realized returns: they track one another closely

[2] ACAFP 3.125% 02/05/2026 REGS Corp; BAC 4.25% 12/10/2026 REGS Corp; BAC 5.875% 02/07/2042 Corp; BAC 7.00% 07/31/2028 REGS Corp; C 2.8% 06/25/2027 Corp; C 3.0% 06/26/2037 Corp; C 3.1% 06/26/2047 Corp; C 4.95% 11/07/2043 Corp; C 5.15% 05/21/2026 REGS Corp; C 5.875% 01/30/2042 Corp; C 6.8% 06/25/2038 REGS Corp; C 7.375% 09/01/2039 REGS Corp; C 8.125% 07/15/2039 Corp; GS 6.25% 02/01/2041 Corp; ISPIM 0.00% 01/08/2027 Corp; JPM 2.875% 05/24/2028 REGS Corp; JPM 3.5% 12/18/2026 REGS Corp; JPM 5.4% 01/06/2042 Corp; JPM 6.4% 05/15/2038 Corp; LLOYDS 6.5% 09/17/2040 REGS Corp; MS 6.375% 07/24/2042 Corp; STANLN 4.375% 01/18/2038 REGS Corp; WFC 3.5% 09/12/2029 REGS Corp; WFC 4.625% 11/02/2035 Corp.

Table 4 Point-VaR backtests for the three real-data VaR forecasts, with the summary of the corresponding QDtACI 95% interval widths

Metric	CAViaR	GARCH	Copula
Point-VaR backtests			
Violation rate (nominal 5%)	4.98%	5.02%	4.85%
Kupiec p value	0.965	0.965	0.810
Christoffersen p value	0.264	0.413	0.055
QDtACI interval width (95%)			
Mean width	0.307	0.270	0.278
Median width	0.266	0.238	0.214
Minimum width	0.092	0.071	0.072
Maximum width	0.714	0.631	0.681

Point-VaR backtests are computed on the centres and are independent of the conformal configuration. GARCH produces the narrowest intervals on average, CAViaR the widest

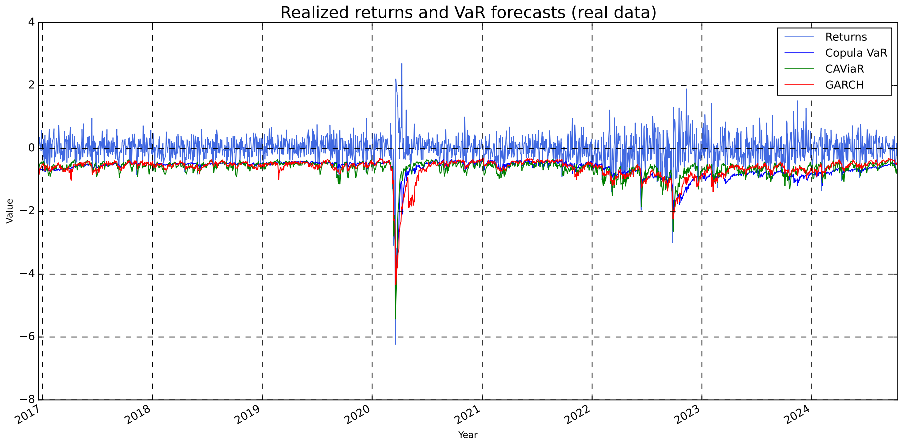

Fig. 4 VaR forecasts from the CAViaR, DCC-GARCH and copula models with realized portfolio returns, 2017–2024

and respond similarly to major events, yet the most extreme realized losses fall below all three, illustrating the residual tail risk that motivates interval construction.

Intervals across regimes. Figures 5, 6 and 7 show VaR forecasts with their 95% QDtACI intervals and realized returns, with market regimes shaded. The intervals behave consistently across all three specifications: narrow in the calm period, widening sharply during the COVID-19 shock, normalizing through the recovery, and widening persistently again through the 2022–23 tightening.

Width adapts to volatility. On real data the true VaR is unobservable; hence PICP and coverage of true VaR cannot be measured. Instead, we assess whether the intervals respond to market conditions. Table 5 reports mean interval width by regime. Width expands significantly in the stress regimes, by a factor of roughly 2.0–2.9 in both the COVID-19 shock and the 2022–23 tightening period, relative to the calm period, while returning toward calm levels as volatility normalizes. Thus, the adaptation runs in both directions. This is not an artefact of the regime partition: across

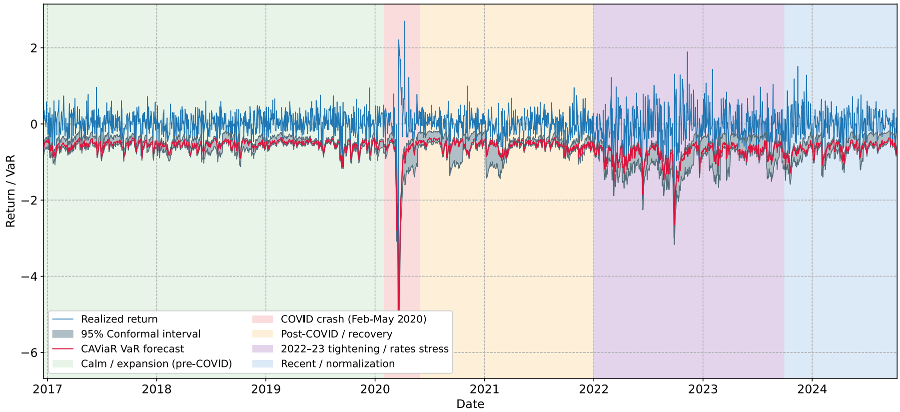

Fig. 5 CAViaR VaR forecast with 95% QDtACI conformal bounds

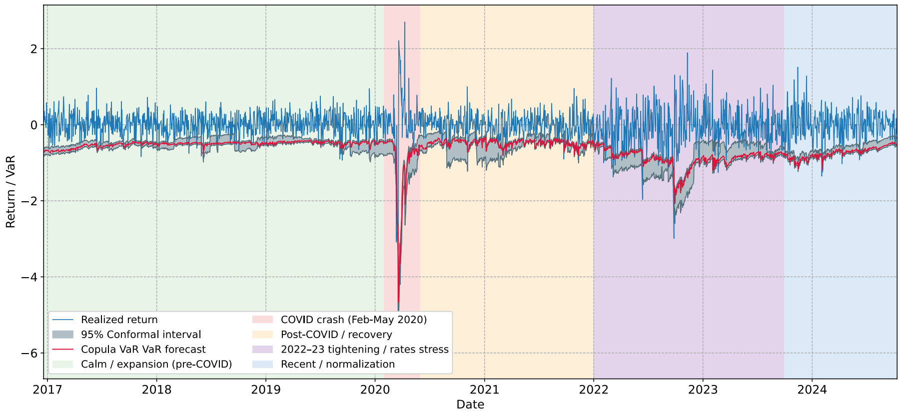

Fig. 6 Copula VaR forecast with 95% QDtACI conformal bounds

the full sample, interval width is positively correlated with realized 30-day volatility for all three forecasts ($r = 0.49$–0.60), day by day and independent of any regime boundary. The intervals therefore respond to genuine changes in market conditions rather than forming a fixed band around the point forecast. The moderate correlation ($r \approx 0.5$) is consistent with width being driven by the dispersion of returns around the forecast, which volatility captures in large part but not entirely.

Overall, the real-data results indicate that QDtACI provides a practically meaningful layer of uncertainty quantification for VaR: the intervals are narrow enough to be informative in calm periods and widen in a manner appropriate with genuine market stress, tracking volatility without being tuned to any particular regime.

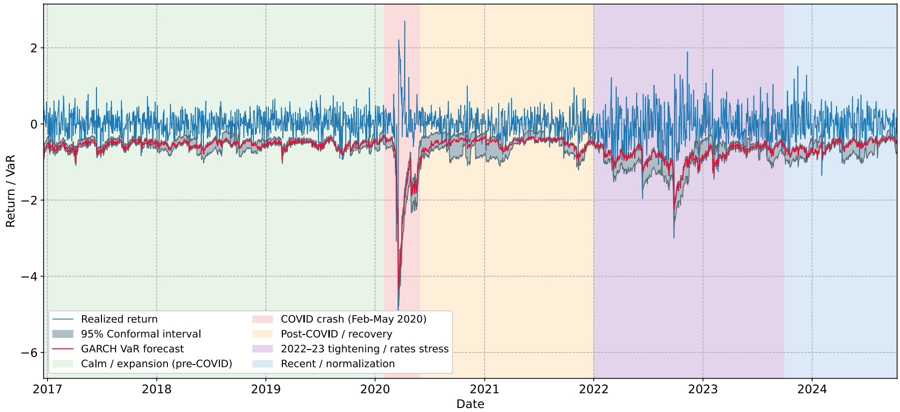

Fig. 7 DCC-GARCH VaR forecast with 95% QDtACI conformal bounds

Table 5 Mean QDtACI interval width by market regime and VaR forecast methodology, with the ratio to the calm regime and the full-sample correlation between width and realized 30-day volatility

Regime	N	Copula	CAViaR	GARCH
Calm/expansion (pre-COVID)	1065	0.187	0.204	0.177
COVID crash (Feb–May 2020)	85	0.539	0.496	0.470
Post-COVID/recovery	415	0.304	0.297	0.331
2022–23 tightening	455	0.420	0.465	0.346
Recent/normalization	270	0.193	0.306	0.267
Full sample	2290	0.278	0.307	0.270
Ratio, COVID/calm		2.9	2.4	2.7
Ratio, tightening/calm		2.2	2.3	2.0
$\mathrm{corr}(\mathrm{width},\ \sigma_{30})$		0.54	0.60	0.49

Discussion and conclusion

This study develops Quantile Dynamically-tuned Adaptive Conformal Inference (QDtACI), a model-agnostic conformal calibration layer for Value-at-Risk. The motivation for QDtACI is twofold. On one hand, the parametric interval methods established in the literature, e.g. the Delta method, bootstrap, asymptotic, and Bayesian approaches, are each tied to a particular model class or set of assumptions, and become unreliable when those assumptions fail. That is, they are not applicable across the full range of VaR forecasts used in practice, while for several common forecast methods no analytical intervals exist at all. On the other hand, the natural distribution-free alternative, adaptive conformal inference in the form of DtACI, is designed for mean regression: it centres intervals on a conditional-mean forecast and calibrates with symmetric nonconformity scores, penalizing over and under-estimation of the risk threshold equally and producing excessively wide intervals when applied to a one-sided tail quantile. QDtACI is designed to resolve both limitations at once: a model-agnostic calibration layer, applicable to any VaR forecast, that is at the same time aligned with the quantile nature of the target through.

On synthetic data, where the true VaR is observable, QDtACI delivers near-nominal coverage of the true VaR when the underlying forecast is well-specified, using a single procedure across structurally different centres. Its coverage degrades in a controlled and interpretable way as the forecast becomes misspecified: the calibration layer absorbs estimation error through the nonconformity-score distribution, but it cannot compensate for a fundamentally misspecified point forecast. That is, coverage of the true VaR is inherited from, and bounded by, the quality of the forecast. This dependence on forecast quality is fundamental: reliable coverage of the true VaR requires a well-specified point forecast, and this is as true for QDtACI as for the parametric alternatives. What distinguishes QDtACI is not that it escapes this requirement (without an effective way to calculate the difference to real VaR - no calibration layer can) but that it applies to any forecast, including historical simulation and copula constructions for which the Delta method and asymptotic intervals cannot be defined. Against these benchmarks and the DtACI baseline, QDtACI attains coverage closer to nominal, at comparable or better interval quality as measured by the Winkler score.

There is an underlying idea and distinction that runs through the analysis, and it is worth making it explicit here. The calibration layer controls an observable quantity, the rate at which returns breach the interval's lower bound, whereas coverage of the true VaR is latent and cannot be measured on real data. Because the two align only when the forecast is well-specified, we complement the intervals with a return-only backtest diagnostic that indicates, from observable data alone, the regime in which interval calibration can be trusted as a proxy for true-VaR coverage.

In the real-data application to a portfolio of 24 fixed-income assets over the period from 2016 to 2024, the intervals respond meaningfully to market conditions across all three VaR models. They are narrow in calm periods and widen sharply during the COVID-19 shock, by a factor of roughly two to three times relative to calm levels, remaining elevated through the 2022–2023 tightening and geopolitical stress. Given the portfolio's long duration, its sensitivity to rate movements was pronounced in this period, and the sustained interval width reflects genuine forecast uncertainty in a rate environment without precedent in the calibration sample. Across regimes, the dominant driver of width is market-wide volatility rather than model-specific structure: interval width correlates with realized volatility for all three forecasts, and the three structurally different forecasts produce broadly similar widths, consistent with the calibration layer responding to the realized dispersion of returns rather than to the internal form of the forecast.

A natural question is how QDtACI relates to the coherence of the underlying risk measure, given that VaR is not subadditive. Subadditivity is a property of the risk *measure*, whereas QDtACI is a calibration layer operating on a single series of portfolio returns and the corresponding univariate series of VaR forecasts. It does not aggregate positions, define a new functional of the loss distribution, or alter the risk measure. Coherence and interval calibration are therefore separate concerns. Aggregation across positions is handled by the underlying VaR models, while QDtACI quantifies forecast uncertainty in a single quantile series. As with any interval method, QDtACI neither inherits nor repairs the non-subadditivity of VaR: that property belongs to the input model producing the point forecast and persists unchanged whether or not a conformal interval is placed around it. There is also a reason the method targets VaR specifically. Conformal calibration needs a target that can be scored by a consistent loss function, and VaR has one: the pinball loss, which is minimised exactly at the quantile (Gneiting 2011). Expected Shortfall, the coherent alternative, has no such standalone loss, thus cannot be calibrated this way. It can, however, be scored jointly with VaR (Fissler and Ziegel 2016), which points to a natural extension of our approach toward a coherent risk measure, a direction we leave to future work.

Two limitations bound the scope of our conclusions. First, the empirical application is confined to a single portfolio of fixed-income assets. While the model-agnostic construction means QDtACI is not tailored to any particular asset class, as it wraps any VaR forecast using only returns and the forecast series, we do not empirically verify its behavior on other classes. Replication across equity, commodity, or cryptocurrency portfolios could provide stronger evidence of general applicability. Second, coverage of the true VaR is verifiable only on synthetic data. On real data, true coverage cannot be measured directly. The calibration is carried over from the synthetic setting, and the variation in coverage we observe across the three forecasts provides some evidence on how sensitive this transfer is, though it cannot confirm that the calibration holds exactly for the real return distribution.

Beyond broader empirical validation, two directions are promising. The joint elicitability of $(\mathrm{VaR}, \mathrm{ES})$ as noted above suggests an extension of conformal framework targeting a coherent risk measure. Separately, the observation that interval width tracks forecast uncertainty raises the question of whether widening intervals could serve as an early-warning signal of impending VaR-model failure, preceding formal rejection under Kupiec or Christoffersen backtests, an extension of direct relevance to risk managers operating under regulatory capital requirements.

Appendix 1: Pinball loss and quantile estimation

Definition For a quantile level $a \in (0, 1)$, the *pinball loss* (or *quantile loss*) for a prediction q and observation Y is defined as:

$$L_a(q, Y) = \begin{cases} a(Y - q), & \text{if } Y \geq q, \\ (1 - a)(q - Y), & \text{if } Y < q. \end{cases}$$

A compact form is

$$L_a(q, Y) = \big(a - \mathbb{I}\{Y < q\}\big)(Y - q).$$

This loss function was introduced by Koenker and Bassett (1978) and forms the basis of quantile regression.

Population (theoretical) case. Assume Y has density f and cumulative distribution function F. We seek the value of q minimizing the expected loss:

$$\min_q \mathbb{E}[L_a(q, Y)].$$

Expanding the expectation,

$$\mathbb{E}[L_a(q,Y)] = \int_{-\infty}^{q} (1-a)(q-y)f(y)\,dy + \int_{q}^{\infty} a(y-q)f(y)\,dy.$$

Differentiating with respect to q:

$$\frac{d}{dq}\mathbb{E}[L_a(q,Y)] = (1-a)\int_{-\infty}^{q} f(y)\,dy - a\int_{q}^{\infty} f(y)\,dy.$$

Since

$$\int_{-\infty}^{q} f(y)\,dy = F(q) \quad \text{and} \quad \int_{q}^{\infty} f(y)\,dy = 1-F(q),$$

we have

$$\frac{d}{dq}\mathbb{E}[L_a(q,Y)] = F(q) - a.$$

Setting this derivative to zero gives the minimizer:

$$F(q^*) = a.$$

Thus, the solution q^* is the *a*–quantile of *Y*:

$$q^* = \inf\{q \in \mathbb{R} : F(q) \geq a\}.$$

Empirical (sample) case. Given a sample $\{\, Y_1, \ldots, Y_n \,\}$, define the empirical risk:

$$R_n(q) = \frac{1}{n}\sum_{i=1}^{n} L_a(q, Y_i).$$

This function is piecewise linear and convex in *q*, with kinks at the sample points y_i . Any minimizer $\hat{q}_a$ satisfies:

$$\#\{Y_i \leq \hat{q}_a\} \geq an \quad \text{and} \quad \#\{Y_i \geq \hat{q}_a\} \geq (1-a)n.$$

A common choice of minimizer is the order statistic $Y_{(\lceil an \rceil)}$ (or any value between adjacent order statistics when the minimizer is not unique).

Conclusion: Minimizing the pinball loss yields the theoretical (population) *a*–quantile for *Y*, and its sample version yields an empirical *a*–quantile for data $\{Y_i, ..., Y_n\}$.

Appendix 2: Quantile equivariance under strictly increasing transformations

We now show that empirical quantiles and equivalently, minimizers of the pinball loss– are preserved under any strictly monotone transformation of the data in finite samples. This equivariance property of quantiles is well known in the literature (see, e.g., Koenker 2005; Serfling 1980).

Theorem Appendix 2.1 *(Quantile Equivariance) Let $S = \{s_1, \ldots, s_m\}$ be a finite sample and let $T : \mathbb{R} \to \mathbb{R}$ be strictly increasing. Let $s_{(1)} \leq \cdots \leq s_{(m)}$ denote the order statistics of S, and define $S' = T(S) = \{T(s_1), \ldots, T(s_m)\}$ with order statistics $s'_{(i)}$. Define the empirical a–quantile by $\hat{Q}_S(a) := s_{(\lceil am \rceil)}$. Then for any $a \in (0, 1)$,*

$$\hat{Q}_{S'}(a) = T\big(\hat{Q}_S(a)\big).$$

Proof Since T is strictly increasing, it preserves order: $s_i \leq s_j \iff T(s_i) \leq T(s_j)$. Hence $s'_{(i)} = T(s_{(i)})$ for all i. Let $k = \lceil am \rceil$. Then

$$\hat{Q}_{S'}(a) = s'_{(k)} = T(s_{(k)}) = T\big(\hat{Q}_S(a)\big).$$

□

Corollary Appendix 2.2 *(Pinball-loss Minimizer Equivariance) Let L_a be the pinball loss. Then*

$$\operatorname{argmin}_q \sum_{i=1}^{m} L_a\big(q, T(s_i)\big) = \{\, T(u) : u \in \operatorname{argmin}_q \sum_{i=1}^{m} L_a(q, s_i) \,\}.$$

Proof Minimizers of $\sum_i L_a(q, s_i)$ coincide with empirical a–quantiles of the sample. Applying the theorem to S and S' maps empirical a–quantiles via T, hence the corresponding argmin sets are mapped via T. □

Remark Appendix 2.1 If T is only non-decreasing, ties may be introduced, and the empirical quantile (and hence the argmin) may be non-unique, but the set of valid quantiles remains consistent with the induced rank structure.

Appendix 3: CAViaR model

One of the models use in a process of VaR modeling is CAViaR (the Conditional Autoregressive Value at Risk) proposed by Engle and Manganelli (2004). This model directly estimates quantile instead of assuming specific parametric distribution. As such, it utilizes quantile regression technique to determine the evolution of VaR over time. The process is summarized in the following steps below.

Using the notation from the article of Engle and Manganelli (2004) let y_t be the portfolio returns at time *t*, and VaR_t represent the estimated Value at Risk. Hence, we have the probability that the return will fall below VaR_t at a given confidence level θ as:

$$P(y_t < -VaR_t|\mathcal{F}_{t-1}) = \theta,$$

where $\mathcal{F}_{t-1}$ denotes the information set available at time $t-1$.
The general specification of the Asymmetric Slope CAViaR model is:

$$VaR_t = \beta_1 + \beta_2 VaR_{t-1} + \beta_3 \max(Y_{t-1}, 0) + \beta_4 \min(Y_{t-1}, 0),$$

where:

- β_1 represents the intercept term,
- β_2 captures the persistence of VaR,
- β_3 models the impact of positive returns at time $t-1$, and
- β_4 models the impact of negative returns at time $t-1$.If we have a sample of T observations $\{Y_1, Y_2, \ldots, Y_T\}$ generated by the model:

$$Y_t = X_t\beta_0 + \varepsilon_{\theta t}, \quad \text{where} \quad \text{Quant}_\theta(\varepsilon_{\theta t}|X_t) = 0.$$

where X_t is a p-dimensional vector of regressors, and $\text{Quant}_\theta(\varepsilon_{\theta t}|X_t)$ represents the conditional θ-quantile of $\varepsilon_{\theta t}$ conditional on X_t. Then, the parameters of the model $\beta_1, \beta_2, \beta_3$, and β_4 are estimated by minimizing the quantile loss function, introduced by Koenker and Bassett (1978) as:

$$\min_\beta \frac{1}{T} \sum_{t=1}^{T} [\theta - \mathbb{I}(Y_t < f_t(\beta))] [Y_t - f_t(\beta)]$$

where:

- $f_t(\beta) \equiv x_t\beta$.
- Y_t is the portfolio return at time t.
- θ is the quantile (0.05 for a 5% quantile/VaR).
- $\mathbb{I}(Y_t < f_t(\beta))$ is an indicator function that takes the value 1 if Y_t is less than $f_t(\beta)$ or 0 otherwise.

Appendix 4: Dynamic conditional correlation GARCH (DCC-GARCH) model

This section briefly discusses the Dynamic Conditional Correlation GARCH (DCC-GARCH) Model.

Univariate GARCH(p,q) model

Each asset is modelled with generalised autoregressive conditional heteroscedasticity, GARCH (p,q) model as (tested for best model based on AIC and BIC criteria up to p=4 and q=4):

$$\sigma_{t,i}^2 = \omega + \alpha r_{t-1,i}^2 + \beta \sigma_{t-1,i}^2$$

where $r_{t,i}$ is return of asset *i* at time *t*, $\sigma_{t,i}^2$ is conditional variance of $r_{t,i}$, and ω, α, β are model parameters.

Dynamic conditional correlation (DCC) model

The dynamic conditional covariance matrix is estimated as:

$$Q_t = (1 - a - b)\bar{Q} + a\epsilon_{t-1}\epsilon_{t-1}^\top + bQ_{t-1}$$

where:

- Q_t: Time-varying covariance matrix,
- $\bar{Q}$: Unconditional covariance matrix of standardized residuals,
- ϵ_{t-1}: Standardized residuals from the GARCH model,
- a, b: DCC parameters.Which is then converted to a time-varying correlation matrix as:

$$R_t = \text{diag}(Q_t)^{-1/2} Q_t \text{diag}(Q_t)^{-1/2}$$

Time-varying covariance matrix

The conditional covariance matrix is calculated as:

$$H_t = D_t R_t D_t$$

where D_t is a diagonal matrix of conditional volatilities.

Value-at-Risk (VaR)

To estimate VaR at confidence level α, a Monte Carlo simulation with 10,000 scenarios is conducted, deriving with the standard normal returns Z_t. These samples are transformed using the estimated covariance matrix H_t to obtain simulated portfolio returns:

$$r_t^{\text{sim}} = w^\top L Z_t$$

where L is the lower-triangular Cholesky factor of H_t, such that $H_t = LL^\top$.

Subsequently, the empirical quantile of the simulated portfolio returns is computed to estimate VaR:

$$\mathrm{VaR}_{\alpha,t} = \mathrm{Quantile}_\alpha(r_t^{\mathrm{sim}})$$

This approach is, thus, a data-driven and distribution-free estimation of VaR, capturing non-linear dependencies and fat-tailed behavior in returns.

Appendix 5: Copulas modelling for value-at-risk

This section outlines the Copula-based approach used in the study. The overall process involves fitting GARCH models to individual assets, transforming residuals into a uniform space, estimating a Student-t Copula, generating Monte Carlo simulations, and computing portfolio VaR.

Fit marginal distributions

Before applying the Copula, the volatility of each asset is modeled using a GARCH process, similar as in the previous segment.

From this, standardized residuals are obtained:

$$z_{t,i} = \frac{r_{t,i} - \mu}{\sigma_{t,i}}$$

where μ is the mean return and $\sigma_{t,i}$ is the conditional volatility.

And transformed into a uniform distribution $[0, 1]$ using the cumulative distribution function (CDF) of the Student-t distribution $U_{t,i} = F_t(z_{t,i})$, where F_t is the Student-t CDF.

Fit the Student t copula

The dependence structure between assets is modeled using the Student-t Copula, as suggested by the literature regarding the financial returns: The Copula captures dependencies by estimating:

- The correlation matrix of the transformed residuals,
- The degrees of freedom parameter for tail dependence.The dynamic correlation structure follows $Q_t = (1 - a - b)\bar{Q} + a\epsilon_{t-1}\epsilon_{t-1}^\top + bQ_{t-1}$ where Q_t is the time-varying correlation matrix and $\bar{Q}$ is the unconditional correlation matrix.

Monte Carlo sampling from the Copula

Using the fitted Copula, synthetic asset returns are generated using 10,000 Monte Carlo simulations. The process involves:

- Generating correlated uniform samples $U_{t,i}$ from the Student-t copula.
- Transforming uniform samples into Student-t distributed residuals:

$$t_{\mathrm{sim}} = \frac{Z}{\sqrt{\chi^2_{\mathrm{df}}/\mathrm{df}}}$$

- And convert residuals into real-world returns using the estimated GARCH volatility.*Transforming Simulated Returns Back to Real-World Space*

The simulated Copula samples are transformed back into return space

$$R_t^{\text{sim}} = \mu + \sigma_t \cdot Z_t^{\text{sim}},$$

where σ_t is the GARCH-estimated volatility.

Finally, VaR is computed as the empirical quantile of the simulated returns:

$$\text{VaR}_{\alpha,t} = \text{Quantile}_\alpha(r_t^{\text{sim}}).$$

Author contributions M.I. developed the QDtACI methodology, conducted all empirical analyses, implemented the computational framework, and wrote the main manuscript text. K.A.N. and Z.L. provided conceptual guidance, contributed to the theoretical framework, and reviewed and revised the manuscript. All authors approved the final version.

Funding Funding was provided by British Scholarship Trust.

Data availability The real-world data used in this study consist of daily returns on 24 fixed-income corporate bonds sourced from Bloomberg Terminal. Access to Bloomberg data requires an institutional or commercial subscription and cannot be made publicly available by the authors. The specific bond identifiers are listed in the manuscript footnote, allowing readers with Bloomberg access to replicate the dataset. The synthetic data used in the study are generated from a GARCH(1,1) process with parameters fully specified in the manuscript $(\alpha = 0.10, \beta = 0.85)$, and can be reproduced exactly using the data-generating procedure described in "Synthetic experimental design" section. The code implementing the QDtACI methodology and the synthetic data experiments is available from the corresponding author upon reasonable request.

Code availability An anonymized reference implementation of QDtACI is available at https://anonymous.4open.science/r/QDtACI-9D0F/. The repository contains the core method and the scripts that reproduce the reported statistics and tables. The full repository, with citation details, will be released publicly upon acceptance.

Declarations

Conflict of interest The authors declare no conflict of interests.

References

Basel Committee on Banking Supervision. 2016. *Minimum capital requirements for market risk*. Standards d352, Bank for International Settlements, Basel, Switzerland. Available at https://www.bis.org/bcbs/publ/d352.pdf.

Beutner, E., A. Heinemann, and S. Smeekes. 2024. A residual bootstrap for conditional value-at-risk. *Journal of Econometrics* 238 (2): 1–16.
Casella, G., and R. L. Berger. 2002. *Statistical inference*, 2nd ed. Duxbury Press.
Chen, J. M. 2014. Measuring market risk under the Basel accords: VaR, stressed VaR, and expected shortfall. *Aestimatio: The IEB International Journal of Finance* 8:184–201.
Christoffersen, P., and S. Gonçalves. 2004. Estimation risk in financial risk management. Technical report, CIRANO.
Christoffersen, P., and S. Gonçalves. 2005. Estimation risk in financial risk management. *Journal of Risk* 7(3):1–28.
Contreras, P., and A. Bhattacharjee. 2004. *A Bayesian confidence interval for value-at-risk*. Cambridge working papers in economics 0348, faculty of economics, University of Cambridge.
Danielsson, J. 2011. *Financial risk forecasting: The theory and practice of forecasting market risk with implementation in R and Matlab*, vol. 588. Wiley.
Duan, J.-C. 1994. Maximum likelihood estimation using price data of the derivative contract. *Mathematical Finance* 4 (2): 155–167.
Engle, R. F., and S. Manganelli. 2004. Caviar: Conditional autoregressive value at risk by regression quantiles. *Journal of Business & Economic Statistics* 22 (4): 367–381.
Fissler, T., and J. F. Ziegel. 2016. Higher order elicitability and Osband's principle. *The Annals of Statistics* 44 (4): 1680–1707.
Francq, C., and J.-M. Zakoian. 2004. Maximum likelihood estimation of pure GARCH and ARMA-GARCH processes. *Bernoulli* 10 (4): 605–637.
Gibbs, I., and E. Candès. 2021. Adaptive conformal inference under distribution shift. In *Advances in Neural Information Processing Systems (NeurIPS)*, vol. 34, 1660–1672.
Gibbs, I., and E. J. Candès. 2024. Conformal inference for online prediction with arbitrary distribution shifts. *Journal of Machine Learning Research* 25 (162): 1–36.
Gneiting, T. 2011. Making and evaluating point forecasts. *Journal of the American Statistical Association* 106 (494): 746–762.
Goh, Y.-L., and A.-H. Pooi. 2013. Confidence intervals for multivariate value at risk. *ScienceAsia* 39S:70–74.
Gradu, P., E. Hazan, and E. Minasyan. 2023. Adaptive regret for control of time-varying dynamics. In *Proceedings of The 5th annual learning for dynamics and control conference*, vol. 211, 560–572. PMLR.
He, Y., Y. Hou, L. Peng, and H. Shen. 2020. Inference for conditional value-at-risk of a predictive regression. *Annals of Statistics* 48 (6): 3442–3464.
Koenker, R. 2005. *Quantile regression*. Cambridge University Press.
Koenker, R., and G. Bassett. 1978. Regression quantiles. *Econometrica* 46 (1): 33–50.
Lahiri, S. N. 2003. *Resampling methods for dependent data*. New York: Springer.
McNeil, A. J., R. Frey, and P. Embrechts. 2015. *Quantitative risk management: Concepts, techniques and tools*. Revised. Princeton: Princeton University Press.
Pritsker, M. 2006. The hidden dangers of historical simulation. *Journal of Banking & Finance* 30 (2): 561–582.
Serfling, R. J. 1980. *Approximation theorems of mathematical statistics*. New York: Wiley.
van der Vaart, A. W. 1998. *Asymptotic statistics*. Cambridge: Cambridge University Press.
Vovk, V., A. Gammerman, and G. Shafer. 2005. *Algorithmic learning in a random world*. Springer.
Vovk, V.G. 1990. Aggregating strategies. In *Proceedings of the third annual workshop on Computational Learning Theory (COLT '90)*, 371–386. San Francisco: Morgan Kaufmann Publishers Inc.
Wang, H. 2003. Simulation and extreme VaR and VaR confidence interval estimation for a class of heavy-tailed risk factors. *Chinese Journal of Applied Probability and Statistics* 19 (3): 267–276.
Zaffran, M., A. Dieuleveut, O. Féron, Y. Goude, Josse, and J. 2022. Adaptive conformal predictions for time series. arXiv preprint arXiv:2202.07282.

Authors and Affiliations

Milo Ivancevic[1,2] · Khuong An Nguyen[2] · Zhiyuan Luo[2]

✉ Milo Ivancevic
miloivancevic@gmail.com

Khuong An Nguyen
Khuong.Nguyen@rhul.ac.uk

Zhiyuan Luo
Zhiyuan.Luo@rhul.ac.uk

[1] University of Belgrade, Belgrade, Serbia

[2] Royal Holloway University of London, Egham, UK