Reliable Value at Risk Estimations with Conformal Prediction

Milo Ivancevic[1][0009−0005−2455−8812], Khuong An Nguyen[2][0000−0001−6198−9295], and Zhiyuan Luo[2][0000−0002−3336−3751]

[1] Faculty of Economics, University of Belgrade, Serbia
miloivancevic@gmail.com
[2] Computer Science Department, Royal Holloway University of London, UK
{Khuong.Nguyen, Zhiyuan.Luo}@rhul.ac.uk

Abstract. This paper introduces a novel algorithm to enhance the reliability and interpretability of Value at Risk (VaR) estimates in financial markets, using Conformal Prediction. We address the issue of lack of general mechanism to quantify uncertainty of VaR estimates, especially under volatile market conditions. This is done by employing Adaptive Conformal Inference (ACI) methodology on both synthetic and real data. Two ACI techniques, Aggregation-based ACI (AgACI) and Dynamically-tuned ACI (DtACI) suggested that Conformal Prediction can successfully construct valid predictive intervals around VaR forecasts. At 95% confidence level, DtACI achieved consistently narrower prediction intervals than AgACI across all models, with up to a 33% reduction in median width while maintaining nominal coverage. Additionally, we demonstrate that these intervals dynamically adapt to changes in market volatility, widening during periods of financial stress, such as the COVID-19 crisis and the 2022 geopolitical shocks.

Keywords: Value at Risk · Conformal Prediction · GARCH · Copulas · CAViaR.

1 Introduction

VaR is defined as a quantifiable threshold, indicating the worst expected loss under normal market conditions at a given confidence level [4]. While it is relatively simple to comprehend what it is, VAR is not straightforward to calculate. It generally involves predicting an unknown parameter of the return distribution, providing a unique statistical challenge. In the real world, calculating VaR on different sample sizes and different volatilities could introduce unavoidable errors. For this reason, relying solely on point estimates could be misleading, and one needs also to consider interval estimates around this point prediction [12]. One possible solution is to define a lower confidence bound for VaR. However, providing a reliable prediction confidence in statistics has proven to be a significant challenge.

Thus, in this paper, we adopt a Machine Learning approach to estimate the confidence levels of VaR. Specifically, we employ conformal prediction, a powerful technique which quantifies predictive uncertainty while making minimal assumptions [15]. Notably, it offers finite-sample coverage guarantees, a highly desirable property in any uncertainty quantification framework.

Next section presents a literature review, outlining traditional approaches for estimating the uncertainty around VaR and highlighting their limitations. Section 3 introduces the theoretical background of conformal prediction, with a particular focus on its application in time series and risk estimation settings. Section 4 provides an overview of the main empirical results obtained using both synthetic and real-world financial data. Finally, Section 5 concludes the paper.

2 Literature Review

Usually, estimation of VaR is done on the basis of parameters that are derived from the distribution of returns. However, it is also associated with the uncertainty of those estimates, which can be circumvented with the construction of confidence intervals. There are several methods of uncertainty estimation of VaR proposed by the literature. One of those is the Delta method, which is the mathematical approach to approximating variance [5]. Another method is the bootstrap method, relying on repeating the sample in a simulation process to estimate uncertainty (e.g. [2]), or a method that relies on analytical confidence intervals constructed around the VaR, based on its asymptotic distribution [7]. Goh et al. [10] examined different methods for VaR confidence intervals. Their findings suggest that the hypothesis testing approach yields the most reliable results, outperforming normal approximation or bootstrap method. However, one of the key limitations of this method is that it still relies on specific distributional assumptions, which may not fully capture the complexities of the financial market, especially during periods of extreme downturns and volatile conditions. Hao [16] uses sophisticated predictive models to enhance the accuracy of VaR estimates. However, they noted that these models are associated with significant computational complexity, making them impractical for the real-time applications, while it is not clear how they perform under extreme market conditions. Also, Contreras et al. [3] employed Bayesian confidence intervals. However, one of the biggest weaknesses of this model is that they require strong prior assumptions. Thus, while there are several methods in existence related to the uncertainty of VaR estimates, they come with significant limitations. They usually rely on the assumption that model errors follow a normal distribution, but this assumption is not always satisfied. Recent work has addressed confidence intervals around conditional VaR, though often with methodological constraints. He et al. [11] use a residual bootstrap under fixed-design assumptions, while the Beutner et al. [1] paper leverages metamodeling but may struggle with tail representation.

In addition, some of the most widely accepted models for VaR estimation, like GARCH or CAViaR, while effective for point prediction, lack a general framework for constructing valid predictive intervals. While literature on such interval construction exists, it is often highly model-specific, analytically complex, and difficult to generalize across different market regimes. When the underlying assumptions of the interval prediction do not hold, these methods become highly unreliable.

3 Conformal Prediction for Value at Risk estimations

Conformal Prediction (CP), initially developed by Vovk et al. [15], offers a principled, efficient, and flexible way to obtain predictions that guarantee a given error rate under minimal assumptions. The basic idea of CP is to predict a label to the given test observation based on past experience. The framework of CP rests on the notion of nonconformity measure (NCM), a score function $S(X, y) \in \mathbb{R}$. The NCM expresses how much it appears not to conform to a collection of samples. Larger scores indicate worse agreement between X and y. Typically, NCM is designed based on the underlying point prediction model. Given a set of n observed training data $(X_1, y_1), ..., (X_n, y_n)$ and a new test point (X_{n+1}, y_{n+1}) which we only observed X_{n+1}, where $(X_i, y_i) \in \mathbb{R}^d \times \mathbb{R}$ for $i = 1, ..., n+1$, and a point prediction model $\hat{\mu}(X)$, the common NCM for a regression task can be defined as $S_i = |y_i - \hat{\mu}(X_i)|$ for the observation (X_i, y_i).

For any specific miscoverage rate $\alpha \in (0, 1)$, a prediction interval produced by the CP is

$$\hat{C}_\alpha(X_{n+1}) = [\hat{\mu}(X_{n+1}) - q_\alpha(S), \quad \hat{\mu}(X_{n+1}) + q_\alpha(S)]$$

where $q_\alpha(S)$ is defined as the $\lceil (1-\alpha)(n+1) \rceil$ smallest of $S = \{S_1, ..., S_n\}$. This prediction set satisfies the following coverage guarantee:

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

if the observed training data and the new test point are exchangeable [15]. In other words, the true value y_{n+1} falls within the interval $\hat{C}_\alpha(X_{n+1})$ at least $100(1-\alpha)\%$ of the time, averaged over all instances.

The above validity property (Equation (1)) is for any (possibly uninformative) NCM function and (possibly unknown) distribution of the data. Although the guarantee always holds, the usefulness of the prediction sets is primarily determined by the NCM function. Roughly, if the NCM S_i correctly ranks the inputs from lowest to highest magnitude of model error, then the resulting sets $\hat{C}_\alpha(X_{n+1})$ will be smaller for easy inputs and bigger for hard ones. If the NCMs are bad in the sense that they do not approximate this ranking, then the sets will be useless.

However, in the settings with many real-world time series data, where the observed $(X_1, y_1), ...,\ (X_t, y_t), (X_{t+1}, y_{t+1})$ exhibit distributional shift and temporal dependence, this might violet the assumption of data exchangeability and require the development of adaptive techniques to handle such complexities. As a result, various studies have explored the application of CP to time series data using randomisation, ensembles, and other meta-algorithms to produce valid prediction sets [17, 9]. In this paper, the popular Adaptive Conformal Inference proposed by Gibbs et al. [9] is considered.

3.1 Adaptive Conformal Inference (ACI)

Instead of using a fixed miscoverage level at each time step, ACI can dynamically adjust the miscoverage level α and the prediction interval width in order to accommodate observed data and achieve the required guarantee.

At each time step t, ACI calculates the prediction interval $C_{\alpha_t}(X_t)$ according to α_t and then tries to estimate the optimal α_t^* in an online fashion by updating α_t

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

where $\gamma > 0$ is learning rate (step size) and $\mathbb{I}$ is the indicator function taking the value 1 if $Y_t \notin \hat{C}_{\alpha_t}(X_t)$ and 0 otherwise. If the prediction interval fails to cover Y_t, then $\alpha_{t+1} < \alpha_t$, which increases the interval size. On the other hand, if the prediction interval successfully covers Y_t, then $\alpha_{t+1} > \alpha_t$, which shrinks the interval size. This adaptive mechanism ensures that the model dynamically adjusts to shifts in the data distribution. Thus, ACI ensures that a CP is asymptotically valid, even if the data is not exchangeable.

The learning rate γ controls the speed at which the prediction interval width adapts to the observed data and is the primary tuning hyper-parameter.

Unfortunately, the optimal choice of γ requires an in-depth knowledge of the distribution shift a priori. Furthermore, the size of the distribution shift can change over time, and a single fixed γ value can perform poorly. Two approaches have been developed recently to enhance ACI by removing the need for manual tuning γ, making them highly adaptive to non-stationary data distribution: Online Expert Aggregation on ACI by Zaffran et al. [18] and Dynamically tuned ACI by Gibbs et al. [8].

Online Expert Aggregation for ACI (AgACI). The AgACI method of Zaffran et al. [18] aims to learn the value of γ in ACI and introduces an *adaptive expert aggregation* approach. Instead of selecting a fixed learning rate γ, AgACI dynamically aggregates K multiple ACI models with different values of γ_k using an online learning framework. These parallel sequences, in convex optimization literature, are commonly referred to as experts [8].

Dynamically tuned ACI (DtACI). To enhance ACI, Dynamically-Tuned Adaptive Conformal Inference (DtACI). considers the ACI update as a gradient descent step concerning the pinball loss [8]. This method introduces a self-adjusting mechanism for the learning rate γ. Similarly to AgACI, DtACI selects the optimal γ by leveraging an online expert aggregation approach. However, DtACI tends to be more dynamic, as it continuously updates K expert weights using an exponential re-weighting mechanism based on the historical performance of a set of candidate values, thus, allowing it to adapt more quickly to distribution shifts and abrupt changes in the data distribution, which is often observed in the case of financial data.

3.2 Application of ACI for Value at Risk

Value at Risk, as a measure of the downside risk of an asset or portfolio at a certain confidence level, has certain peculiarities that require some adjustment to this methodology. Namely, first of all, in order to properly measure the downside risk and nonconformity to the measure of VaR, the adjusted quantile score function was used as nonconformity score, as below [14]:

$$S(q_t(\alpha), y_t) = (\alpha - \mathbb{I}(y_t \leq q_t(\alpha)))(y_t - q_t(\alpha)) \tag{2}$$

where $q_t(\alpha)$ represents the **quantile forecast** at probability level α (in this case, it represents the modeled VaR), y_t represents the **realized value** at time t, $S(q_t(\alpha), y_t)$ is the **quantile scoring function**, which measures the accuracy of the quantile estimate $q_t(\alpha)$, and $\mathbb{I}(y_t \leq q_t(\alpha))$ is the **indicator function**, taking the value 1 if $y_t \leq q_t(\alpha)$ and 0 otherwise. That is, this function determines whether the observed value y_t falls below the predicted quantile $q_t(\alpha)$.

In summary, Equation (2) could be divided into the following two terms:

- The term $(\alpha - \mathbb{I}(y_t \leq q_t(\alpha)))$ acts as a weight, assigning different penalties depending on whether the observation is below or above the quantile forecast. For the violations, it assigns the value of 0.95, while for nonviolations, it assigns the values of 0.05, in case of α=5% VaR.
- The term $(y_t - q_t(\alpha))$ measures the difference between the actual and predicted values, ensuring that the score reflects the magnitude of any prediction error.

Having in mind that we are interested only in the downside risk, this equation is modified to capture only violations with the weight of 1 (leaving out the non-violation returns). Thus, with an attempt to isolate the violations, the above scoring function is modified to be:

$$S^*(q_t(\alpha), y_t) = (-\mathbb{I}(y_t \leq q_t(\alpha)))(y_t - q_t(\alpha))$$

This modification allows us to have more interpretable results (e.g. the ACI captures the downside volatility at certain confidence level), and to easily track the quality of the VaR estimates. Consequently, the bounds for the final prediction interval is constructed as:

$$\hat{C}_{\text{ACI},t} = [q_t(\alpha) - L_t(\alpha), q_t(\alpha)]$$

where $L_t(\alpha)$ represents the width for the lower bound constructed at time t.

4 Empirical Results

This section presents the results of adopting CP on the synthetic data and a real-world data to calculate VaR.

4.1 Synthetic Data

Synthetic return series are generated using a GARCH(1,1)-style model[3]. We consider three models: GARCH (1,1), Historical VaR (which estimates the values of 5% quantile in the rolling window), and Trivial VaR (which calculates the 5% quantile from an initial period and uses it as a constant VaR estimate, and their results are shown in Table 1.

While none of these models are perfect, GARCH(1,1) model seems to satisfy the main backtests (Kupiec test and Christoffersen test for independence) in all market regimes. Trivial VaR satisfies these tests for the baseline model, Mean-Reverting markets and Highly Volatilite markets; while at 5% significance level, Historical model satisfies the tests only for Mean-Reverting simulations. Thus, during the modelling of VaR in a Mean-Reverting market, one cannot sufficiently justify using the GARCH(1,1) model, only based on the results of conventional backtesting procedures.

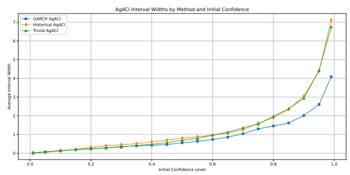
Table 1: Violation Rates and Backtesting p-values for Different Market Conditions. Significance levels: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Baseline Model	**GARCH Model**	**Historical**	**Trivial**
Violation rates (%)	5.25	5.50	5.75
Kupiec test p-value	0.6107	0.3123	0.1324
Christoffersen test p-value	0.8143	0.0220**	0.5790
Mean-Reverting Markets			
Violation rates (%)	5.30	5.30	5.80
Kupiec test p-value	0.5419	0.5419	0.1090
Christoffersen test p-value	0.7785	0.0754*	0.6130
Highly Volatile Markets			
Violation rates (%)	5.35	5.65	5.60
Kupiec test p-value	0.4774	0.1909	0.2267
Christoffersen test p-value	0.5858	0.0334**	0.2758
Extreme Volatility Clustering			
Violation rates (%)	5.40	5.95	6.25
Kupiec test p-value	0.4175	0.0581*	0.0134**
Christoffersen test p-value	0.6199	0.0132**	0.0134**

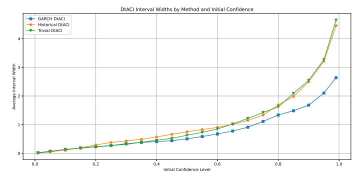
This issue may be exaggerated in real-world settings, in a case all models satisfy the backtests. CP provides a solution in the form of a guarantee of volatility coverage. As it can be seen from the graphs in Figure 1, GARCH(1,1) model

[3] Synthetic return series are generated using a GARCH(1,1)-style model with optional leverage and jump components, where volatility evolves dynamically based on past returns and past volatility. By adjusting the parameters—omega (baseline variance), alpha (shock sensitivity), beta (volatility persistence), and a leverage term (asymmetric impact of negative returns)—different market regimes are simulated, from smooth, mean-reverting behavior to extreme volatility clustering and crisis-like dynamics.

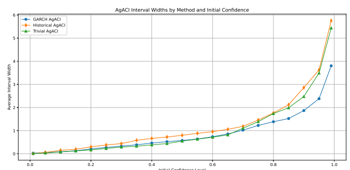
(green line), which is theoretically expected to outperform other models, has narrower intervals, suggesting higher certainty in its predictions, in all market regimes.



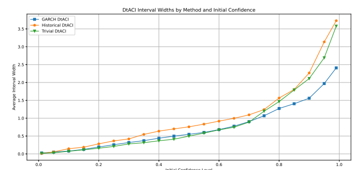
(a) Baseline



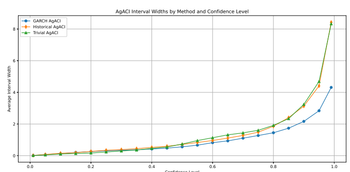
(e) Baseline



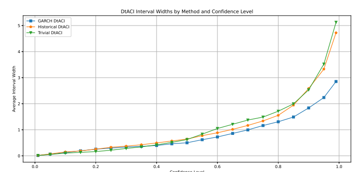
(b) Mean Reverting



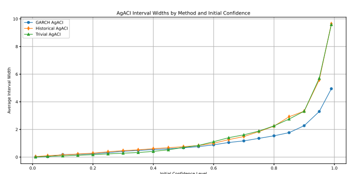
(f) Mean Reverting



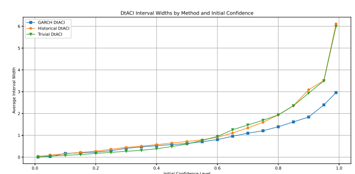
(c) High Volatility



(g) High Volatility



(d) High Clustering



(h) High Clustering

Fig. 1: AgACI (left) and DtACI (right) downside interval width in different market regimes.

4.2 Real-world Data

This section presents the main results of the research based on real-world data. It consists of daily returns on 24 fixed-income assets derived from Bloomberg.[4]

[4] The portfolio includes the following bonds: 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%

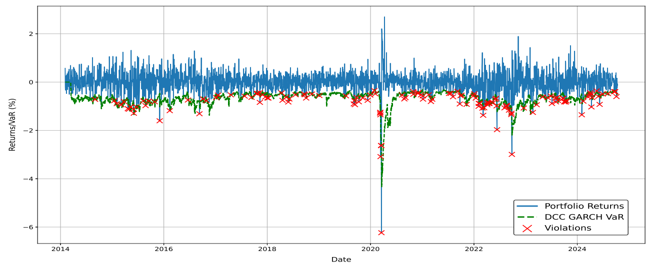
All bonds used for the portfolio are senior, plain vanilla bonds with longer maturities. The starting date of the dataset is January 2014, with the end date being October 2024. Thus, all bonds have maturities of more than 10 years. These are corporate/financial bonds issued in the United States, Great Britain, France, and Italy, in (currency) USD, GBP, EUR, and JPY, respectively.

In this paper, three main methods for estimating Value at Risk have been conducted, CAViaR method (first proposed by Engle and Manganelli [6]), Dynamic Conditional Correlation GARCH model, and Copula model. The main results of these models are presented in Table 2 below, while the process of estimation is described in Appendices I-III. It can be seen from the table that all models are somewhat accurate in predicting VaR (based on Kupiec test) and that the VaR violations are independent at a 5% significance level (based on Christoffersen test). The plots of the models with returns, estimated VaR values and violations for all three models are presented in Figure 2. This particular estimation period consists of different market regimes, moderate to high volatility period (up to 2017), relatively stable period (2017 to 2020), period of big shocks on the market (COVID-19 crisis during 2020), and period of high volatility on the financial markets (period after 2022, that could be linked to the war in Ukraine).

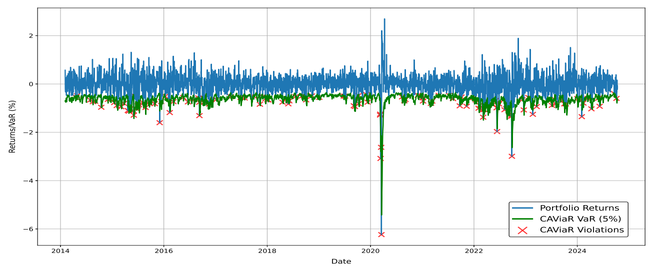
Table 2: Comparison of Risk Model Performance and 95% Conformal Interval Widths.

Metric	**CAViaR**	**GARCH**	**Copula**
Backtesting Results			
Violation Rate (at 5%)	4.98%	5.02%	4.85%
Kupiec t-statistic	0.0019	0.0019	0.0579
Kupiec p-value	0.9654	0.9654	0.8099
Christoffersen t-statistic	1.2497	0.6707	3.6838
Christoffersen p-value	0.2636	0.4128	0.0549
AgACI Interval Width (95% Confidence)			
Average Width	1.0266	1.2413	1.2682
Minimum Width	0.8319	0.8457	0.7424
Maximum Width	1.2258	1.5290	1.5361
Median Width	0.8613	1.2960	1.3490
DtACI Interval Width (95% Confidence)			
Average Width	0.8311	0.9169	1.1391
Minimum Width	0.4819	0.4268	0.5162
Maximum Width	2.3968	3.5396	6.4531
Median Width	0.7456	0.7782	0.9652
AgACI vs DtACI Width Ratio			
Avg. Width Ratio	1.2352	1.3538	1.1133
Median Width Ratio	1.1552	1.6654	1.3976

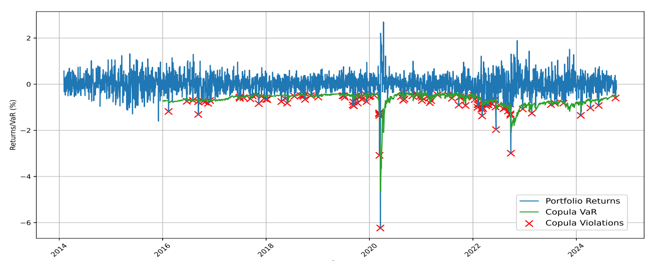
11/07/2043 Corp; C 5.15% 05/21/2026 REGS Corp; C 5.875% 01/30/2042 Corp; C 6.8% 06/25/2038 REGS Corp; C 7.375% 09/01/2039 REGS Corp; C 8.125% 07/15/2039 Corp; GS 6.25% 02/01/2041 Corp; ISPIM 0.00% 01/08/2027 Corp; JPM 2.875% 05/24/2028 REGS Corp; JPM 3.5% 12/18/2026 REGS Corp; JPM 5.4% 01/06/2042 Corp; 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.



(a) GARCH VaR estimates with Violations.



(b) CAViaR VaR estimates with Violations.

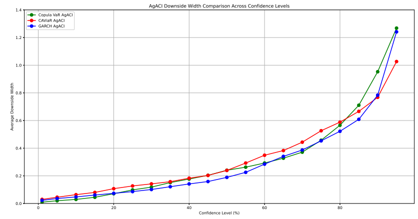


(c) Copulas VaR estimates with Violations.

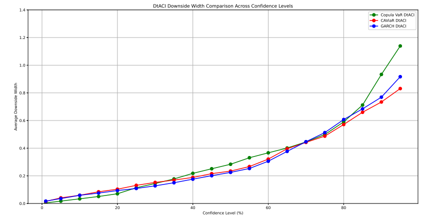
Fig. 2: Comparison of Models: GARCH, CAViaR, and Copulas.

Values for VaR, obtained using the three methodologies, are used to estimate conformal intervals with the AgACI and DtACI methods. They derived somewhat similar results, though, as it can be observed from the plots below (see Figure 3), the DtACI methodology has consistently narrower intervals at each confidence level. This is achieved while at the same time preserving the guaranteed coverage, which is suggested by the DtACI violation rates plot. Moreover, it can be seen from Figure 3 that the models performed very similarly at most confidence levels. Namely, the only significant difference seems to be at higher confidence levels for DtACI, where GARCH and CAViaR models outperform Copula VaR.

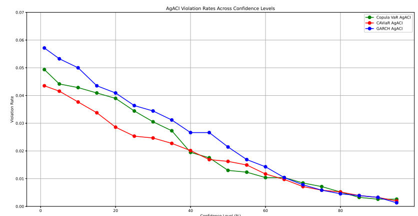
It might also be valuable to observe how the confidence intervals behave during the time, which is presented for DtACI methodology in Figure 4 below.



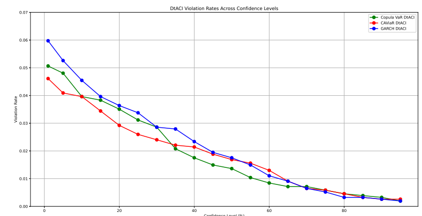
(a) AgACI downside width comparison



(b) DtACI downside width comparison



(c) AgACI violation rates



(d) DtACI violation rates

Fig. 3: Results for AgACI and DtACI.

Namely, at this figure, the 40% and 95% confidence intervals are presented. What can be observed is that all of these methods are deriving similar behaviour at the same confidence levels and at the same time. Namely, during the highly volatile times, at the beginning of 2020 (COVID-19 crisis) and after 2022 (the Ukranian war), there is an obvious and substantial increase in interval widths for all of these methods. This seems to be the definite result of increased volatility on the markets, and, as a result, potentially higher instability of models projections. However, what can be observed is that, surprisingly, the Copula method (which should excel during the time of severe economic downturn) and the GARCH model have higher interval widths at these periods.

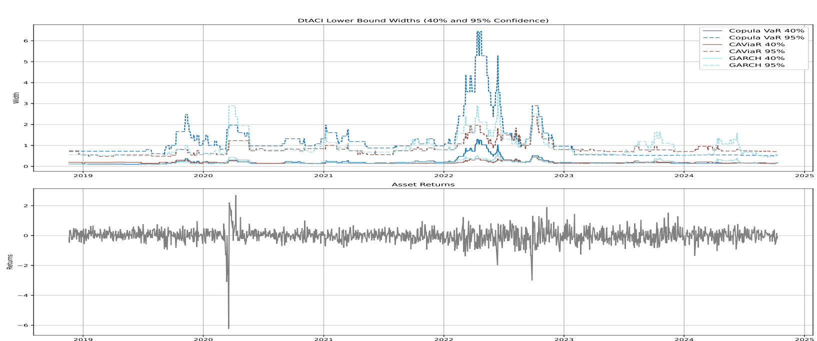


Fig. 4: DtACI lower bound widths at 40% and 95% confidence level.

5 Conclusions

This research utilized the Adaptive Conformal Inference (ACI), more specifically AgACI and DtACI methodologies for constructing confidence intervals around Value at Risk (VaR) estimates. Our proposed methods, AgACI and DtACI, successfully constructed valid intervals around VaR, while at the same time preserving finite-sample coverage guarantees. However, there were some notable differences between the two methods. DtACI seems to exhibit superior adaptability and consistently generates narrower intervals, not at the expense of coverage guarantee. This was particularly notable at higher confidence levels, with some models having as much as 35% narrower intervals on average.

In summary, while both AgACI and DtACI proved effective, these findings suggest that DtACI may be better suited for financial risk estimation, especially during periods of frequent distributional shifts. The model's inherited ability to dynamically adjust expert weights and optimize parameters enables it to be more responsive to market stress. Thus, combing VaR estimation models with DtACI presents a promising avenue for research and improving overall risk management risk metrics. This arises from ACI's ability to provide reliable confidence intervals, which in turn can enhance decision-making, especially under uncertain market conditions.

Furthermore, the results, to some extent, reinforce findings from prior literature, particularly related to the limitations of traditional backtesting methods for VaR evaluation. As noted by Christoffersen et al. [2], while statistical backtests such as Kupiec and Christoffersen tests can confirm the adequacy of coverage and independence under stable conditions, they may fall short in fully assessing the predictive quality during periods of significant regime market shifts. In this context, CP methodology may serve as a more flexible and data-driven mechanism that will extend the traditional risk evaluation framework.

These findings support the view that conformal prediction improves the practical relevance of risk metrics in finance. More specifically, the ability of conformal intervals to dynamically adapt to changing market conditions while not making strong distributional assumptions about the underlying returns offers a distinct advantage over classical confidence levels VaR estimation techniques, which often rely on strong assumptions.

Overall, it seems that the integration of conformal prediction into financial risk modeling aligns with the ongoing shift toward more transparent, distribution-agnostic, and adaptive tools in quantitative finance. This is especially relevant in light of recent market events that have challenged the assumptions of many traditional models for estimation of risk, further motivating the need for methodologies that can retain statistical validity while adapting to non-stationary dynamics. However, it is important to note that this study, due to the nature of the non-conformity score, provides only downside volatility guarantee. This could be considered both as a strength, from the risk management perspective, and a limitation, from the perspective of generalisability and broader use. Hence, future research into this area could investigate the use of alternative non-conformity scores, depending on the purpose of the research.

References

[1] Beutner, E., Heinemann, A., Smeekes, S.: A residual bootstrap for conditional value-at-risk. Journal of Econometrics **238**(2), 1–16 (2024). https://doi.org/10.1016/j.jeconom.2023.105554

[2] Christoffersen, P., Gonçalves, S.: Estimation risk in financial risk management. Tech. rep., CIRANO (2004)

[3] Contreras, P., Bhattacharjee, A.: A bayesian confidence interval for value-at-risk (2004), working paper

[4] Danielsson, J.: Financial Risk Forecasting: The Theory and Practice of Forecasting Market Risk with Implementation in R and Matlab, vol. 588. John Wiley & Sons (2011)

[5] Duan, J.C.: Maximum likelihood estimation using price data of the derivative contract. Mathematical Finance **4**(2), 155–167 (1994)

[6] Engle, R.F., Manganelli, S.: Caviar: Conditional autoregressive value at risk by regression quantiles. Journal of Business & Economic Statistics **22**(4), 367–381 (2004)

[7] Francq, C., Zakoian, J.M.: Maximum likelihood estimation of pure garch and arma-garch processes. Bernoulli **10**(4), 605–637 (2004)

[8] Gibbs, I., Candès, E.J.: Conformal inference for online prediction with arbitrary distribution shifts. Journal of Machine Learning Research **25**(162), 1–36 (2024)

[9] Gibbs, I., Candès, E.: Adaptive conformal inference under distribution shift. In: Advances in Neural Information Processing Systems (NeurIPS). vol. 34, pp. 1660–1672 (2021)

[10] Goh, Y.L., Pooi, A.H.: Confidence intervals for multivariate value at risk. ScienceAsia **39S**, 70–74 (2013)

[11] He, Y., Hou, Y., Peng, L., Shen, H.: Inference for conditional value-at-risk of a predictive regression. Annals of Statistics **48**(6), 3442–3464 (2020). https://doi.org/10.1214/19-AOS1937, https://doi.org/10.1214/19-AOS1937

[12] Huschens, S.: Confidence intervals for the value-at-risk. In: Risk Measurement, Econometrics and Neural Networks: Selected Articles of the 6th Econometric-Workshop in Karlsruhe, Germany, pp. 233–244. Physica-Verlag HD (1998)

[13] Koenker, R., Bassett, G.: Regression quantiles. Econometrica **46**(1), 33–50 (1978)

[14] Taylor, J.W.: Forecasting value at risk and expected shortfall using a model with a dynamic omega ratio. Journal of Banking & Finance **140**, 106519 (2022)

[15] Vovk, V., Gammerman, A., Shafer, G.: Algorithmic Learning in a Random World. Springer (2005)

[16] Wang, H.: 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 (2003)
[17] Xu, C., Xie, Y.: Conformal prediction interval for dynamic time-series. In: Proceedings of the International Conference on Machine Learning. pp. 11559–11569. PMLR (2021)
[18] Zaffran, M., Dieuleveut, A., Féron, O., Goude, Y., Josse, J.: Adaptive conformal predictions for time series. arXiv preprint arXiv:2202.07282 (2022)

Appendix I CAViaR Model

One of the models use in a process of VaR modeling is CAViaR (the Conditional Autoregressive Value at Risk) proposed by [6]. 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 [6] 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, and
- β_4 models the impact of negative returns.

If we have a sample of 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 $\beta_1, \beta_2, \beta_3, \beta_4$ are estimated by minimizing the quantile loss function, introduced by Koanker and Bassett [13] as:

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

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).
- $I(y_t < f_t(\beta))$ is an indicator function that takes the value 1 if y_t is less than VaR_t and 0 otherwise.

Appendix II 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:

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

where L is the Cholesky decomposition matrix of H_t.

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 III 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:

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