POLITECNICO MILANO 1863

SCUOLA DI INGEGNERIA INDUSTRIALE E DELL'INFORMAZIONE

Executive Summary of the Thesis

To Conformalise or not to Conformalise in Growth-at-Risk Analysis? Assessing the Empirical Performance of Calibrated Versus Non-Calibrated Quantile Prediction Models.

Laurea Magistrale in Mathematical Engineering - Ingegneria Matematica

Author: Pietro Bogani

Advisor: Prof. Simone Vantini

Co-advisors: Matteo Fontana, Luca Neri

Academic year: 2023-2024

1. Introduction

The study examines the critical task of computing robust extremal quantiles of empirical distributions, a task that is crucial for applications in policymaking and the financial industry, particularly for risk measures such as Growth-at-Risk (GaR) and Value-at-Risk (VaR). We conduct an extensive simulation study and a real-world analysis of GaR to assess the effectiveness of implementing Conformal Prediction for quantile estimation. This work represents the first application of Conformal Prediction, traditionally used for generating prediction intervals, to the estimation of quantiles.

2. Methodology

Our algorithm is based on an adaptation of the Conformalized Quantile Regression (CQR) [1], an algorithm which integrates conformal prediction methods with quantile regression to provide robust quantile estimates with valid coverage guarantees. Traditionally, conformal prediction is used to construct prediction intervals that contain the true response variable with a certain probability, regardless of the underlying data distribution. However, this study extends the application of conformal prediction to directly estimate quantiles, obtaining also a related coverage guarantee

The CQR algorithm begins by splitting the dataset into two disjoint subsets: a training set and a calibration set. The training set is used to fit two conditional quantile functions, $\hat{Q}(\alpha/2, x)$ and $\hat{Q}(1-\alpha/2, x)$, where α represents the desired miscoverage rate. These quantile functions can be computed using various methods, such as Quantile Regression (QR) or Quantile Random Forest (QRF). Next, nonconformity scores are computed on the calibration set to measure how unusual each observation is relative to the fitted quantile functions. Specifically, the nonconformity scores E_t are calculated as:

$$E_t = \max\left\{\hat{Q}(\alpha/2, X_t) - Y_t,\, Y_t - \hat{Q}(1-\alpha/2, X_t)\right\},$$

for each observation (X_t, Y_t) in the calibration set. The empirical quantile of these scores is then used to adjust the quantile functions and produce a prediction interval.

To adapt this algorithm for quantile estimation, we modified the CQR procedure. Instead of producing two quantile functions, we generate only one conditional quantile function, $\hat{Q}(\alpha, x)$, which corresponds to the quantile level of interest. The nonconformity scores are then computed as:

$$E_t = \hat{Q}(\alpha, X_t) - Y_t.$$

The final quantile estimate is obtained from the related prediction interval

$$[\hat{Q}(\alpha, x) - Q_E(1-\alpha), +\infty),$$

where $Q_E(1-\alpha)$ is the $(1-\alpha)(1+1/|T_2|)$-th empirical quantile of the nonconformity scores E_t. This modification allows us to estimate specific quantiles levels directly. Starting from the coverage guarantee in the original work of Romano et al.:

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

We derive a key inequality, related to our quantile estimation:

$$\mathbb{P}\{Y_{n+1} \leq \hat{Q}(\alpha, x) - Q_E(1-\alpha)\} \leq \alpha. \qquad (1)$$

Our result guarantees that the estimated quantile $\hat{Q}(\alpha, x) - Q_E(1-\alpha)$ will exceed the next observation Y_{n+1} at most α% of the time. This coverage guarantee is particularly beneficial in risk management scenarios, such as Growth-at-Risk analysis, where policymakers need to quantify the likelihood of adverse outcomes. For example, central banks can set a predefined threshold for GDP growth (e.g., 2%) and use this inequality to determine the maximum probability of an outcome worse than that threshold. Conversely, it can also be used to identify the GDP growth value associated with a specific risk level, α. This dual functionality provides a robust framework for managing economic risks by aligning growth projections with acceptable risk thresholds.
However, it's important to note that the coverage guarantee provided by the modified CQR algorithm relies on the assumption of exchangeability, which is often violated in real-world time series data due to inherent temporal dependencies. When the exchangeability condition is not met, the theoretical coverage guarantee may no longer hold. In this study, we empirically assess the behaviour of the coverage guarantee when the exchangeability assumption does not hold. Addressing this limitation remains an area of current and future research, as most conformal prediction methods rely on exchangeability assumptions.
By modifying the CQR algorithm for quantile estimation, this study offers a novel approach that extends the utility of conformal prediction methods beyond their traditional use, providing a powerful tool for both theoretical research and practical applications in economic and financial contexts.

3. Simulation Study

A key performance indicator in our thesis is the mean absolute error between the estimated coverages and the quantile level:

$$\text{MAE} := \frac{1}{2000} \sum_{i=1}^{20} \sum_{j=1}^{100} \left| \hat{P}(\hat{Q}_{ij}) - L_i \right|. \qquad (4)$$

The quantity is averaged across the 100 iterations and the 20 quantile levels L_i. $\hat{P}(\hat{Q}_i)$ is the unconditional empirical coverage, which is simply the proportion of test data that falls below the empirical quantile value $\hat{Q}_i$ estimated by our model. Perfectly calibrated models have $\text{MAE} = 0$ and a low value of MAE is associated to a well calibrated model.

3.1. AR(2) with Cauchy Distributed Errors

The first simulation considers an AR(2) model with Cauchy-distributed errors:

$$Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \epsilon_t$$

where

$$\phi_1 = 0.5, \quad \phi_2 = -0.2, \quad \epsilon_t \sim \text{Cauchy}(0, 1).$$

The use of a Cauchy error term significantly impacts the model's characteristics by introducing heavy tails and frequent outliers, which poses substantial challenges to standard quantile estimation methods. This scenario is particularly relevant for financial data, where accurately capturing the probability of extreme losses is crucial.
The results show that while the Quantile Regression (QR) model provides satisfactory calibration even without conformalisation, the Quantile Random Forest (QRF) model benefits significantly from the conformalisation step. The CQR QRF model, which combines the conformal approach with QRF, shows a marked improvement

in calibration, particularly at the extreme quantiles, where standard methods typically struggle. Table 1 presents the MAE for each model across different sample sizes n. The table illustrates that the CQR methods, especially CQR QRF, substantially reduce the MAE compared to their non-conformalised counterparts, thereby confirming the effectiveness of the conformalisation procedure.

MAE	98	198	998
QR	0.011	0.008	0.006
QRF	0.026	0.021	0.018
CQR QR	0.015	0.008	0.005
CQR QRF	0.017	0.009	0.006

Table 1: MAE (4) for QR, QRF, CQR QR, and CQR QRF models across different sample sizes n.

Overall, the results highlight that the conformalisation step is effective in improving calibration for QRF models while not compromising the good performance of the QR method, confirming the robustness of CQR methods in dealing with heavy-tailed distributions.

3.2. AR(2) with Exogenous Variables

In this simulation, we implement an autoregressive model with exogenous variables:

$$Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \boldsymbol{\beta}^\top \mathbf{X}_t + \epsilon_t$$

where

$$\phi_1 = 0.5, \quad \phi_2 = -0.2, \quad \epsilon_t \sim t_2,$$

$$\beta_i \sim U(0,1) \quad \text{for each} \quad i = 1, \dots, p,$$

$$p/n \in \{0.1, 0.2, 0.3, 0.4\}$$

$$\mathbf{X}_t \sim \mathcal{N}_p(\mathbf{m}, \mathbf{v})$$

$$\mathbf{m} \sim \mathcal{N}_p(\mathbf{0}, \mathbf{I}), \quad \mathbf{v} = \begin{bmatrix} \sigma_1 & & \\ & \ddots & \\ & & \sigma_p \end{bmatrix}, \quad \sigma_i \sim U(0,10)$$

This simulation introduces a t-Student error term with two degrees of freedom, which maintains heavy tails while having a finite mean. We also introduce exogenous variables to test the models' calibration when the number of covariates increases, with different ratios of covariates to observations $p/n : 0.1, 0.2, 0.3, 0.4$.

From Figure 1, we observe that the calibration of QR models worsens significantly as the ratio p/n increases, particularly for extreme quantiles, highlighting the limitations of QR in high-dimensional settings. On the other hand, CQR QR maintains good calibration across all quantile levels, showing its robustness in handling a large number of covariates. The MAE results in Table 2 further confirm that the conformalisation step substantially improves the performance of the CQR QR and CQR QRF models, providing lower MAE values compared to their non-conformal counterparts. CQR methods also show better calibration as the sample size n increases, whereas the improvement for standard QR and QRF is either less evident or negligible.

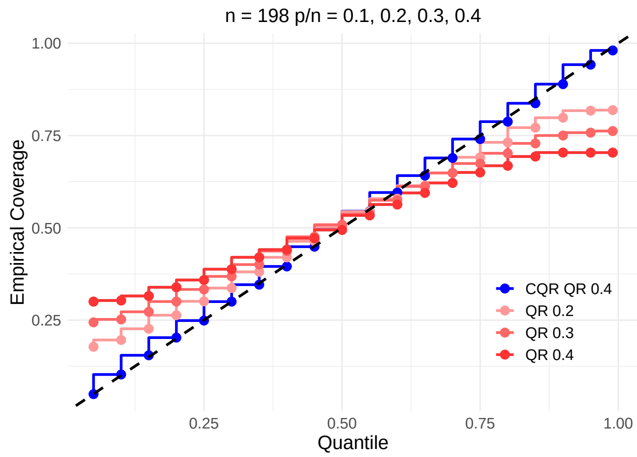


Figure 1: This plot provides an intuitive overview of the calibration of QR models as the ratio p/n increases, compared to a single CQR QR model with $p/n = 0.4$, which is very well-calibrated.

MAE	98	198	998
QR	0.08	0.077	0.073
QRF	0.019	0.02	0.015
CQR QR	0.012	0.007	0.004
CQR QRF	0.013	0.005	0.003

Table 2: MAE (4) for the four models across different sample sizes n, averaged over the four p/n ratios.

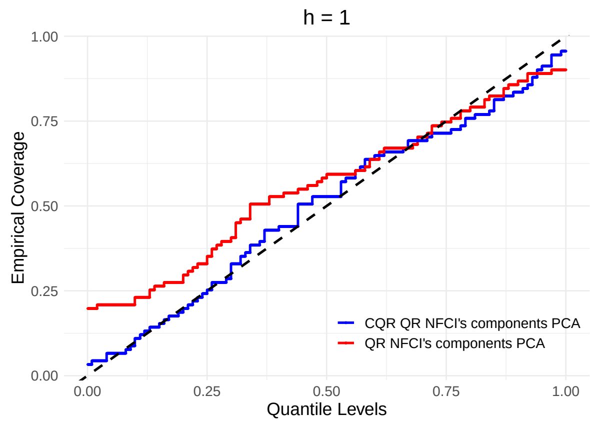
Overall, the results demonstrate that the conformalisation step significantly improves calibration for both QR and QRF models in high-dimensional settings. The CQR methods outperform their non-conformal counterparts, particularly for extreme quantiles, confirming their effectiveness and robustness in scenarios with many covariates and heavy-tailed distributions.

3.3. AR(1) with Nearly Unit Root Coefficients

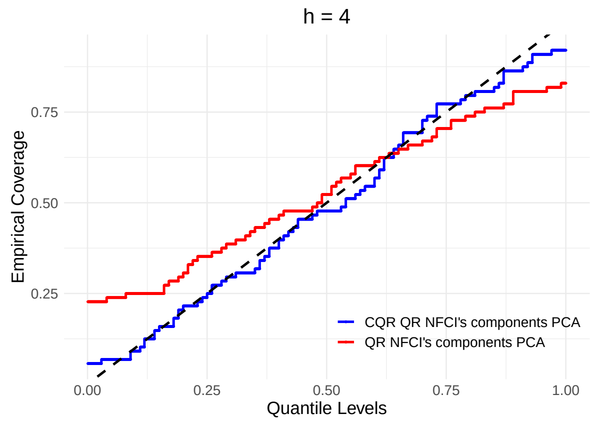
We also examine AR(1) processes with coefficients close to a unit root. The results indicate that none of the models, conformal or non-conformal, performed well in these settings. This highlights a limitation of the proposed method when dealing with nearly unit root time series, which exhibit persistence and autocorrelation.

4. Real Case Study

For the real-world analysis, we use the quarterly values of the National Financial Conditions Index (NFCI) [2] and U.S. GDP growth from 1973 to 2016, as provided by Adrian et al. [3]. Their study serves as a benchmark for our models' performance, focusing on predicting GDP growth using an autoregressive model that incorporates the NFCI, which captures both financial and macroeconomic conditions. Building on our simulation findings, where CQR methods demonstrated strong performance in handling a large number of covariates, we apply CQR algorithm using the individual components of the NFCI rather than the aggregated index. This approach is also designed to exploit the potential of the individual components in providing more granular and precise information about future GDP growth, potentially enhancing the calibration of quantile estimates.
Figure 2 shows the calibration curves for QR and CQR QR using the NFCI components, highlighting the superior performance of CQR QR in terms of calibration. The CQR QR model consistently provides better-calibrated estimates across different quantile levels, particularly at the tails of the distribution.



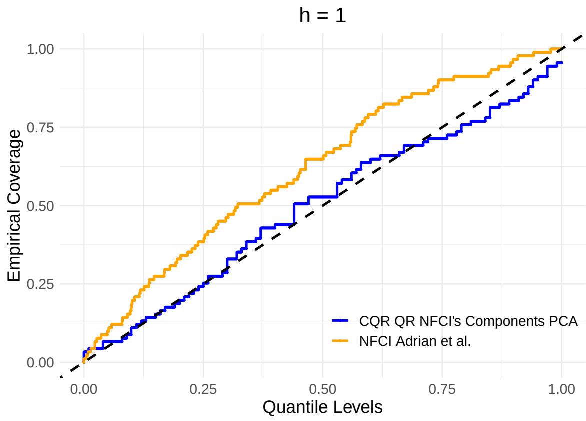
(a)



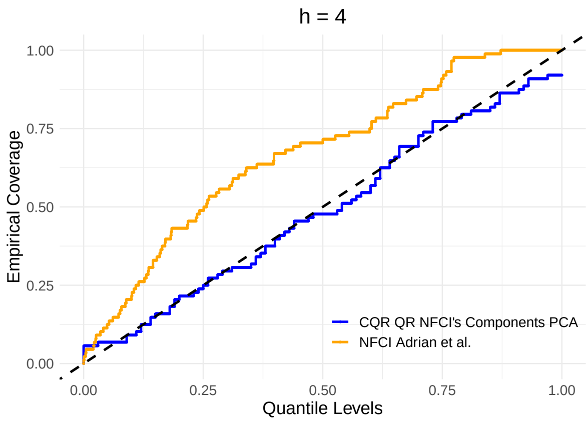
(b)

Figure 2: Calibration curves for QR and CQR QR models using the individual components of the NFCI as predictors for GDP growth. The values $h = 1$ and $h = 4$ correspond to one-quarter-ahead and one-year-ahead estimates of GDP growth, respectively.

To benchmark against the original work, Figure 3 compares our CQR QR results with those from Adrian et al. The comparison demonstrates that our conformalised method provides superior calibration.

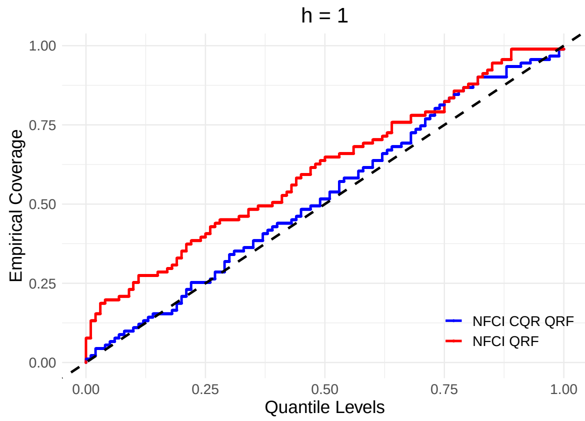


(a)

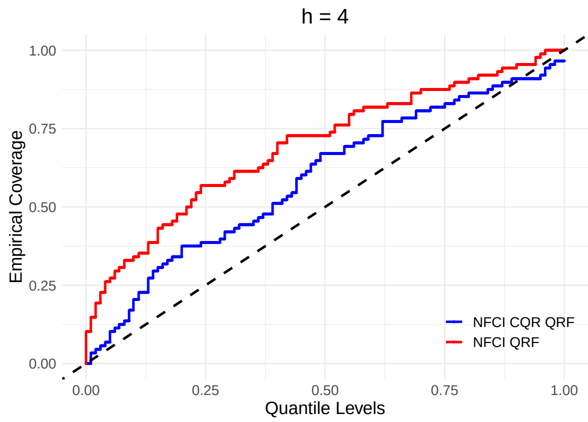


(b)

Figure 3: Comparison between the calibration curves from Adrian et al. and our CQR QR model using the NFCI components. The values $h = 1$ and $h = 4$ correspond to one-quarter-ahead and one-year-ahead estimates of GDP growth, respectively.



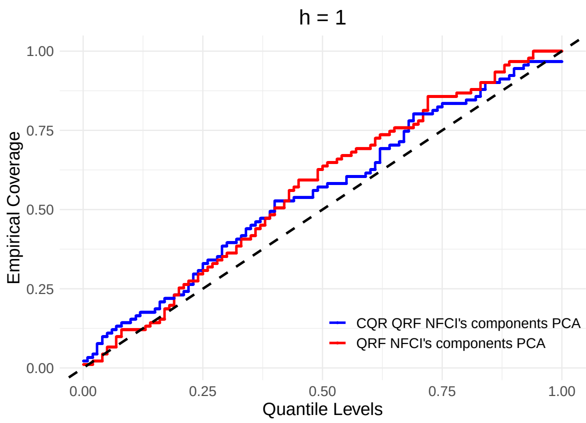
(a)



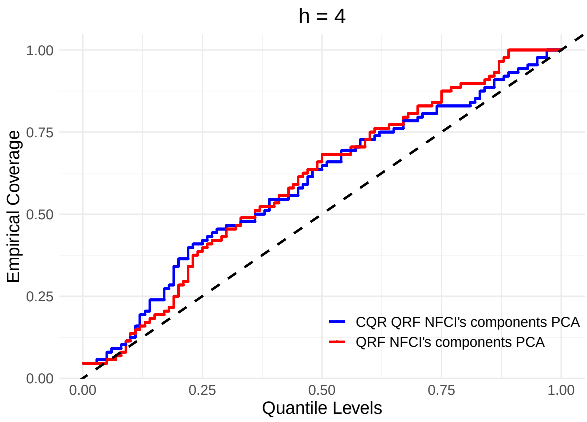
(b)

Figure 4: Calibration curves for QRF and CQR QRF models using the NFCI index as predictors for GDP growth. The values $h = 1$ and $h = 4$ correspond to one-quarter-ahead and one-year-ahead estimates of GDP growth, respectively.

For QRF models, Figure 4 below illustrates the comparison of QRF and CQR QRF models using the individual components of the NFCI for predicting U.S. GDP growth. Our findings indicate that the conformalisation procedure significantly improves calibration, particularly for h = 1, even without using NFCI components. Figure 5 illustrates the performance of QRF versus CQR QRF using the individual components of the NFCI. CQR QRF shows a modest improvement in calibration compared to its non-conformal counterpart, which aligns with our simulation findings where standard QRF demonstrated strong performance in handling a large number of covariates.



(a)



(b)

Figure 5: Calibration curves for QRF and CQR QRF models using the individual components of the NFCI as predictors for GDP growth. The values $h = 1$ and $h = 4$ correspond to one-quarter-ahead and one-year-ahead estimates of GDP growth, respectively.

Table 3 and 4 provide a comprehensive comparison of the MAE for all models across different time horizons. The conformalised methods consistently achieve lower MAE values, confirming their superior performance in real-world scenarios.

MAE	QR	CQR QR
NFCI h = 1	0.125	0.075
NFCI h = 4	0.177	0.115
PCA on NFCI components h = 1	0.075	0.028
PCA on NFCI components h = 4	0.079	0.023

Table 3: MAE (4) for QR and CQR QR models across different horizons and datasets.

MAE	QRF	CQR QRF
NFCI h = 1	0.104	0.017
NFCI h = 4	0.192	0.091
PCA on NFCI components h = 1	0.064	0.051
PCA on NFCI components h = 4	0.094	0.086

Table 4: MAE (4) for QRF and CQR QRF models across different horizons and datasets.

Overall, the real-case analysis confirms the advantages of applying conformalised methods to improve the calibration of quantile estimates in economic and financial contexts, offering a more accurate tool for risk assessment and policy formulation.

5. Conclusions

This study demonstrates the value of CQR in improving the calibration and robustness of quantile estimates across various economic and financial contexts, both in simulation and in a real case scenario. Our results reveal that CQR methods consistently enhance the accuracy of quantile predictions, especially at extreme levels, which are critical for risk assessment in Growth-at-Risk analysis.
A key finding of this research is the introduction of a coverage guarantee for quantile estimates under the exchangeability assumption. In our extensive simulation study, despite the ab-

sence of exchangeability, this coverage guarantee nearly holds. In the real-world case study, although the coverage property did not hold as robustly as in the simulations, the performance of CQR methods remained strong. The time-dependent nature of GDP growth rates and the violation of the exchangeability assumption may lead to some overestimation of coverage, but CQR still provided significant improvements in accuracy and calibration, demonstrating its value even in real-world scenarios. Overall, this thesis contributes to the literature by providing the first application of conformal prediction methods to quantile estimation, offering a novel perspective on how these methods can be adapted and utilized for economic and financial forecasting. Our findings underscore the potential of CQR in enhancing the robustness of risk assessments.

Future research could explore methods to overcome the limitations posed by the exchangeability assumption in time-series data, as well as adapt other conformal prediction techniques for quantile estimation. Additionally, the proposed method could be applied to other economic and financial contexts, such as Value-at-Risk estimation, and in any setting where accurate quantile estimation is crucial.

References

[1] Yaniv Romano, Evan Patterson, and Emmanuel J. Candès. Conformalized Quantile Regression. *Advances in Neural Information Processing Systems (NeurIPS)*, 2019.

[2] Federal Reserve Bank of Chicago. National Financial Conditions Index (NFCI), 2024.

[3] Tobias Adrian, Nina Boyarchenko, and Domenico Giannone. Vulnerable Growth. *American Economic Review*, 2019.

[4] International Monetary Fund. Global financial stability report: Is growth at risk? Technical report, International Monetary Fund, Washington, DC, 2017.

[5] Rina Foygel Barber, Emmanuel J. Candès, Aaditya Ramdas, and Ryan J. Tibshirani. Conformal Prediction Beyond Exchangeability. *The Annals of Statistics*, 2023.

[6] James H. Stock and Mark W. Watson. Testing for common trends. *Journal of the American Statistical Association*, 1988.