1. Introduction

Correct specification of a loss/returns distribution is key to the accuracy of a risk measure such as Value at Risk. As noted in [9], the major difference among many estimators of Value at Risk lies in estimation of the distribution of returns. Complexity in modeling financial data is due to its failure to exhibit standard statistical properties such as normality, independence, and identical distribution [18]. Statistical tests have revealed that returns exhibit fat-tails, time-varying volatility, and volatility clustering. Moreover, [7] showed that returns exhibit serial correlation over long time horizons. Models based on mean autoregression coupled with results from extreme value theory such as the AR(1)-GARCH(1,1) model in [30] incorporated most of the aforementioned characteristics of financial data but suffer from lack of robustness due to the effect of extreme observations on the mean. Extreme quantile autoregression in [2, 24, 25] among others leads to a more robust model. This is because they combine regression quantiles introduced by [17] in an autoregressive fashion while using extreme value techniques on the resulting residuals to capture the tail behaviour. A major challenge of this approach is possible quantile crossing.

The challenge of quantile crossing has been addressed by smoothing suggestions in [5, 6, 16] among others in a nonparametric setting. Equally, the conditional location-scale model used in obtaining Restricted Regression Quantiles (RRQ) in [11] averts possible crossing in extreme quantiles but can suffer from the same when estimating the median. To avert quantile crossing even at the middle, [1] added a forced ordering constraint in the estimation of multiple quantiles. Simulation results revealed that, based on standard error of the estimates, noncrossing quantile regression in [1] produces better estimates in the middle. However, the RRQ estimator produced better estimates at the tails, especially when the sample size was large.

The lack of monotonicity in estimation of conditional quantiles is addressed in [3] through sorting of originally estimated nonmonotone quantile curves using a functional delta method. The monotonic quantile functions obtained were found to be closer to the true quantile than the nonmonotonic quantiles. Function limit theory for the rearranged estimators was also derived. The resulting
monotonic quantile functions were then used in estimating economic functions using Vietnam data. However, the model was not extended to extreme cases to cover heavy tails beyond the sample.

Parametric quantile regression is used in [27] to estimate percentiles in positive valued datasets. Specifically, a linear quantile regression model was used with the error term assumed to follow a generalized gamma distribution. The idea of quantile regression was achieved by allowing parameters for the error distribution to depend on the univariate covariate leading to a location-scale model. The four-, five-, and six-parameter generalized conditional gamma distributions were considered and likelihood ratio test was used in selecting the best-fit model for each dataset. Asymptotics for the three resulting models were also derived. However, the use of generalized gamma distribution limits the model to applications where the covariate is greater than zero. This together with the fact that some financial datasets have heavier tails than the Gamma distribution limits the application of the model in finance.

We seek to improve the extreme conditional quantile estimator in [25] using an interquantile dispersion from the central conditional quantile.

2. Methods and Estimation

Let \( \{S_t, t \in \mathbb{R}^+\} \) be a real valued financial time series on a complete probability space \((\Omega, \mathcal{F}, P)\). We assume that \( S_t \in \mathbb{R}^+ \) and it is \( \mathcal{F}_t \)-measurable where \( \mathcal{F}_t, t \in \mathbb{R}^+ \) is an increasing sequence of \( \sigma \)-algebras representing information available up to time \( t \). In particular, let \( S_t \) be the value of a portfolio at trading time \( t \). The return on the portfolio at time \( t \), used to quantify the gain in value of the portfolio from trading time \( t - \Delta t \) to trading time \( t \), is given by

\[
R_t = \frac{S_t - S_{t-\Delta t}}{S_{t-\Delta t}}.
\]

so that

\[
X_t = -R_t,
\]

is the corresponding loss return of \( S_t \).

**Definition 1 (Risk Measure).** A risk measure is a function \( \rho \) from a set \( \mathcal{L} \) of risks in a financial position (in this case, the loss distribution) to \( \mathbb{R} \); that is, \( \rho : \mathcal{L} \rightarrow \mathbb{R} \).

We assume that \( X_t \) can be expressed using a linear heteroscedastic model of the following form:

\[
X_t = \mu_t + \epsilon_t,
\]

where \( \mu_t \equiv f: \mathbb{R}^d \rightarrow \mathbb{R} \) is the conditional mean function of \( X_t \) given \( \mathcal{F}_{t-1} \) and it is defined as \( \mu_t = Y_t \beta \). \( \epsilon_t \) are errors and \( Y_t \) is a \( d \)-dimensional process which is \( \mathcal{F}_{t-1} \)-measurable. In particular, \( Y_t \) has 1 as the first element and a collection of the last observed returns up to time \( t - 1 \); that is, \( Y_t = (1, X_{t-1}, \ldots, X_{t-d}) \). To ensure that the model is smooth and obeys some of the financial norms such as clustering of shocks, we further assume that \( \epsilon_t \) can be decomposed into

\[
e_t = \epsilon_t \sigma_t,
\]

where \( \epsilon_t \) are independent and identically distributed random variables and \( \sigma_t > 0 \) is the conditional volatility. In this case, \( X_t \) is said to assume a location-scale model of the form

\[
X_t = \mu_t + \sigma_t \epsilon_t.
\]

The corresponding \( \alpha \) quantile of \( X_t \) under this formulation is given by

\[
\mu_{t, \alpha}^m = \mu_t + \sigma_t \alpha Z_{t, \alpha},
\]

where \( Z_{t, \alpha} \) is the \( \alpha \)-quantile of \( \epsilon_t \). Let us now define a conditional quantile autoregressive model on \( X_t \) of the form

\[
X_t = \mu_{t, \theta} + \epsilon_t,
\]

where \( \mu_{t, \theta} \) is the central conditional \( \theta \)-quantile of \( X_t \) and \( \epsilon_t \) are errors with zero \( \theta \)-quantile. Let \( \epsilon_t = \sigma_{t, \theta} Z_{t, \theta} \), where \( \sigma_{t, \theta} \) is the central conditional scale of \( X_t \) and \( Z_{t, \theta} \) are i.i.d. residuals.

Using an approach similar to Points Over Threshold (POT), we propose an extreme conditional quantile of the form given in equation (8) and refer to it as the adjusted extreme conditional quantile. That is, suppose that we are interested in extreme quantiles, \( \mu_{t, \alpha, \theta} \), for some \( \alpha \approx 1 \) or \( 0 \); the idea is to estimate the central quantile, \( \mu_{t, \theta} \), and scale \( \sigma_{t, \theta} \), for some level \( \theta \) in the middle and approximate the extreme quantile as

\[
\mu_{t, \alpha, \theta} = \mu_{t, \theta} + \sigma_{t, \theta} (\bar{q}_{t, \alpha} - \bar{q}_{t, \theta}),
\]

where \( \bar{q}_{t, \alpha} \) and \( \bar{q}_{t, \theta} \) are the \( \alpha, \theta \)-quantile of \( Z_t \), respectively, for \( \alpha, \theta \in (0,1) \). If the parametric distribution of \( Z_t \) is known, then \( \bar{q}_{t, \alpha} \) and \( \bar{q}_{t, \theta} \) are easily determined as the inverse of the cumulative distribution of \( Z_t \) at probability levels \( \alpha \) and \( \theta \), respectively; otherwise, appropriate estimates are determined. Note that \( \mu_{t, \alpha, \theta} \) is \( \mu_{t, \alpha} \) in conditional quantile autoregressive form. Observe that when \( \alpha = \theta \), equation (8) reduces to

\[
\mu_{t, \theta, \theta} = \mu_{t, \theta},
\]

which is the central conditional quantile of \( X_t \) given \( Y_t = y \). This confirms that indeed \( \epsilon_t \) has a zero \( \theta \)-quantile. From equation (8), we obtain the following estimator for the extreme conditional quantile:

\[
\hat{\mu}_{t, \alpha, \theta} = \hat{\mu}_{t, \theta} + \hat{\sigma}_{t, \theta} (\bar{q}_{t, \alpha}^{	heta} - \bar{q}_{t, \theta}^{	heta}),
\]

where \( \bar{q}_{t, \alpha}^{	heta} \) and \( \bar{q}_{t, \theta}^{	heta} \) are appropriate estimates of the \( \theta, \alpha \)-quantiles of the i.i.d. residuals, respectively. We compare this estimator with

\[
\hat{\mu}_{t, \theta} = \hat{\mu}_{t, \theta} + \hat{\sigma}_{t, \theta} Z_{t, \theta},
\]

proposed in [25], and

\[
\hat{\mu}_{t, \alpha} = \hat{\mu}_{t, \alpha} + \hat{\sigma}_{t, \alpha} c_{\alpha},
\]

where \( c_{\alpha} \) is the resulting coefficient from quantile regression of \( \epsilon_t \) against \( \hat{\epsilon}_t \) at 100\(\alpha\)% confidence level. Note that \( \hat{\mu}_{t, \alpha, \theta} \) is the estimator proposed in [11].
2.1. Estimation of Central Quantiles. Let \( \mu_t : \mathbb{R}^d \to \mathbb{R} \) be an unknown smooth function and define the loss function \( M_\theta \):

\[
M_\theta(x, \mu_t) = \theta |x - \mu_t|^+ + (1 - \theta) |x - \mu_t|^{-},
\]

where \( |x - \mu_t|^+ \) and \( |x - \mu_t|^{-} \) represent absolute positive and negative values, respectively, and \( 1_{x - \mu_t \leq 0} \) is the indicator function. Assuming that the conditional quantile process is well defined, we expect

\[
E[M_\theta(X_t, \mu_t)|Y_t = y] = 0.
\]

So the \( \theta \)-conditional quantile of \( X_t \) is given by

\[
\mu_{t, \theta} = \arg \min_{\mu_t} E[M_\theta(X_t, \mu_t)|Y_t = y].
\]

Note that equation (14) can be used to check whether the conditional quantile process is correctly specified or not. We impose the following regularity assumptions to ensure consistency of the conditional quantiles.

Assumption 1. \( \{X_t, Y_t\} \) are identically distributed with the joint probability density \( f_{XY}(x, y) \) and a continuous conditional probability density \( f_{X|Y}(x|y) \) of \( X_t \) given \( Y_t = y \).

Assumption 2. There exists \( \delta > 1 \) such that \( E(\|Y_t\|^{\delta}) < \infty \).

Assumption 3. \( \mu_{t, \theta} \equiv f : B \to \mathbb{R} \), where \( B \) is a compact subset of \( \mathbb{R}^d \).

Assumption 4. \( f_{X|Y}(y|x) > 0 \forall x \), where \( f_{X|Y}(y|x) \) is the conditional probability density of \( \varepsilon_t = X_t - \mu_{t, \theta} \) given \( Y_t = y \).

From the sample analog of equation (15), we obtain

\[
\bar{\mu}_{t, \theta} = \arg \min_{\mu_t} \frac{1}{n} \sum_{i=1}^{n} M_\theta[(X_t, \mu_t)|Y_t = y],
\]

which is the \( \theta \)-conditional quantile estimator for a sample of size \( n \). To overcome the limitation of quantile crossing, we used the approach in [1] where required quantiles were estimated simultaneously with a noncrossing constraint using the optimization problem:

\[
\min_{\mu_t} \sum_{i=1}^{k} w(\theta_i) \sum_{i=1}^{n} M_{\theta_i}[(X_t, \mu_t)|Y_t = y]
\]

subject to \( \mu_{t, i} \geq \mu_{t, i+1}, Y_t \in B, i = 1, \ldots, k, \)

for some weight function \( w(\theta_i) > 0 \). A conveniently used practical choice of the weight function which was also adopted for this study is \( w(\theta_i) = 1, \forall i = 1, \ldots, k. \)

Lemma 1 (consistency central quantiles’ estimator). Given that Assumptions 1, 2, 3, and 4 hold, \( \bar{\mu}_{t, \theta} - \mu_{t, \theta} = o_p(1). \)

2.1.1. The Scale Function. To still maintain dependence and ensure positivity of the scale function, this study incorporated a scale function in the form of a quantile autoregressive (QAR) function on the absolute of the nonstandardized residuals. This was achieved by replacing \( \mu_{t, \theta} \) with its corresponding estimate in equation (7) so that \( \varepsilon_t = X_t - \bar{\mu}_{t, \theta} \) and

\[
|\varepsilon_t| = \sigma_{t, \theta} + \bar{\theta}_{t, \theta} \eta_t,
\]

where \( \sigma_{t, \theta} \) is the central conditional \( \theta \)-quantile of \( |\varepsilon_t|, \bar{\theta}_{t, \theta} \) is the conditional scale of \( |\varepsilon_t|, \eta_t \) are i.i.d. residuals, and \( \bar{\theta}_t = \bar{\theta}_{t, \theta} \eta_t \) are errors with zero \( \theta \)-quantile. Similarly, as is the case in Section 2.1, we let \( \sigma_t : \mathbb{R}^d \to \mathbb{R} \) be an unknown smooth function and define the loss function \( M_\theta(x, \sigma_t) \). We assume that the QAR process in equation (18) obeys the four regularity assumptions given earlier so that, using noncrossing quantile regression approach, we obtain the following estimate of the scale:

\[
\bar{\sigma}_{t, \theta} = \arg \min_{\sigma_t} \frac{1}{n} \sum_{i=1}^{n} M_{\theta}[(|\varepsilon_t|, \sigma_t)|Y_t = y],
\]

2.2. Extreme Value Theory (EVT). Most of the financial datasets are heavy-tailed [12]. Therefore, it is fundamentally important to incorporate extreme value theory in the estimation of extreme quantiles. A basic requirement for the application of EVT is independence in the particular distribution. The study in [25] observed that it is appropriate to assume that, at high (low) levels of \( \alpha \), the standardized residuals in equation (7) given by

\[
Z_t = \frac{\varepsilon_t}{\bar{\sigma}_{t, \theta}}
\]

where \( \bar{\sigma}_{t, \theta} \) is the estimate of the scale, are approximately independent, which allows us to apply EVT. Let \( Z_1, Z_2, \ldots, Z_n \) follow a common distribution function \( F \). Consider a sample \( (Z_i)_{i=1}^{n} \) from which \( M_n = \max[Z_1, \ldots, Z_n] \) and is such that \( M_n \to z_F \) (a.s) where \( z_F = \sup\{z \in \mathbb{R}|F(z) < 1\} \). Pursuant to Fisher–Tippett’s theorem in [8], the random variable \( Z \) (or alternatively the distribution \( F \) of \( Z \)) is said to belong to the Maximum Domain of Attraction (MDA) of the extreme value distribution \( H \) if there exist normalizing constants \( c_n > 0 \) and \( d_n \in \mathbb{R} \) such that

\[
\frac{M_n - d_n}{c_n} \to H.
\]

Consequently, \( H \) is referred to as the Generalized Extreme Value (GEV) distribution.

This study applied Points Over Threshold (POT) method because it uses more data leading to better estimates compared to the Block Maxima method. POT models the distribution
\(F_u(z) = P(Z - u \leq z | Z > u) = \frac{F(z + u) - F(u)}{1 - F(u)}, \quad (22)\)
of all excesses above a particular threshold \(u\), where \(0 \leq z \leq z_F - u\).

**Theorem 1** (Pickands–Balkema–de Haan) (see [22]). We can find (positive-measurable function) \(\beta(u)\) such that
\[
\lim_{u \to z_F} \sup_{u \leq z < z_F - u} |F_u(z) - G_{\lambda, \beta}(z)| = 0, \quad (23)
\]
if and only if \(F \in \text{MDA}(H_1), \lambda \in \mathbb{R}_+,\) and \(G_{\lambda, \beta}(z)\) is the Generalized Pareto Distribution (GPD) given by
\[
G_{\lambda, \beta}(z) = \begin{cases} 
1 - \left(1 + \frac{(z - u)}{\beta}\right)^{-\frac{1}{\lambda}}, & \lambda \neq 0, \\
1 - \exp\left(-\frac{(z - u)}{\beta}\right), & \lambda = 0,
\end{cases} \quad (24)
\]
with \(z \geq u\) when \(\lambda > 0\) and \(0 \leq z \leq u - (\beta/\lambda)\) when \(\lambda < 0\). \(\lambda\) and \(\beta\) are the shape and scale parameters, respectively.

A major challenge in POT is accuracy in choosing a threshold to separate extreme observations from the center of the distribution [30]. Among the methods discussed in [32], most authors such as those of [21, 28, 30] prefer the conventional method in which a threshold is chosen such that between 5% and 10% of the sample data is classified as extreme observations is chosen. Although conventional method is subjective, the choice of the threshold can be checked for appropriateness using a mean excess plot. The mean excess function for the GPD is given by
\[
E[Z - u | Z > u] = \frac{\beta - \lambda u}{1 + \lambda}, \quad (25)
\]
where \(1 + \lambda > 0\). This implies that an optimal threshold corresponds to start of approximate linearity of the mean excess plot with the sign of the slope, \(\lambda/1 - \lambda\), indicating the specific family of the GPD. A positive sign corresponds to the Frechet family, while a negative sign implies the Weibull family [8].

For convenience in making inferences on variability of the estimated quantiles, a recommendation in [13] on the use of Probability Weighted Moments (PWM) method in estimating parameters of the GPD was adopted. Using the first and second PWMs, we obtain the corresponding parameters estimates as
\[
\lambda = \frac{M_0}{M_0 - 2M_1} - 2, \\
\beta = \frac{2M_0M_1}{M_0 - 2M_1}, \quad (26)
\]
where \(M_0\) and \(M_1\) are obtained by replacing for \(k = 0\) and \(k = 1\) in

\[
M_{1,0,k} = M_k = E[|X - F(X)|^k] = \frac{\beta}{(k + 1)(k + 1 + \lambda)}, \quad (27)
\]
which is the PWM of GPD with \(\lambda > -1\). See [10, 23] for details on PWM method. For a sample of size \(n\), the corresponding PWM estimates are given by
\[
\tilde{M}_k = \frac{1}{n} \sum_{i=1}^{n} (1 - P_{x_i, n})^k x_{i, n}, \quad (28)
\]
for \(z > u\) and \(G_{\lambda, \beta}(z - u)\) the Generalized Pareto Distribution. When \(F(u)\) is approximated empirically, we obtain the following estimate of \(F(z)\):
\[
\tilde{F}(z) = 1 - \frac{m}{N} \left(1 + \frac{\bar{\lambda}(z - u)}{\bar{\beta}}\right)^{(-\frac{1}{\bar{\lambda}})}, \quad (30)
\]
where \(N\) is the sample of size, \(m\) is the number of exceedances above the threshold \(u\), and \(\bar{\beta}\) together with \(\bar{\lambda}\) are the estimated GPD parameters.

**Lemma 3** (consistency of Probability Density Function (PDF) estimator). Let \(Z_1, Z_2, \ldots, Z_n\) be i.i.d. random variables from a Cumulative Distribution Function (CDF) \(F(z)\) belonging to the MDA of \(H(\lambda)\). Suppose that \(F(z)\) has a right endpoint at \(z_F \leq \infty\); then \(\sup_z \tilde{F}(z) - F(z) \to 0\) as \(n \to \infty\).

From equation (30), we get the following estimate for the quantile of the standardized residuals at level \(\alpha\):
\[
\tilde{q}_n^\alpha = u + \frac{\bar{\beta}}{\bar{\lambda}} \left[\left(\frac{N}{m} (1 - \alpha)\right)^{\frac{\bar{\lambda}}{\bar{\beta}}} - 1\right]. \quad (31)
\]

**Lemma 4** (consistency of error quantiles). Let \(Z_1, Z_2, \ldots, Z_n\) be i.i.d. random variables from a CDF \(F(z)\) belonging to the MDA of \(H(\lambda)\) satisfying \(\alpha < F(q_n^\alpha + \varepsilon)\) for any \(\varepsilon > 0\). Then, for every \(\varepsilon > 0\) and \(n = 1, 2, \ldots\),
\[
\lim_{n \to \infty} P(\tilde{q}_n^\alpha - q_n^\alpha < \varepsilon) = 1. \quad (32)
\]

Using equation (10), we obtained the one-step VaR predictions as
\[
\tilde{\nu}_n^{\alpha + 1} = \tilde{\nu}_{n+1, \beta, \alpha} = \tilde{\nu}_{n+1, \beta} + \tilde{\vartheta}_{n+1, \beta}(\tilde{q}_n^\alpha - \tilde{q}_n^\alpha), \quad (33)
\]
where \( \hat{\mu}_{t+1, \theta} \) and \( \hat{\sigma}_{t+1, \theta} \) are the corresponding one-step \( \theta \)-quantile and scale estimates, respectively, from the linear conditional quantile process.

**Theorem 2** (consistency of extreme quantile estimator). Given that the QAR processes defined satisfy Assumptions 1, 2, 3, and 4 and \( F(z) \) belongs to the MDA of \( H(\lambda) \), \( \hat{\mu}_{t, \alpha} - \mu_{t, \alpha} = o_p(1) \).

### 3. Simulations

To evaluate the accuracy of our estimators, a sample of size \( T = 4250 \) was generated using the model

\[ X_t = 0.5 + 0.3X_{t-1} + \sqrt{1 + 0.35X_{t-1}^2}Z_t, \]

where \( Z_t \) follows Student’s \( t \)-distribution with 4 degrees of freedom. The sample was partitioned into design data of size \( n \) and test data of size \( T - n \). Figure 1 represents a sample path of the model superimposed with the median. Note that, for simulation purposes, \( \theta = 0.5 \) was used in estimating the central quantile.

Clearly, from the sample path, there is some level of volatility clustering, which is common in most financial data. An ACF plot of the resulting standardized residuals in Figure 2 confirms that indeed they are independent.

\[ z = 2.3283 \] was chosen as the threshold to ensure that 10% of the resulting ordered standardized errors were classified as extremes. This was confirmed by approximate linearity of the mean excess plot after the threshold as shown in Figure 3.

The corresponding shape and scale parameter estimates from the GPD fit were 0.1156182 and 1.272937, respectively. Figure 4 shows the corresponding quantile estimates at different levels of \( \alpha \).

The accuracy of our extreme quantile estimator was evaluated using the Average Root Mean Squared Error (ARMSE). The RMSE seeks to return the Mean Squared Error (MSE) to the original scale of the sample. For \( k \) sample paths of simulations of size \( n \) of the extreme quantiles, the average RMSE is given by

\[
\text{A.R.M.S.E.} = \frac{1}{m} \sum_{i=1}^{k} \frac{1}{n} \sum_{t=1}^{n} (\hat{\mu}_{t, \alpha} - \mu_{t, \alpha})^2, \tag{34}
\]

where \( \hat{\mu}_{t, \alpha} \) is replaced by \( \hat{\mu}_{t, \alpha, m} \) or \( \hat{\mu}_{t, \alpha, h} \) when considering extreme conditional quantiles or restricted regression quantiles, respectively. To check how our model behaves under different choices of central quantile, we computed ARMSE for the extreme conditional quantile at \( \alpha = 0.95 \), where \( \theta = 0.25, 0.5 \) and 0.75 were considered. 1000 sample paths, each of sizes 250, 500, 1000, 2000, and 4000, were used in the computation of ARMSE. Table 2 reports the obtained ARMSE values.

We note that, for a large enough sample (2000 observations and above), ARMSE is lowest when \( \theta = 0.5 \) and thus this choice of theta is maintained in investigating the accuracy of our estimator and forecasting the one-day-ahead VaR. Table 3 outlines the obtained ARMSE for the extreme conditional quantile at \( \alpha = 0.95 \) under three different models. The sample sizes and number or replications are still maintained.

Based on ARMSE, both RRQ and ECQ perform better than AECQ for small samples. However, as the sample size increases, AECQ outperforms both RRQ and ECQ. The decreasing ARMSE with increase in the sample size for AECQ and ECQ confirms that both are consistent estimators of the extreme conditional quantile. Also, for sample size above 2000, the rate of convergence of the AECQ estimator is higher than that of the ECQ estimator. It was not possible to comment on the consistency of the RRQ estimator, since its ARMSE fluctuated with increase in sample size. The consistent reduction in ARMSE when noncrossing constraint is added during estimation, confirming that indeed this constraint increases accuracy of resulting estimators.

### 4. Evaluating VaR Forecasts

In Section 3, we evaluated accuracy of our in-sample quantile estimates. In this section, we extend this by evaluating the out-of-sample VaR forecasts from our quantile estimator. To achieve this, we carry out backtests on 250 one-day-ahead VaR forecasts (as recommended in the Basel Accord) using coverage tests in [4, 31]. Coverage tests were adopted due to their popularity in literature and practice [26]. Consider the failure process \( I_t^n = I[X_t > \text{VaR}^n_t] \), where \( I \) is the indicator function such that

\[
I_t^n = \begin{cases} 1, & \text{if } X_t > \text{VaR}^n_t, \\ 0, & \text{otherwise} \end{cases}, \tag{35}
\]

and \( t = n + 1, n + 2, \ldots, T \). By Lemma 1 in [4], \( I_t^n \sim \text{Binom}(a) \) which is tested using the conditional coverage test that combines both the unconditional coverage test in [19] and the test for independence in [4] to under the null hypothesis \( E[I_t^n|I_{t-1}^n] = \alpha \). The likelihood under the null hypothesis is

\[
L(\alpha; I_t^n, I_{t+1}^n, \ldots, I_T^n) = (1-\alpha)^{n_0}\alpha^{n_1}, \tag{36}
\]

where \( n_1 \) is the number of VaR exceedances and \( n_0 = T-n-n_1 \). Now consider the first-order Markov chain generated by the transition probabilities of \( I_t^n \):

\[
\Pi_1 = \begin{bmatrix} \pi_{00} & \pi_{01} \\ \pi_{10} & \pi_{11} \end{bmatrix}, \tag{37}
\]

where \( \pi_{ij} = \Pr(I_t^n = j|I_{t-1}^n = i) \). This has an approximate likelihood function:

\[
L(\Pi_1; I_{t+1}^n, I_{t+2}^n, \ldots, I_T^n) = \pi_{00}^{n_0}\pi_{01}^{n_1}\pi_{10}^{n_0}\pi_{11}^{n_1}, \tag{38}
\]

where \( n_{ij} \) is the number of times observation \( i \) is followed by \( j \) in the failure process \( I_t^n \). From equation (38), we obtain the maximum likelihood estimate of \( \pi_{ij} \) as

\[
\hat{\pi}_{ij} = \frac{n_{ij}}{\sum_{j=0}^{1} n_{j}}, \tag{39}
\]
Figure 1: Sample path superimposed with median.

Figure 2: Autocorrelation plot of resulting residuals.

Figure 3: Mean excess plot.

Table 1: Sample statistics.

|   | $\alpha$ | 0 | 0.50 | 0.75 | 0.95 | 0.99 |
|---|----------|---|------|------|------|------|
| Minimum |          | -15.81 | -3.58 | -2.80 | -0.73 | 1.76 |
| Median ($Q_1, Q_3$) |          | 0.67 (0.01, 1.38) | 0.65 (0.48, 0.83) | 1.15 (0.99, 1.32) | 2.47 (2.34, 2.61) | 4.07 (3.97, 4.17) |
| Mean ± SD |          | 0.69 ± 1.41 | 0.66 ± 0.36 | 1.16 ± 0.34 | 2.48 ± 0.27 | 4.07 ± 0.20 |
| Maximum |          | 10.69 | 3.22 | 3.55 | 4.42 | 5.47 |

$Q_1$, lower quartile; $Q_3$, upper quartile; SD, standard deviation.
Figure 4: Quantile estimates at various $\alpha$-levels for $\theta = 50\%$.

Table 2: ARMSE for $\tilde{\mu}_{t,\theta,\alpha}$ for $\alpha = 95\%$.

| Sample size ($n$) | $\theta = 0.25$ | $\theta = 0.5$ | $\theta = 0.75$ |
|-------------------|-----------------|-----------------|-----------------|
| 250               | 0.96844         | 0.74493         | 0.62015         |
| 500               | 0.79632         | 0.60788         | 0.56179         |
| 1000              | 0.71974         | 0.54056         | 0.53254         |
| 2000              | 0.71416         | 0.49585         | 0.50037         |
| 3000              | 0.69751         | 0.47862         | 0.49713         |
| 4000              | 0.69608         | 0.46439         | 0.48776         |

Table 3: ARMSE for $\tilde{\mu}_{t,\theta,\alpha}$ under different models for $\theta = 50\%$ and $\alpha = 95\%$.

| Sample size ($n$) | Without noncrossing constraint | With noncrossing restriction |
|-------------------|--------------------------------|-----------------------------|
|                   | RRQ   | ECQ   | AECQ  | RRQ   | ECQ   | AECQ  |
| 250               | 0.58236 | 0.64552 | 0.73307 | 0.58200 | 0.64230 | 0.72728 |
| 500               | 0.56989 | 0.61336 | 0.60766 | 0.56816 | 0.61083 | 0.60341 |
| 1000              | 0.53386 | 0.58654 | 0.53386 | 0.55780 | 0.58529 | 0.53186 |
| 2000              | 0.56051 | 0.58204 | 0.49562 | 0.55800 | 0.57880 | 0.49360 |
| 3000              | 0.55478 | 0.57663 | 0.48235 | 0.55235 | 0.57369 | 0.48073 |
| 4000              | 0.55271 | 0.57619 | 0.47032 | 0.55013 | 0.57013 | 0.46832 |

RRQ, restricted regression quantiles in [11]. ECQ, extreme conditional quantiles in [25]. AECQ, proposed adjusted extreme conditional quantiles.

Table 4: $P$ values for the likelihood ratio tests.

| Sample size ($n$) | RRQ model | ECQ model | AECQ model |
|-------------------|------------|-----------|-------------|
|                   | $LR_{uc}$  | $LR_{ind}$| $LR_{cc}$  |
|                   | $LR_{uc}$  | $LR_{ind}$| $LR_{cc}$  |
|                   | $LR_{uc}$  | $LR_{ind}$| $LR_{cc}$  |
| 250               | 0.000      | 0.001     | 0.000       |
| 500               | 0.000      | 0.001     | 0.000       |
| 1000              | 0.000      | 0.000     | 0.000       |
| 2000              | 0.000      | 0.001     | 0.000       |
| 3000              | 0.000      | 0.000     | 0.000       |
| 4000              | 0.000      | 0.000     | 0.000       |

RRQ, restricted regression quantiles in [11]. ECQ, extreme conditional quantiles in [25]. AECQ, proposed adjusted extreme conditional quantiles.
Therefore, the conditional coverage hypothesis can be assessed using the likelihood ratio:

$$\text{LR}_{cc} = -2 \log \left( \frac{L(\theta; I_1^n, I_{n+1}^n, \ldots, I_T^n)}{L(\Pi_1; I_1^n, I_{n+1}^n, \ldots, I_T^n)} \right)^{\frac{1}{2y^2}} \chi^2(2). \quad (40)$$

Table 4 reports the obtained P values for the likelihood ratios of the three tests in [4]. The tests were conducted on 250 one-day-ahead 5% VaR forecasts from the three considered models.

$\text{LR}_{uc}$ denotes likelihood ratio for the unconditional coverage test. $\text{LR}_{ind}$ denotes likelihood ratio for the independence test. $\text{LR}_{cc}$ denotes likelihood ratio for the conditional coverage test. Models accepted at 5% level of significance are highlighted in bold. Note that $n$ is size of the sample used in estimation, while testing was done using a sample of size 250 for all $n$.

Observe that, as a consequence of consistency, the accuracy of ECQ and AECQ forecasts improves with increase in sample size. It can also be seen that all the three models perform poorly under LRind and LRcc due to dependence in sample size. It can also be seen that all the three models perform poorly under LRind and LRcc due to dependence in sample size. Based on $\text{LR}_{uc}$, the RRQ model performed significantly better than the other two models.

5. Conclusions and Recommendations

We have derived the extreme conditional quantile estimator and used it to obtain the one-step-ahead conditional Value at Risk forecast for a simulated financial distribution. Consistency of our estimators has been proved and illustrated through Monte Carlo simulations. We noticed that adding the noncrossing restriction during estimation improves accuracy of the resulting extreme conditional quantile estimator. Backtesting results from the one-step-ahead conditional Value at Risk forecasts indicate that independence and conditional coverage tests in [4] are not appropriate for our estimators due to dependence in autoregressive models.

6. Proofs

Proof of Lemma 1. Let $\zeta(\mu_{t,\beta}) = M_\beta(X_t, \mu_{t}) - M_\beta(X_t, \mu_{t,\beta})$, $\zeta_n(\mu_{t,\beta}) = (1/n) \sum_{i=1}^{n} \zeta(\mu_{i,\beta})$, and $\bar{\zeta}(\mu_{t,\beta}) = E(\zeta_n(\mu_{t,\beta}))$. Note that, by Assumption 1, $\zeta_n(\mu_{t,\beta})$ does not depend on $n$. Since $M_\beta(X_t, \mu_{t})$ does not depend on $\mu_{t,\beta}$, then $\mu_{t,\beta} = \arg\min_{\beta \in \mathbb{R}} (1/n) \sum_{i=1}^{n} \zeta(\mu_{i,\beta})$. We need to show that the objective function $\zeta(\mu_{t,\beta})$ satisfies the following conditions for application of Theorem 1.2 in [34]:

1. $\zeta(\mu_{t,\beta})$ is measurable for each $\beta \in \mathbb{B}$
2. $\zeta(\mu_{t,\beta})$ is continuous on $B$ almost surely
3. $\exists$ a measurable function $m: \mathbb{R}^{d} \rightarrow \mathbb{R}$ such that
   (i) $|\zeta(\mu_{t,\beta})| \leq m(Y_j) |\beta| \in B$
   (ii) $\exists \delta > 0: E|m(Y_j)|^{1+\delta} \leq M \leq \infty$

   (4) $\bar{\zeta}(\mu_{t,\beta})$ has a unique minimum at $\mu_{t,\beta}$

The functional form of $\zeta(\mu_{t,\beta})$ and Assumption 3 guarantees measurability of $M_\beta(X_t, \mu_{t,\beta})$. To prove condition (2), we first show that $\zeta(\mu_{t,\beta})$ is Lipschitz continuous. By definition,

$$\bar{\zeta}(\mu_{t,\beta}) = \zeta(\mu_{t,\beta}) + \zeta(X_t - \mu_{t,\beta}) = \zeta(\mu_{t,\beta}) - \zeta(X_t - \mu_{t,\beta}) \leq 0$$

Considering the possible range $X_t$, we have the following:

(i) For $\mu_t < X_t < \mu_{t,\beta}$, it follows that $I(X_t - \mu_t \leq 0) = 0$ and $I(X_t - \mu_{t,\beta} \leq 0) = 1$; hence, equation (41) reduces to

$$\bar{\zeta}(\mu_{t,\beta}) = \zeta(\mu_{t,\beta}) + \zeta(X_t - \mu_{t,\beta})$$

(ii) Similarly, when $\mu_t < \mu_{t,\beta} < X_t$, $I(X_t - \mu_t \leq 0) = 0$ and $I(X_t - \mu_{t,\beta} \leq 0) = 1$; hence,

$$\bar{\zeta}(\mu_{t,\beta}) = \zeta(\mu_{t,\beta}) - \zeta(X_t - \mu_t)$$

(iii) $X_t < \mu_t < \mu_{t,\beta}$, then $I(X_t - \mu_t \leq 0) = 1$ and $I(X_t - \mu_{t,\beta} \leq 0) = 1$; hence,

$$\bar{\zeta}(\mu_{t,\beta}) = \zeta(\mu_{t,\beta})$$

Combining equations (43)–(45), we have

$$\bar{\zeta}(\mu_{t,\beta}) \leq \max(\theta, 1 - \theta) |\mu_{t,\beta} - \mu_t| \leq |\mu_t - \mu_{t,\beta}| \leq |\mu_{t,\beta} - \mu_t|.$$

Thus, $\zeta(\mu_{t,\beta})$ is Lipschitz continuous for $K = 1$ and hence is differentiable almost everywhere by Rademacher’s theorem in [14] which implies that $\zeta(\mu_{t,\beta})$ is continuous everywhere.

To prove Condition 3(i), let $m(Y_j) = \|Y_j\| (L\|\beta\|^{\delta}) (\theta + 1)$, where $\|\beta\| = \max_{\|\beta\|} \|\beta\|$. Existence of $\theta$ is guaranteed by Assumption 3. Clearly, $m(Y_j)$ is measurable and $|\zeta(\mu_{t,\beta})| \leq m(Y_j)$. Assumption 2 ensures that Condition 3(ii) is satisfied.
To verify Condition 4, let $\delta = \mu - \mu_G$; we need to show that $E(\zeta(\mu_G)) > 0$ for any $\delta > 0$. By Knight’s identity in [15], we have

$$E[\zeta(\mu_G)] = E\left[M_G(X_1, \mu_G) - M_G(X_1, \mu_G)\right]$$

$$= E\left(\mu_G - \mu_G\right)\left(1 \left\{ X_1 - \mu_G \leq 0 \right\} - \theta\right) + \int_0^{\mu_G - \mu_G} \left[\left\{ X_1 - \mu_G \leq t\right\} - I\left\{ X_1 - \mu_G \leq 0\right\}\right] dt. \tag{47}$$

Thus, $E[\zeta(\mu_G)] > 0$ for all $\delta > 0$ by monotonicity of the CDF. Therefore, $\bar{\mu}_G - \mu_G = o_P(1)$ by Theorem 12.2 in [34].

Proof of Lemma 2. The proof proceeds in a similar way to the proof of Lemma 1.

Proof of Lemma 3. Observe that

$$\sup_z |\bar{F}(z) - F(z)| = \sup_z \left|1 - \bar{F}(u)G_{\lambda, \beta}(z - u) + \bar{F}(u) - 1 - F(u)G_{\lambda, \beta}(z - u) - F(u)\right|$$

$$= \sup_z \left|\bar{F}(u) - F(u)\right| + G_{\lambda, \beta}(z - u) \left(1 - \bar{F}(u)G_{\lambda, \beta}(z - u) + F(u)G_{\lambda, \beta}(z - u)\right)$$

$$+ \bar{F}(u)G_{\lambda, \beta}(z - u) - F(u)G_{\lambda, \beta}(z - u)$$

$$\leq \sup_u |\bar{F}(u) - F(u)| + \sup_z \left|\bar{G}_{\lambda, \beta}(z - u) - G_{\lambda, \beta}(z - u)\right|$$

$$- G_{\lambda, \beta}(z - u) \sup_u |\bar{F}(u) - F(u)| - F(u) \sup_u \left|\bar{G}_{\lambda, \beta}(z - u) - G_{\lambda, \beta}(z - u)\right|. \tag{48}$$

Data Availability

The data used in the article were simulated, and the data generating process (DGP) is included within the article.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

The authors acknowledge the Pan-African University of Basic Sciences, Technology and Innovation (PAUSTI) for funding this research.

References

[1] H. D. Bondell, B. J. Reich, and H. Wang, “Noncrossing quantile regression curve estimation,” Biometrika, vol. 97, no. 4, pp. 825–838, 2010.
[2] V. Chernozhukov, “Extremal quantile regression,” The Annals of Statistics, vol. 33, no. 2, pp. 806–839, 2005.
[3] V. Chernozhukov, I. Fernández-Val, and A. Galichon, “Quantile and probability curves without crossing,” Econometrica, vol. 78, no. 3, pp. 1093–1125, 2010.
