# CAViaR Model Selection Via Adaptive Lasso*

Zongwu Cai[a], Ying Fang[b], Dingshi Tian[c,†]

[a]Department of Economics, University of Kansas, Lawrence, KS 66045, USA.
[b]Wang Yanan Institute for Studies in Economics, Department of Statistics and Data Science, School of Economics, Xiamen University, Xiamen, Fujian 361005, China.
[c]Department of Statistics, School of Statistics and Mathematics, Zhongnan University of Economics and Law, Wuhan, Hubei 430073, China.

January 27, 2024

**Abstract:** The estimation and model selection of conditional autoregressive value at risk (CAViaR) model may be computationally intensive and even impractical when the true order of the quantile autoregressive components or the dimension of the other regressors are high. On the other hand, automatic variable selection methods cannot be directly applied to this problem because the quantile lag components are latent. In this paper, we propose to identify the optimal CAViaR model using a two-step approach. The estimation procedure consists of an approximation of the conditional quantile in the first step, followed by an adaptive Lasso penalized quantile regression of the regressors as well as the estimated quantile lag components in the second step. We show that under some mild regularity conditions, the proposed adaptive Lasso penalized quantile estimators enjoy the oracle properties. Finally, the proposed method is illustrated by Monte Carlo simulation study and applied to analyzing the daily data of the S&P500 return series.

**Keywords:** CAViaR model; Adaptive Lasso; Model selection; Tail risk.

**JEL classification:** C32, C51, C58

## 1 Introduction

To find an effective risk measure which is both responsive to financial or political news and easy to grasp even in the complex situations, Engle and Manganelli (2004) proposed the conditional autoregressive value at risk (CAViaR) model. Since then, the CAViaR model has soon become popular for its simplicity and the capability to specify the evolution of the quantile over time and it has been applied to various fields. As for

*The authors gratefully acknowledge the financial supports, in part, from the Natural Science Foundation under grant numbers *72103207, 72133002 and 72033008*, "the Fundamental Research Funds for the Central Universities", Zhongnan University of Economics and Law with grant *#2722023BY018*, and the Humanities and Social Science Research Foundation of Ministry of Education of China(*#22YJAZH038*).

financial risk management, for example, Kuester et al. (2006) shown that an extension to a particular CAViaR model outperforms the other alternative strategies in the value at risk (VaR) prediction. Recently, Laporta et al. (2018) investigated different VaR forecasts for daily energy commodities returns, and found that the CAViaR model and dynamic quantile regression model perform relatively better than other approaches such as GARCH, EGARCH, and GJR-GARCH, while Gao et al. (2022) proposed a regime-switching CAViaR model to jointly forecast the VaR and expected shortfall (ES) of Bitcoin series. Especially, from theoretical perspectives, the CAViaR model was extended to multivariate case by White et al. (2015). Besides, it was incorporated into the unobserved components model by Harvey (2013) and the generalized autoregressive score model by Creal et al. (2013), respectively. To overcome the problem of "elicitability" for ES, Taylor (2019) and Patton et al. (2019) considered using the loss function proposed by Fissler and Ziegel (2016) to jointly estimate the CAViaR model along with the dynamic ES model.

Although CAViaR model has attracted a great deal of research attention, the inclusion of autoregressive components may bring new challenge to the estimation of the model with appropriate orders. To alleviate the heavy computation burden in the optimization routines, Taylor (2008) proposed using the expectile regression to estimate the CAViaR model by the one-to-one relationship between quantile and expectile. Meanwhile, an iterative Kalman filter method introduced by De Rossi and Harvey (2009) can also be applied to calculate the CAViaR model. By establishing the relationship between the linear GARCH model of Taylor (1986) and the CAViaR model, Xiao and Koenker (2009) proposed a robust and easy-to-implement two-step approach for quantile regression on GARCH models.

However, none of the aforementioned papers addresses the issue about the model selection of CAViaR model. When the true orders of quantile lags or the dimension of the other regressors are high, the implementing of the CAViaR model can be computationally intensive. On the other hand, the traditional variable selection method such as adaptive Lasso in Zou (2006) for mean regression and in Wu and Liu (2009) for quantile regression cannot be directly applied because of the latent variable. In this paper, we propose a two-step procedure to fill this gap. Our estimation procedure consists of an approximation of the conditional quantiles in the first step, followed by an adaptive Lasso penalized quantile regression of the regressors as well as the estimated quantile lag components.

The rest of this paper is organized as follows: Section 2 introduces the CAViaR model and our two-step estimation and model selection procedures. Section 3 investigates the asymptotic properties of the two-step CAViaR estimators, and the Oracle properties of the adaptive Lasso penalized quantile regression estimators are also studied. Monte Carlo experiment and empirical analysis results of a real data example are reported in Sections 4 and 5, respectively. Section 6 concludes the paper. All technical proofs are deferred to the supplementary file.

## 2 Model Framework

Assume that $(Y_t, \mathbf{X}_t, \mathbf{M}_t), t = 1, 2, \ldots, T$, is a sequence of strictly stationary random vectors, where $Y_t$ is a scalar variable of interest, $\mathbf{X}_t \in \mathcal{R}^p$ is a vector of covariates defining the risk factors, and $\mathbf{M}_t \in \mathcal{R}^d$ denotes the tail risk drivers, which may contain lagged macroeconomic state variables, lagged firm-specific characteristics and lagged returns. Let $\mathcal{F}_t = \sigma\{(Y_1, \mathbf{X}_1, \mathbf{M}_1), \ldots, (Y_{t-1}, \mathbf{X}_{t-1}, \mathbf{M}_{t-1})\}$ be the information set available at time $t$, $\mathcal{M}_\tau$ denote a model of the $\tau$th conditional quantile of $Y_t$, $\mathcal{M}_\tau \equiv \{q_\tau(\mathbf{W}_t, \boldsymbol{\theta}_\tau^*)\}$, in which $\mathbf{W}_t \equiv h(\mathbf{X}_t, \ldots, \mathbf{X}_{t-n_0}) : \Omega \to \mathcal{R}^l$ with $h(\cdot)$ denoting a measurable random vector for some finite number $n_0$, $\boldsymbol{\theta}_\tau^*$ be an unknown parameter in the parameter space $\Theta$, and $q_\tau(\mathbf{W}_t, \cdot) : \Theta \to \mathcal{R}$ denote some real function. The $\tau$th conditional quantile of $Y_t$ given $\mathbf{W}_t = \mathbf{w}_t$ is defined as

$$q_\tau(\mathbf{w}_t, \boldsymbol{\theta}_\tau^*) = \arg\min_{u \in \mathcal{R}} E\{\mathrm{Q}_\tau(Y_t - u) | \mathbf{W}_t = \mathbf{w}_t\},$$

where $\mathrm{Q}_\tau(Y_t - u) = (\tau - \mathrm{I}(Y_t \leq u))(Y_t - u)$ with $\mathrm{I}(\cdot)$ denoting the indicator function. To specify the evolution of the quantile over time using an autoregressive process, Engle and Manganelli (2004) considered a class of CAViaR$(p, q)$ models as

$$q_\tau(\mathbf{w}_t, \boldsymbol{\theta}_\tau^*) \equiv q_{t,\tau} \;=\; \alpha_{0,\tau} + \sum_{i=1}^{p} \alpha_{i,\tau} X_{t,i} + \sum_{j=1}^{q} \beta_{j,\tau} q_{t-j,\tau} = \alpha_{0,\tau} + \boldsymbol{\alpha}_\tau^\top \mathbf{X}_t + B(L) q_{t,\tau}, \quad (1)$$

where $\boldsymbol{\theta}_\tau^* = (\alpha_{0,\tau}, \alpha_{1,\tau}, \ldots, \alpha_{p,\tau}, \beta_{1,\tau}, \ldots, \beta_{q,\tau})^\top \in \mathcal{R}^{p+q+1}$ with $p > 0, q > 0$ and $\beta_{j,\tau} \geq 0$ for $j = 1, \ldots, q$, $\boldsymbol{\alpha}_\tau = (\alpha_{1,\tau}, \ldots, \alpha_{p,\tau})^\top$, $q_{t-j,\tau}$ is the conditional quantile of $\{Y_t\}_{t=m}^T$ at time $t - j$ with a finite number $m \equiv \max(p, q) + 1$, and $B(L) = \sum_{j=1}^{q} \beta_{j,\tau} L^j$ with $L$ denoting the lag operator. Here, $\mathbf{X}_t$ is allowed to include the past returns of $Y_t$. Note that for simplicity of notation, $\tau$ is dropped from $\boldsymbol{\theta}_\tau^*$, $\alpha_{i,\tau}$, $i = 0, \ldots, p$ and $\beta_{j,\tau}$, $j = 1, \ldots, q$, if doing so does not cause confusion.

The key issue of implementing this method is how to select an appropriate and parsimonious model, which is similar to choose an appropriate orders $p$ and $q$ in an ARMA$(p, q)$ model. Here, we propose using the adaptive Lasso for the model selection purpose. The two-step estimation procedures are introduced below. In the first step, we model the conditional quantile of $Y_t$ using an approximately parametric quantile regression model of the tail risk drivers $\mathbf{M}_t$ as follows:

$$q_{t,\tau} \approx \mathbf{a}_\tau^\top \overline{\mathbf{M}}_t,$$

where $\mathbf{a}_\tau = (a_{0,\tau}, a_{1,\tau}, \ldots, a_{d,\tau})^\top$ and $\overline{\mathbf{M}}_t$ is an known function of $\mathbf{M}_t$. To capture possible nonlinearity, we recommend using B-spline (for such a case, $\overline{\mathbf{M}}_t$ is the B-spline functions of the tail risk drivers $\mathbf{M}_t$) or other type of parametric approximation approaches in the above approximation. For simplicity of exposition, $\overline{\mathbf{M}}_t$ is taken to be $\overline{\mathbf{M}}_t = (1, \mathbf{M}_t^\top)^\top$. Then, the conditional quantiles can be approximated using the following equation:

$$\widehat{q}_{t,\tau} \approx \widehat{a}_{0,\tau} + \sum_{j=1}^{d} \widehat{a}_{j,\tau} M_{t,j}.$$

In the second step, given $\mathbf{Z}_{t,\tau} = (1, X_{t,1}, \ldots, X_{t,p}, q_{t-1,\tau}, \ldots, q_{t-q,\tau})^\top$, then the optimal subset CAViaR model can be selected by the adaptive Lasso penalized quantile regression of $Y_t$ on $\widehat{\mathbf{Z}}_t = (1, X_{t,1}, \ldots, X_{t,p}, \widehat{q}_{t-1,\tau}, \ldots, \widehat{q}_{t-q,\tau})^\top$ and the adaptive Lasso penalized quantile regression estimator of $\boldsymbol{\theta}^*$ can be estimated by:

$$\widehat{\boldsymbol{\theta}}^{(T)} = \min_{\boldsymbol{\theta} \in \mathcal{R}^{p+q+1}} \frac{1}{T} \sum_{t=m}^{T} \mathrm{Q}_\tau \left( Y_t - \alpha_0 - \sum_{i=1}^{p} \alpha_i X_{t,i} - \sum_{j=1}^{q} \beta_j \widehat{q}_{t-j,\tau} \right) + \lambda_T \sum_{i=1}^{p+q+1} \widehat{w}_i |\theta_i|, \tag{2}$$

where $\lambda_T \in (0, \infty)$ is the regularization parameter, $\widehat{\mathbf{w}} \equiv (\widehat{w}_1, \ldots, \widehat{w}_{p+q+1}) = |\widetilde{\boldsymbol{\theta}}|^{-\eta}$ consists of $p+q+1$ data-driven weights with some appropriately chosen $\eta > 0$, and $\widetilde{\boldsymbol{\theta}}$ denotes the quantile regression estimator of $\boldsymbol{\theta}^*$. Specifically, the estimation procedures in the second step are formulated in the following algorithm.

**Algorithm**:

1. Calculate the initial value of adaptive weights $\frac{1}{|\widetilde{\theta}_s|}(s = 1, \ldots, p+q+1)$ by least square regression of $Y_t$ on $\widehat{\mathbf{Z}}_t$.

2. Update the adaptive weights $\frac{1}{|\widehat{\theta}_s|}(s = 1, \ldots, p+q+1)$ by adaptive Lasso regression of $Y_t$ on $\widehat{\mathbf{Z}}_t$.

3. Find the solution path of the adaptive Lasso regression.

4. The optimal $\lambda_T$ is selected using the cross validation technique.

5. Keep iterating 2-4 until convergence achieved.

# 3 Asymptotic Theory

In this section, the oracle property of the adaptive Lasso penalized quantile regression estimator is derived.

## 3.1 Notations and Assumptions

Now, the assumptions for deriving asymptotic results are listed below. Note that these assumptions given in this paper are sufficient conditions but not necessarily the weakest.

**Assumptions**:

A1. For the true system (1), the polynomial $B(z) \neq 0$ for $|z| \leq 1$.

A2. Conditional on the random vector $\mathbf{W}_t = \mathbf{w}_t$, the error term $\varepsilon_{t,\tau} \equiv Y_t - q_{t,\tau}$ form a stationary process, with continuous conditional density $h_t(\varepsilon|\mathbf{w}_t) \geq h$ for some $h > 0$ and for all $t$. Further, for some constant $N$ and for any $t$, $h_t(\varepsilon|\mathbf{w}_t) \leq N < \infty$.

A3. $h_t(\varepsilon|\mathbf{w}_t)$ satisfies Lipschitz condition, i.e., $|h_t(\lambda_1|\mathbf{w}_t) - h_t(\lambda_2|\mathbf{w}_t)| \leq L|\lambda_1 - \lambda_2|$ for some constant $L < \infty$ and for any $t$;

A4. $\|\mathbf{M}_t\| \leq A(\mathbf{w}_t)$ for all $t$, where $A(\mathbf{w}_t)$ is some stochastic function of variables that belong to the information set $\mathbf{w}_t$, such that $E(|A^2(\mathbf{w}_t)|) < \infty$; $\|\mathbf{X}_t\| \leq C(\mathbf{w}_t)$ where $C(\mathbf{w}_t)$ is some stochastic function of variables that belong to the information set $\mathbf{w}_t$, such that $E(|C^2(\mathbf{w}_t)|) < \infty$ and $E(|A(\mathbf{w}_t)C(\mathbf{w}_t)|) < \infty$; $q_{t,\tau} \leq H(\mathbf{w}_t)$ where $H(\mathbf{w}_t)$ is some stochastic function of variables that belong to the information set $\mathbf{w}_t$, such that $E(|H^2(\mathbf{w}_t)|) < \infty$ and $E(|C(\mathbf{w}_t)H(\mathbf{w}_t)|) < \infty$;

A5. Let $\mathbf{B}_T = E\left[T^{-1}\sum_{t=m}^{T} h_t(0|\mathbf{w}_t)\overline{\mathbf{M}}_t\overline{\mathbf{M}}_t^\top\right]$, $\mathbf{D}_T = E\left[T^{-1}\sum_{t=m}^{T} h_t(0|\mathbf{w}_t)\mathbf{Z}_{t,\tau}\mathbf{Z}_{t,\tau}^\top\right]$ and $\mathbf{\Psi}_T = E\left[T^{-1}\sum_{t=m}^{T}\left(\mathbf{Z}_{t,\tau} - \Gamma\mathbf{B}_T^{-1}\overline{\mathbf{M}}_t\right)\left(\mathbf{Z}_{t,\tau} - \Gamma\mathbf{B}_T^{-1}\overline{\mathbf{M}}_t\right)^\top\right]$, and assume that their inverse functions are uniformly bounded.

A6. $\sqrt{T}(\widehat{\mathbf{a}}_\tau - \mathbf{a}_\tau) = O_p(1)$, i.e., the estimators in the first step are consistent and the usual normalized difference is stochastically bounded.

**Remark 1.** *Assumption A1 is an invertibility condition which ensures that $q_{t,\tau}$ is a stationary process and that appropriate limiting theory can be applied. Assumptions A2-A3 are equivalent to Assumption AN2 in Engle and Manganelli (2004). Assumptions A4-A5 are used to establish the asymptotic normality of the quantile regression estimator in the second step, which is similar to Assumptions AN1(a) and AN3 in Engle and Manganelli (2004). Assumption A6 is a standard condition in the literature of two-step estimation method, e.g., Powell (1983) and Hautsch et al. (2015). Note that the minimum distance estimator of Xiao and Koenker (2009) satisfies this condition.*

## 3.2 Oracle Properties

The asymptotic normality and model selection consistency of the adaptive Lasso estimators are provided in this subsection. To simplify the presentation, we only describe the asymptotic results here, with all technical details relegated to the supplementary file. Next, we present the asymptotic representation of the quantile regression estimator $\widetilde{\boldsymbol{\theta}}$ and the oracle property of $\widehat{\boldsymbol{\theta}}^{(T)}$ in Theorems 1 and 2, respectively, as follows.

**Theorem 1.** *Recall that $\widetilde{\boldsymbol{\theta}}_\tau$ is the estimator of the quantile regression of $Y_t$ on $\widehat{\mathbf{Z}}_t$ for the quantile level $\tau$. Under Assumption A, one has*

1. *Asymptotic representation:*

$$\sqrt{T}(\widetilde{\boldsymbol{\theta}}_\tau - \boldsymbol{\theta}^*_\tau) = \mathbf{D}_T^{-1}\left(\frac{1}{\sqrt{T}}\sum_{t=m}^{T}\psi_\tau(Y_t - \mathbf{Z}_{t,\tau}^\top\boldsymbol{\theta}^*)\mathbf{Z}_{t,\tau}\right) - \mathbf{D}_T^{-1}\Gamma\sqrt{T}(\widehat{\mathbf{a}}_\tau - \mathbf{a}_\tau) + o_p(1),$$

*where* $\Gamma \equiv \sum_{i=1}^{q} E(h_t(0|\mathbf{w}_t)\mathbf{Z}_{t,\tau}\overline{\mathbf{M}}_{t-i}^\top)\beta_i^*$.

2. *Asymptotic normality:* $\sqrt{T}(\widetilde{\boldsymbol{\theta}}_\tau - \boldsymbol{\theta}^*_\tau) \xrightarrow{\mathcal{L}} N(0, \tau(1-\tau)\boldsymbol{\Omega})$, *where* $\boldsymbol{\Omega} \equiv \mathbf{D}_T^{-1}\boldsymbol{\Psi}_T\mathbf{D}_T^{-1}$.

To establish the model selection consistency of $\widehat{\boldsymbol{\theta}}^{(T)}$, some notations are provided. To this end, let $\mathcal{A} = \{j : \theta_j^* \neq 0\}$ and $\mathcal{A}_c = \{j : \theta_j^* = 0\}$. Suppose $\mathbf{C}$ is a matrix, then $\mathbf{C}_{\mathcal{AB}}$ is defined as the sub-matrix of $\mathbf{C}$ whose rows and columns are chosen from $\mathbf{C}$ according to the row index set $\mathcal{A}$ and column index set $\mathcal{B}$. For simplicity, we may write $\mathbf{C}_{\mathcal{AA}} = \mathbf{C}_{\mathcal{A}}$ when $\mathbf{C}$ is a square matrix. For a vector $\mathbf{C}$, $\mathbf{C}_{\mathcal{A}}$ denotes the sub-vector of $\mathbf{C}$ whose elements are chosen from $\mathbf{C}$ with index set $\mathcal{A}$. Then, we have the following theorem.

**Theorem 2.** *(Oracle) Suppose that Assumptions A1-A6 are satisfied. If* $\lambda_T T^{(\eta-1)/2} \to \infty$ *and* $\lambda_T/\sqrt{T} \to 0$, *then one has*

1. *Asymptotic normality:* $\sqrt{T}\left(\widehat{\boldsymbol{\theta}}_{\mathcal{A}}^{(T)} - \boldsymbol{\theta}^*_{\mathcal{A}}\right) \xrightarrow{\mathcal{L}} N(0, \tau(1-\tau)\boldsymbol{\Omega}_{\mathcal{A}})$.

2. *Sparsity:* $\widehat{\boldsymbol{\theta}}_{\mathcal{A}_c}^{(T)} = 0$.

**Remark 2.** *If the* $\sqrt{T}$*-consistent estimator in the first step is difficult to obtain in practical applications, it is easy to use a quantile autoregressive approximation such as* $\widehat{q}_{t,\tau} = \sum_{i=1}^{n}\widehat{\boldsymbol{\gamma}}_{i,\tau}^\top\mathbf{X}_{t-i}$, *where* $n$ *is the truncation parameter usually set to be a sufficiently large constant* $n = \log(T)$. *Xiao and Koenker (2009) proved that* $\|\widehat{\boldsymbol{\gamma}}_\tau - \boldsymbol{\gamma}_\tau\|^2 = O_p(n/T)$ *in the Theorem 1 of their paper. Under this condition, it is easy to show that the convergence rate of* $\widehat{\boldsymbol{\theta}}^{(T)}$ *is* $\sqrt{n/T}$ *with the asymptotic representation of the Theorem 1 here. However, although* $\widehat{\boldsymbol{\theta}}^{(T)}$ *is not* $\sqrt{T}$*-consistent, the sparsity property of the penalized estimator is maintained. The same situation for adaptive Lasso in mean regression has been discussed in Remark 1 in Zou (2006).*

# 4 Monte Carlo Simulation Study

In this section, a simulated example is used to illustrate the finite sample performance of the proposed model and the penalized CAViaR estimators. We consider sample sizes of $T = 400$ and 800, and simulations are repeated $M = 500$ times for each of the given sample sizes. Different quantile levels $\tau = 0.01$, 0.05, and 0.1 are considered. When generating the series of $Y_t$, the initial value is set to zero, and the first 200 observations are dropped to reduce the impact of the initial value. To measure the performance, the median and the standard deviation (SD) of the absolute deviation of errors (ADE) are reported, where $\text{ADE}_{\theta_j}^{(k)} \equiv \left|\widehat{\theta}_j^{(k)} - \theta_j\right|$ for $1 \le j \le (p+q+1)$, and $\widehat{\theta}_j^{(k)}$ is the estimator in the $k$-th simulation replication. Further, the positive rate PR is also reported, which is defined as the relative frequencies of picking the correct model.

**Example 1**. The DGP is given by

$$Y_t = q_{t,\tau} + \varepsilon_{t,\tau}, \quad \text{and} \quad q_{t,\tau} = \alpha_0 + \alpha_1 X_{t-1} + \alpha_6 X_{t-6} + \beta_3 q_{t-3,\tau}, \quad t = 1, \dots, T,$$

where $\alpha_0 = -0.2$, $\alpha_1 = 0.2$, $\alpha_6 = -0.2$, $\beta_3 = 0.6$, and $\varepsilon_{t,\tau}$ follows an i.i.d. tick-exponential family of $\tau$ with density function:

$$f(\varepsilon_{t,\tau}) = \frac{1}{\sigma_{\varepsilon_{t,\tau}}} \exp\left\{ \frac{\varepsilon_{t,\tau}}{\tau\sigma_{\varepsilon_{t,\tau}}} I(\varepsilon_{t,\tau} \le 0) - \frac{\varepsilon_{t,\tau}}{(1-\tau)\sigma_{\varepsilon_{t,\tau}}} I(\varepsilon_{t,\tau} > 0) \right\},$$

and $\sigma_{\varepsilon_{t,\tau}} = 1$, which can be found in Komunjer (2005). Clearly, the $\tau$th quantile of $\varepsilon_{t,\tau}$ equals 0, which satisfies the model identification condition. We then use the two-step estimation procedures introduced in Section 2 to estimate the parameters of the model. To save time, here, we use the quantile autoregressive approximation discussed in Remark 2. The truncation parameter $n$ is chosen to be $10\log_{10}(T)$, and the maximum of the order $(p, q)$ is pre-determined as $p = 10$ and $q = 3$, respectively. The median and SD (in parentheses) of the ADE values of the corresponding penalized CAViaR estimators $\widehat{\alpha}_0$ ($\text{ADE}_{\alpha_0}$), $\widehat{\alpha}_1$ ($\text{ADE}_{\alpha_1}$), $\widehat{\alpha}_6$ ($\text{ADE}_{\alpha_6}$), and $\widehat{\beta}_3$ ($\text{ADE}_{\beta_3}$) in all cases are reported in Table 1. Table 1 also reports the accuracy of the proposed method for selecting the correct model.

Table 1: Median(SD) of the ADE values and PR values under three quantiles.

| | $\tau = 0.01$ | | $\tau = 0.05$ | | $\tau = 0.1$ | |
|---|---|---|---|---|---|---|
| $T$ | 400 | 800 | 400 | 800 | 400 | 800 |
| $\text{ADE}_{\alpha_0}$ | 0.0766 | 0.0326 | 0.0581 | 0.0323 | 0.0537 | 0.0299 |
| | (0.0409) | (0.0175) | (0.0290) | (0.0158) | (0.0448) | (0.0166) |
| $\text{ADE}_{\alpha_1}$ | 0.0158 | 0.0074 | 0.0149 | 0.0123 | 0.0169 | 0.0153 |
| | (0.0167) | (0.0076) | (0.0172) | (0.0096) | (0.0230) | (0.0127) |
| $\text{ADE}_{\alpha_6}$ | 0.0169 | 0.0088 | 0.0168 | 0.0142 | 0.0216 | 0.0174 |
| | (0.0197) | (0.0087) | (0.0165) | (0.0114) | (0.0262) | (0.0142) |
| $\text{ADE}_{\beta_3}$ | 0.0503 | 0.0195 | 0.0512 | 0.0287 | 0.0713 | 0.0370 |
| | (0.0509) | (0.0207) | (0.0416) | (0.0226) | (0.0636) | 0.0263 |
| PR | 0.962 | 1.000 | 0.994 | 1.000 | 0.972 | 1.000 |

First, we find that the median and SD of ADE values for all the penalized CAViaR estimators decrease as the sample size increases. For example, when $T = 400$, the median and SD of the $\text{ADE}_{\alpha_1}$ are 0.0158 and 0.0167 under the quantile level $\tau = 0.01$, and they decrease to 0.0074 and 0.0076, respectively, when the sample size increases to 800. Clearly, the same pattern for $\text{ADE}_{\alpha_6}$ can also be observed. Indeed, when $\tau = 0.1$ and the sample size is 400, the median and the corresponding SD are 0.0216 and 0.0262. When the sample size increases to 800, the median and the corresponding SD decrease to 0.0174 and 0.0142, respectively. Finally, when the sample size is 400, the median and SD of $\text{ADE}_{\beta_3}$ are 0.0512 and 0.0416 under the quantile level $\tau = 0.05$, and they decrease to 0.0287 and 0.0226 as the sample size doubled.

Under all quantile levels considered, we find that the accuracy of the method for selecting the true model increases with the sample size. For example, when $T = 400$, the accuracy for selecting the correct model are 96.2% under the quantile level $\tau = 0.01$, and they increase to 100%, when the sample size increases to 800. Under the quantile level $\tau = 0.05$ and 0.1, the accuracy for selecting the correct model when $T = 400$ are 99.4% and 97.2%, and when the sample size is doubled, they increase to 100% and 100%, respectively.

# 5 Empirical Example

To illustrate the practical usefulness of the application of our proposed model, we consider the daily data of S&P500 from April 19, 2017, to March 31, 2023, with 1500 observations in total. Note that the first 1000 observations are used for in-sample model fitting, and the remaining 500 observations are for out-of-sample forecasting. The data are downloaded from CSMAR Database, and the daily returns are computed as the difference of the log transformation of the index; that is, $Y_t = \log(p_t/p_{t-1})$, where $p_t$ is the daily price. Table 2 reports the summary statistics of the return series. It clearly shows that, for the S&P500

Table 2: Summary statistics of return series.

| Mean | Min | Median | Max | S. Dev. | Skew. | Kurt. |
|---|---|---|---|---|---|---|
| 0.0004 | −0.1277 | 0.0008 | 0.0897 | 0.0129 | −0.8312 | 14.4950 |

return series, the sample mean is close to zero but the distribution is slightly negatively skewed and has fat tail. Figure 1 presents the histogram (left panel) and time series plot (right panel) for the series, and it shows that extreme values mainly occur during the beginning of 2020, the period of the outbreak of the Covid-19 epidemic. However, the return series is less volatile from 2017 to 2019.

Here, a study is conducted by comparing the accuracy of VaR predictions calculated by our new model with those calculated by some established methods. The benchmark models include the most popular techniques in both academia and industry: RiskMetrics (RM), the GARCH$(1,1)$ model with Gaussian innovations (GGARCH), the GARCH$(1,1)$ model with student's $t(4)$ innovations (TGARCH), the Gaussian GARCH$(1,1)$-EVT model with the threshold as the $100\tau\%$ unconditional quantile for the lower tail (GGARCH-EVT), student's $t(4)$ GARCH$(1,1)$-EVT model with the threshold as the $100\tau\%$ unconditional quantile for the lower tail (TGARCH-EVT), the quantile autoregressive model as $q_{t,\tau} = \alpha_{0,\tau} + \alpha_{1,\tau}Y_{t-1}$, and the symmetric absolute value CAViaR(1,1) model as $q_{t,\tau} = \alpha_{0,\tau} + \alpha_{1,\tau}|Y_{t-1}| + \beta_{2,\tau}q_{t-1,\tau}$.

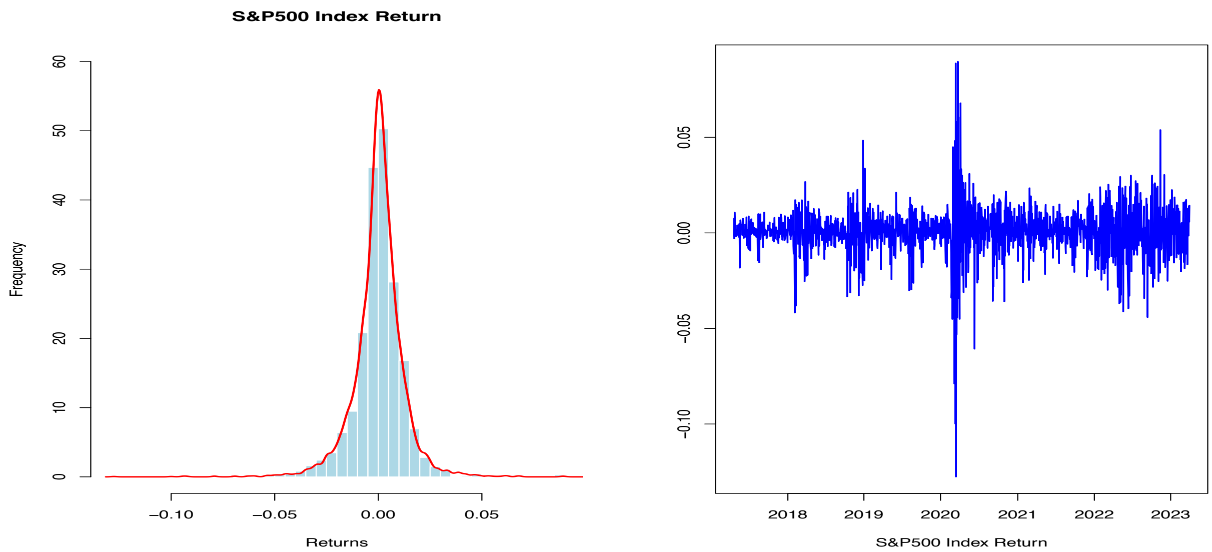


Figure 1: Time series and histogram plot of stock return series: S&P500.

Finally, we consider using the CAViaR$(p, q)$ model with $X_{t,i} = |Y_{t-i}|$, $i = 1, \ldots, p$, i.e.,

$$q_{t,\tau} = \alpha_{0,\tau} + \sum_{i=1}^{p} \alpha_{i,\tau}|Y_{t-i}| + \sum_{j=1}^{q} \beta_{j,\tau}e_{t-j,\tau}.$$

It is worthy noting that we set $p = 10$ and $q = 5$, and the model is selected through the estimation procedures introduced in Section 2.

To compare the relative performance of these methods in terms of predictive ability, all considered models are estimated on a rolling window of length 1000. For each of the windows, the day-ahead post-sample VaR predictions are computed from every method as the basis for our comparison purposes. Finally, the coverage ratio for each of the methods is computed, where the coverage ratio assesses the proportion of observations falling below the VaR predictions. Ideally, for estimation of the conditional $\tau$ quantile, the coverage ratio should be $\tau$. Therefore, the significant difference from the ideal case could be examined by implementing a test based on a binomial distribution. Besides, we also report the results of out-of-sample dynamic quantile (DQ) test proposed in Engle and Manganelli (2004) for validation.

Table 3 reports the coverage ratio and $p$-value (in parentheses) for post-sample predictions of conditional quantiles under three quantile levels, $\tau = 1\%, 5\%$, and $10\%$, in which the $p$-value is computed based on the significance test with perfect coverage ratio as null percentage. As shown in this table, the TGARCH model performs well for the 1% quantile, while the CAViaR$(p, q)$ model outperforms all the other models for the cases of $\tau = 5\%$ and $\tau = 10\%$. In all cases, only the QAR model, the CAViaR(1,1) model and the CAViaR$(p, q)$ model are not rejected by the binomial test for the 5% significance level. The DE test results validates

Table 3: Coverage ratio($p$-value) and $p$-value of DQ test statistics for 500 post-sample predictions of different levels of conditional quantiles

| | Coverage ratio test | | | DQ test | | |
|---|---|---|---|---|---|---|
| VaR% | 1% | 5% | 10% | 1% | 5% | 10% |
| RM | 2.2% (0.0118) | 7.8% (0.0038) | 12.6% (0.0525) | 0.0000 | 0.0015 | 0.0458 |
| GGARCH | 2.4% (0.0085) | 6.8% (0.0646) | 11.8% (0.1811) | 0.0071 | 0.3609 | 0.4108 |
| TGARCH | 1.0% (1.0000) | 7.4% (0.0132) | 12.8% (0.0366) | 0.7545 | 0.0329 | 0.2904 |
| GGARCH-EVT | 0.4% (0.1905) | 1.2% (0.0001) | 4.4% (0.0000) | 0.8889 | 0.0142 | 0.0056 |
| TGARCH-EVT | 0.2% (0.0709) | 1.2% (0.0001) | 4.4% (0.0000) | 0.7244 | 0.0142 | 0.0015 |
| QAR | 0.4% (0.1905) | 4.8% (0.8400) | 12.6% (0.0525) | 0.8553 | 0.0990 | 0.2388 |
| CAViaR(1,1) | 0.4% (0.1905) | 5.8% (0.4174) | 11.8% (0.1811) | 0.9087 | 0.3945 | 0.0519 |
| CAViaR($p$,$q$) | 0.4% (0.1905) | 5.0% (1.0000) | 11.6% (0.2350) | 0.9361 | 0.7636 | 0.2161 |

the result of the coverage ratio test, with the CAViaR($p, q$) model being the only model which is not rejected by the out-of-sample DE test for the 10% significance level. The empirical results suggest the usefulness of our method for the given dataset.

# 6 Conclusion

In this paper, we introduce a two-step approach for the model selection of the conditional autoregressive value at risk model. In the first step, the quantile lag components are approximated by a linear quantile regression model of the tail risk drivers. Then, the optimal CAViaR model can be selected by the adaptive Lasso penalized quantile regression. The asymptotic normality and oracle properties of the penalized CAViaR estimators are established. Finally, the proposed method is applied to the prediction of VaR for a real empirical example. The empirical results demonstrate that the CAViaR model with adaptive Lasso outperforms other benchmark models based on coverage ratio test and dynamic quantile test.

# Disclosure Satements

The authors claim that there is no conflict of interests about the manuscript submitted. Also, the authors declare that they do not use any generative AI and AI-assisted technologies in the writing process, to analyze and draw insights from data as part of the research process.

# References

Creal, D., Koopman, S. J., and Lucas, A. (2013). Generalized autoregressive score models with applications. *Journal of Applied Econometrics*, 28(5):777–795.

De Rossi, G. and Harvey, A. (2009). Quantiles, expectiles and splines. *Journal of Econometrics*, 152(2):179–185.

Engle, R. F. and Manganelli, S. (2004). CAViaR: conditional autoregressive value at risk by regression quantiles. *Journal of Business & Economic Statistics*, 22(4):367–381.

Fissler, T. and Ziegel, J. F. (2016). Higher order elicitability and Osband's principle. *Annals of Statistics*, 44(4):1680–1707.

Gao, L., Ye, W., and Guo, R. (2022). Jointly forecasting the value-at-risk and expected shortfall of Bitcoin with a regime-switching CAViaR model. *Finance Research Letters*, 48:102826.

Harvey, A. C. (2013). *Dynamic models for volatility and heavy tails: with applications to financial and economic time series*. Cambridge University Press.

Hautsch, N., Schaumburg, J., and Schienle, M. (2015). Financial network systemic risk contributions. *Review of Finance*, 19(2):685–738.

Komunjer, I. (2005). Quasi-maximum likelihood estimation for conditional quantiles. *Journal of Econometrics*, 128(1):137–164.

Kuester, K., Mittnik, S., and Paolella, M. S. (2006). Value-at-risk prediction: A comparison of alternative strategies. *Journal of Financial Econometrics*, 4(1):53–89.

Laporta, A. G., Merlo, L., and Petrella, L. (2018). Selection of value at risk models for energy commodities. *Energy Economics*, 74:628–643.

Patton, A. J., Ziegel, J. F., and Chen, R. (2019). Dynamic semiparametric models for expected shortfall (and value-at-risk). *Journal of Econometrics*, 211(2):388–413.

Powell, J. L. (1983). The asymptotic normality of two-stage least absolute deviations estimators. *Econometrica: Journal of the Econometric Society*, pages 1569–1575.

Taylor, J. W. (2008). Estimating value at risk and expected shortfall using expectiles. *Journal of Financial Econometrics*, 6(2):231–252.

Taylor, J. W. (2019). Forecasting value at risk and expected shortfall using a semiparametric approach based on the asymmetric laplace distribution. *Journal of Business & Economic Statistics*, 37(1):121–133.

Taylor, S. J. (1986). *Modelling financial time series*. New York: Wiley.

White, H. L., Kim, T. H., and Manganelli, S. (2015). VAR for VaR: measuring tail dependence using multivariate regression quantiles. *Journal of Econometrics*, 187(1):169–188.

Wu, Y. and Liu, Y. (2009). Variable selection in quantile regression. *Statistica Sinica*, 19:801–817.

Xiao, Z. and Koenker, R. (2009). Conditional quantile estimation for generalized autoregressive conditional heteroscedasticity models. *Journal of the American Statistical Association*, 104(488):1696–1712.

Zou, H. (2006). The adaptive lasso and its oracle properties. *Journal of the American statistical association*, 101(476):1418–1429.