Journal of Time Series Analysis

WILEY

ORIGINAL ARTICLE

CAViaR Model Selection via Adaptive Lasso

Zongwu Cai[1] | Ying Fang[2] | Dingshi Tian[3]

[1]Department of Economics, University of Kansas, Lawrence, Kansas, USA | [2]Wang Yanan Institute for Studies in Economics and Department of Statistics and Data Science, School of Economics, Xiamen University, Xiamen, Fujian, China | [3]Department of Statistics, School of Statistics and Mathematics, Zhongnan University of Economics and Law, Wuhan, Hubei, China

Correspondence: Dingshi Tian (dstian@zuel.edu.cn)

Received: 26 March 2024 | **Revised:** 2 November 2024 | **Accepted:** 11 November 2024

Funding: National Natural Science Foundation of China, Grant/Award Numbers: 72103207; 72133002; 72033008; 71988101; Fundamental Research Funds for the Central Universities for Zhongnan University of Economics and Law, Grant/Award Number: 2722023BY018; Humanities and Social Science Research Foundation of Ministry of Education of China, Grant/Award Number: 22YJAZH038

Keywords: adaptive lasso | CAViaR model | model selection | tail risk

ABSTRACT

The estimation and model selection of the 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, conventional automatic variable selection methods cannot be directly applied to this problem because the quantile lag components are latent. In this paper, we propose a two-step approach to select the optimal CAViaR model. 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 regularity conditions, the proposed adaptive Lasso penalized quantile estimators enjoy the oracle properties. Finally, the proposed method is illustrated by a Monte Carlo simulation study and applied to analyzing the daily data of the S& P 500 return series.

1 | Introduction

To find an effective tail risk measure, Engle and Manganelli (2004) proposed seminally a conditional autoregressive value at risk (CAViaR) model, which has been widely applied to various fields due to its simplicity and a capability to grasp the evolution of the quantiles over time. As a financial risk management tool, for example, Kuester, Mittnik, and Paolella (2006) showed that an extension to a particular CAViaR model outperforms other alternative strategies in the value at risk (VaR) prediction. Recently, Laporta, Merlo, and Petrella (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, Ye, and Guo (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 cases by White, Kim, and Manganelli (2015). Besides, it was incorporated into the unobserved components model by Harvey (2013) and the generalized autoregressive score model by Creal, Koopman, and Lucas (2013), respectively. To overcome the problem of "elicitability" for ES, Taylor (2019) and Patton, Ziegel, and Chen (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 the CAViaR model has attracted a great deal of research attention, the inclusion of autoregressive components may bring a 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 calculating the CAViaR model. By establishing the relationship between the linear GARCH model as in 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 the CAViaR model. When the true orders of quantile lags or the dimension of the other regressors are high, the implementation 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 existence of latent variables. In this article, 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 article is organized as follows: Section 2 introduces the CAViaR model and our two-step estimation and model selection procedures. In the same section, the asymptotic properties of the two-step CAViaR estimators are investigated, 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 3 and 4, respectively. Section 5 concludes the paper. All technical proofs are deferred to Appendix.

2 | Model Framework

2.1 | Model Setup

Assume that $\{(Y_t, \mathbf{X}_t)\}_{t=1}^{\infty}$ is a sequence of strictly stationary random vectors, where Y_t is a scalar variable of interest and $\mathbf{X}_t = (X_{t,1}, \dots, X_{t,p})^\top \in \mathcal{R}^p$ is a vector of covariates defining risk factors. Let $\mathcal{F}_t = \sigma\{(Y_s, \mathbf{X}_s); s < t\}$ be the information set available at time t and $q_{t,\tau}(\theta_\tau)$ denotes the τth conditional quantile of Y_t given $\mathcal{F}_t$, which is dynamic to characterize quantile changing over time. That is, the τth conditional quantile of Y_t given $\mathcal{F}_t$ is defined as

$$q_{t,\tau} = q_{t,\tau}(\theta_\tau) = \arg\min_{u\in\mathcal{R}} E\{\mathrm{Q}_\tau(Y_t - u)|\mathcal{F}_t\}$$

where $\mathrm{Q}_\tau(v) = (\tau - \mathrm{I}(v \le 0))v$ with $\mathrm{I}(\cdot)$ denoting an indicator function. To specify an autoregressive evolution of quantiles over time, similar to an ARMA(p, q) model, Engle and Manganelli (2004) proposed a class of CAViaR (p, q) models as

$$\begin{aligned} q_{t,\tau} &= \alpha_{0,\tau} + \sum_{i=1}^{p} \alpha_{i,\tau} X_{t,i} + \sum_{j=1}^{q} \beta_{j,\tau} q_{t-j,\tau} \\ &= \alpha_{0,\tau} + \boldsymbol{\alpha}_\tau^\top \mathbf{X}_t + B(L) q_{t,\tau} \end{aligned} \tag{1}$$

where $\boldsymbol{\theta}_\tau = (\alpha_{0,\tau}, \alpha_{1,\tau}, \dots, \alpha_{p,\tau}, \beta_{1,\tau}, \dots, \beta_{q,\tau})^\top \in \mathcal{R}^{p+q+1}$ with the orders $p \ge 0$ and $q \ge 0$, $\alpha_{0,\tau}$ represents the intercept, $\alpha_{i,\tau}$ is the i-th element of $\boldsymbol{\alpha}_\tau = (\alpha_{1,\tau}, \dots, \alpha_{p,\tau})^\top$ with $\boldsymbol{\alpha}_\tau$ denoting a vector of coefficients for $\mathbf{X}_t$, $\beta_{j,\tau}$ is the coefficient of $q_{t-j,\tau}$, $j = 1, \dots, q$, denoting the jth lag of $q_{t,\tau}$, 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 lags of Y_t and possible transformations of the lags of Y_t such as $X_{t,i_0} = |Y_{t-d_0}|^\gamma$ for some $\gamma > 0$, where $1 \le i_0 \le p$ and $d_0 \ge 1$. Note that for simplicity of notation, τ is dropped from $\boldsymbol{\theta}_\tau$ as $\boldsymbol{\theta}$ without causing confusion.

2.2 | Model Selection Procedure

The key issue of implementing this method is how to select an appropriate and parsimonious model, which is similar to choosing 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, for the given data $\{(Y_t, \mathbf{X}_t)\}_{t=1}^{T}$, we estimate the conditional quantile of Y_t using an approximately parametric quantile regression model of the tail risk drivers $\mathbf{M}_t$, which may contain some or all components of $\mathbf{X}_t$, some lagged macroeconomic state variables, lagged firm-specific characteristics, and lags of Y_t and their functions, as follows:

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

where $\mathbf{a}_\tau = (a_{0,\tau}, a_{1,\tau}, \dots, a_{d,\tau})^\top$ and $\overline{\mathbf{M}}_t$ is a 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 function of the tail risk drivers $\mathbf{M}_t$) or other types 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 quantile $q_{t,\tau}$ can be estimated by the quantile regression of Y_t on $\mathbf{M}_t$:

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

where $\hat{a}_{0,\tau}$ and $\hat{a}_{j,\tau}$ are the quantile regression estimators of the first and $(j+1)$-th elements, respectively, and $M_{t,j}$ is the j-th element of the tail risk drivers $\mathbf{M}_t$. As mentioned earlier, the selection of $\mathbf{M}_t$ is determined by all possible information from the dependent variable Y_t and its risk factors $\mathbf{X}_t$. For example, if Y_t represents the stock return of an individual company, then the lagged macroeconomic state variables, the lagged firm-specific characteristics, and the lagged returns Y_{t-i}, $i = 1, 2, \dots$ and their functions, are recommended to be included in $\mathbf{M}_t$.

In the second step, first, define $\mathbf{Z}_{t,\tau} = (1, X_{t,1}, \dots, X_{t,p}, q_{t-1,\tau}, \dots, q_{t-q,\tau})^\top$ and plug-in $\hat{q}_{t-q,\tau}$ into $\mathbf{Z}_{t,\tau}$ to obtain $\hat{\mathbf{Z}}_{t,\tau} = (1, X_{t,1}, \dots, X_{t,p}, \hat{q}_{t-1,\tau}, \dots, \hat{q}_{t-q,\tau})^\top$. Then, the optimal subset CAViaR model can be selected by the adaptive Lasso penalized quantile regression of Y_t on $\hat{\mathbf{Z}}_{t,\tau}$ and the adaptive Lasso penalized quantile regression estimator of $\boldsymbol{\theta}$ is obtained by:

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

where $m = \max\{p, q\} + 1$, $\lambda_T > 0$ is the tuning parameter, and the adaptive weights $\boldsymbol{\omega} \equiv (w_1, \dots, w_{p+q+1})^\top$ consist of $p + q +$

1 data-driven weights. These weights are calculated iteratively using the formula $\hat{\boldsymbol{\omega}} = |\hat{\theta}|^{-\eta}$, where $\eta > 0$ is an appropriately chosen constant. To derive an initial value of $\boldsymbol{\omega}$, we employ the quantile regression estimator $\tilde{\theta}$ of Y_t on $\hat{\mathbf{Z}}_{t,\tau}$, which can be estimated by

$$\tilde{\theta} = \min_{\theta \in \mathcal{R}^{p+q+1}} \frac{1}{T} \sum_{t=m}^{T} \mathrm{Q}_\tau \left(Y_t - \alpha_0 - \sum_{i=1}^{p} \alpha_i X_{t,i} - \sum_{j=1}^{q} \beta_j \hat{q}_{t-j,\tau} \right)$$

Specifically, the estimation procedures in the second step are formulated in the following algorithm.

Algorithm:

1. Calculate the initial value of the adaptive weights by running quantile regression of Y_t on $\hat{\mathbf{Z}}_{t,\tau}$, which are given by $\hat{\boldsymbol{\omega}}^{(0)} = |\tilde{\theta}|^{-1}$, by choosing $\eta = 1$ for simplicity.
2. Update the adaptive weights by adaptive Lasso regression of Y_t on $\hat{\mathbf{Z}}_{t,\tau}$. Specifically, the adaptive weights of the k-th iteration can be calculated by $\hat{\boldsymbol{\omega}}^{(k)} = |\hat{\theta}^{(k)}|^{-1}$, where $\hat{\theta}^{(k)}_\tau$ is estimated using (2) with adaptive weights $\hat{\boldsymbol{\omega}}^{(k-1)}$.
3. Find the solution path of the adaptive Lasso penalized quantile regression.
4. The optimal λ_T is selected using the cross validation technique, as recommended by Zou (2006).
5. Keep iterating (say, 2-4 steps) until convergence achieved.

2.3 | Asymptotic Theory

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

2.3.1 | Notations and Assumptions

Some necessary assumptions for deriving asymptotic results are listed below. Note that these assumptions given in this article are sufficient conditions but not necessarily the weakest. First, define $\Gamma = \sum_{i=1}^{q} E\left[h_t(0|\mathcal{F}_t)\mathbf{Z}_{t,\tau}\overline{\mathbf{M}}_{t-i}^\top \beta_i\right]$.

Assumptions:

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

A2. The process $\{(Y_t, \mathbf{X}_t)\}_{t=1}^{\infty}$ is (strong) α-mixing with mixing coefficient $\alpha(\cdot)$ of size $-r/(r-1)$, with $r > 1$.

A3. Conditional on $\mathcal{F}_t$, the error term $\varepsilon_{t,\tau} \equiv Y_t - q_{t,\tau}$ forms a stationary process, with a continuous conditional density $h_t(\varepsilon|\mathcal{F}_t) \geq M_0$ for some $M_0 > 0$ and for all t. Further, for a constant M_1 and for any t, $h_t(\varepsilon|\mathcal{F}_t) \leq M_1 < \infty$.

A4. $h_t(\varepsilon|\mathcal{F}_t)$ satisfies the Lipschitz condition, i.e., $|h_t(\lambda_1|\mathcal{F}_t) - h_t(\lambda_2|\mathcal{F}_t)| \leq L_0|\lambda_1 - \lambda_2|$ for a constant $L_0 < \infty$ and for any t.

A5. Define $\boldsymbol{\Psi} = \lim_{T\to\infty} T^{-1}\sum_{t=1}^{T}\sum_{s=1}^{T} E\left[\left(\mathbf{Z}_{t,\tau} - \Gamma\mathbf{B}^{-1}\overline{\mathbf{M}}_t\right)\left(\mathbf{Z}_{s,\tau} - \Gamma\mathbf{B}^{-1}\overline{\mathbf{M}}_s\right)^\top\right]$, where $\mathbf{B} = E\left[h_t(0|\mathcal{F}_t)\overline{\mathbf{M}}_t\overline{\mathbf{M}}_t^\top\right]$. Also, let $\mathbf{D} = E\left[h_t(0|\mathcal{F}_t)\mathbf{Z}_{t,\tau}\mathbf{Z}_{t,\tau}^\top\right]$. We assume that the inverse of three matrices, $\mathbf{B}$, $\boldsymbol{\Psi}$ and $\mathbf{D}$, are bounded.

A6. There exist some stochastic functions of variables from the information set $\mathcal{F}_t$, $A(\mathcal{F}_t)$, $C(\mathcal{F}_t)$, $H(\mathcal{F}_t)$, and $K(\mathcal{F}_t)$ such as for all t, $\|\mathbf{M}_t\| \leq A(\mathcal{F}_t)$, $\|\mathbf{X}_t\| \leq C(\mathcal{F}_t)$, $q_{t,\tau} \leq H(\mathcal{F}_t)$ and $\|\mathbf{Z}_{t,\tau} - \Gamma\mathbf{B}^{-1}\overline{\mathbf{M}}_t\| \leq K(\mathcal{F}_t)$, satisfying $E(|A^2(\mathcal{F}_t)|) < \infty$, $E(|C^2(\mathcal{F}_t)|) < \infty$, $E(|A(\mathcal{F}_t)C(\mathcal{F}_t)|) < \infty$, $E(|H^2(\mathcal{F}_t)|) < \infty$, $E(|C(\mathcal{F}_t)H(\mathcal{F}_t)|) < \infty$, and $E(|K(\mathcal{F}_t)|^{2r+\epsilon}) < \infty$, for some $\epsilon > 0$.

A7. $\sqrt{T}(\hat{\mathbf{a}}_\tau - \mathbf{a}_\tau) = O_p(1)$, that is, the estimators in the first step are consistent and the usual normalized difference is stochastically bounded.

Remark 1. Assumption A1 is an invertibility condition that ensures that $q_{t,\tau}$ is a stationary process so that appropriate limiting theories can be applied. Assumption A2 ensures that the dataset is the realization of an α-mixing stochastic process. Assumptions A3-A4 are equivalent to Assumption AN2 in Engle and Manganelli (2004). Assumptions A5-A6 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 A7 is a standard condition in the literature of the two-step estimation method, e.g., Powell (1983), and Hautsch, Schaumburg, and Schienle (2015). Note that the minimum distance estimator of Xiao and Koenker (2009) satisfies this condition.

2.3.2 | 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 Appendix. Next, we present the asymptotic representation of the quantile regression estimator $\tilde{\theta}$ and the oracle property of $\hat{\theta}$ in Theorems 1 and 2, respectively, as follows.

Theorem 1. *Recall that $\tilde{\theta}$ is the estimator of the quantile regression of Y_t on $\hat{\mathbf{Z}}_{t,\tau}$ for the quantile level τ. Under Assumptions A1–A7, one has*

1. *Asymptotic representation:*

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

where $\psi_\tau(v) = \tau - I(v \leq 0)$.

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

To establish the oracle property of the proposed model selection procedure, some notations are provided. To this end, let $\mathcal{A} = \{j : \theta_{j,\tau} \neq 0\}$ and $\mathcal{A}_c = \{j : \theta_{j,\tau} = 0\}$. Then, we have the following theorem.

Theorem 2. *(Oracle) Suppose that Assumptions A1-A7 are satisfied. If $\lambda_T T^{(\eta-1)/2} \to \infty$ and $\lambda_T/\sqrt{T} \to 0$, then one has*

1. *Asymptotic normality:* $\sqrt{T}\left(\hat{\theta}_{\mathcal{A}} - \theta_{\mathcal{A}}\right) \xrightarrow{\mathcal{L}} N(0, \tau(1-\tau)\mathbf{\Omega}_{\mathcal{A}})$,
2. *Sparsity:* $\hat{\theta}_{\mathcal{A}_c} = 0$,

where $\hat{\theta}_{\mathcal{A}}$ denotes the sub-vector of $\hat{\theta}$ whose elements are chosen from $\hat{\theta}$ with index set $\mathcal{A}$, $\mathbf{\Omega}_{\mathcal{A}}$ is defined as the sub-matrix of $\mathbf{\Omega}$ whose rows and columns are chosen from $\mathbf{\Omega}$ according to the row and column indexes set $\mathcal{A}$, and $\theta_{\mathcal{A}}$ denotes the sub-vector of θ whose elements are chosen from θ with index set $\mathcal{A}$.

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 $\hat{q}_{t,\tau} = \sum_{i=1}^{n_0} \hat{\gamma}_{i,\tau}^{\top} \mathbf{X}_{t-i}$, where n_0 is a truncation parameter usually setting to be a sufficiently large constant in applications.

3 | Monte Carlo Simulation Study

In this section, two simulated examples are used to illustrate the finite sample performance of the proposed model and the penalized CAViaR estimators. Simulations are repeated $M = 500$ times. When generating the series of Y_t, the initial value is set to be zero, and the first 200 observations are dropped to reduce the impact from the initial value. To measure the performance, the median and the standard deviation (SD) of the absolute deviation of errors (ADE) are reported, where $\text{ADE}_{\theta_j}^{(k)} \equiv \left|\hat{\theta}_j^{(k)} - \theta_j\right|$ for $1 \le j \le (p+q+1)$, and $\hat{\theta}_j^{(k)}$ 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

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

where $\alpha_0 = -0.2$, $\alpha_1 = 0.2$, $\alpha_6 = -0.2$, $\beta_3 = 0.6$, X_t is generated from an i.i.d. Gaussian distribution, X_{t-1} and X_{t-6} are lagged variables of X_t, and $\varepsilon_{t,\tau}$ follows an i.i.d. tick-exponential family of τ 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 τth quantile of $\varepsilon_{t,\tau}$ equals 0, which satisfies the model identification condition. We analyze two sample sizes, specifically $T = 400$ and $T = 800$. Additionally, different quantile levels $\tau = 0.01, 0.05$, and 0.10 are considered. We then use the two-step estimation procedures introduced in Section 2 for model selection and parameter estimation. To save computing time, here we apply the quantile autoregressive approximation outlined in Remark 2 for the estimation in the first step, and the truncation parameter n_0 is chosen to be $10\log_{10}(T)$. To examine the sensitivity of our estimation procedure to the maximum order of (p, q), we consider three distinct settings of (p, q). Specifically, in Setting A, we set $p = 10$ and $q = 6$; in Setting B, $p = 8$ and $q = 8$; and in Setting C, $p = 6$ and $q = 10$. The median and SD (in parentheses) of the ADE values of the corresponding penalized CAViaR estimators $\hat{\alpha}_0$ (ADE_{α_0}), $\hat{\alpha}_1$ (ADE_{α_1}), $\hat{\alpha}_6$ (ADE_{α_6}), and $\hat{\beta}_3$ (ADE_{β_3}) for all cases are summarized in Table 1, which also reports the accuracy of the proposed method for selecting the correct model.

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 $T = 400$ and the quantile level is $\tau = 0.01$, the accuracy for selecting the correct model is 91.8% under Setting A, and it increases to 100% when the sample size increases to 800. Under the quantile levels $\tau = 0.05$ and 0.10, the accuracies for selecting the correct model when $T = 400$ are 94.2% and 89.8%, respectively, for Setting B, and when the sample size is doubled, they increase to 100% and 99.8%, respectively.

Further, 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 $T = 400$ and Setting C is considered, the median and SD of the ADE_{α_1} are 0.0148 and 0.0185, respectively, under the quantile level $\tau = 0.01$, and both decrease to 0.0079 and 0.0081, respectively, when the sample size increases to 800. The same pattern for ADE_{α_6} can also be observed. Indeed, when $\tau = 0.10$ and the sample size is 400, the median and SD are 0.0237 and 0.0258, respectively, for Setting B. When the sample size increases to 800, the median and SD decrease to 0.0152 and 0.0130, respectively. Finally, when the sample size is 400, the median and SD of ADE_{β_3} for Setting A are 0.0518 and 0.0766, respectively, under the quantile level $\tau = 0.05$, and they decrease to 0.0289 and 0.0230 as the sample size is doubled.

Finally, the simulation results also conclude that both the selection and estimation accuracy of the model are not significantly affected by the choice of model settings concerning the maximum order (p, q). An analysis of Table 1 reveals that the accuracy of our proposed method for selecting the correct model is highly consistent across the three settings. For instance, with a sample size of $T = 400$, the PRs for Setting A are 91.8%, 96.4%, and 93.6% for the quantile levels at $\tau = 0.01$, 0.05, and 0.10, respectively. Under the same sample size and quantile levels, the PRs are 90.2%, 94.2%, and 89.8%, respectively, for Setting B. The same pattern can be observed for the estimation of the coefficients. For example, for Setting B with a sample size of $T = 800$, the median values of ADE_{α_0}, ADE_{α_1}, ADE_{α_6}, and ADE_{β_3} at $\tau = 0.05$ are 0.0322, 0.0125, 0.0133, and 0.0275, respectively. Under the same sample size, the medians of ADE_{α_0}, ADE_{α_1}, ADE_{α_6}, and ADE_{β_3} are 0.0335, 0.0119, 0.0118, and 0.0300, respectively, for Setting C. In summary, the simulation results indicate that both the selection and estimation accuracy of the model are not significantly affected by the choice of model settings concerning the maximum order (p, q).[1]

Example 2. In this example, the case of time series random errors is considered to evaluate the robustness of our method. The DGP is given by

$$Y_t = q_{t,\tau} + 0.5 Z_t \varepsilon_{t,\tau}, \quad \text{and}$$
$$q_{t,\tau} = \alpha_0 + \alpha_1 X_{t-1} + \alpha_3 X_{t-3} + \beta_3 q_{t-3,\tau}, \quad t = 1, \dots, T$$

TABLE 1 | Median (SD) of the ADE values and PR values under three quantiles.

T	$\tau = 0.01$		$\tau = 0.05$		$\tau = 0.10$	
	400	**800**	**400**	**800**	**400**	**800**
Setting A						
ADE_{α_0}	0.0834	0.0333	0.0574	0.0335	0.0495	0.0306
	(0.0727)	(0.0172)	(0.0609)	(0.0154)	(0.0810)	(0.0175)
ADE_{α_1}	0.0158	0.0077	0.0141	0.0118	0.0198	0.0146
	(0.0177)	(0.0073)	(0.0177)	(0.0097)	(0.0201)	(0.0126)
ADE_{α_6}	0.0191	0.0083	0.0174	0.0125	0.0199	0.0161
	(0.0264)	(0.0081)	(0.0200)	(0.0104)	(0.0294)	(0.0130)
ADE_{β_3}	0.0527	0.0238	0.0518	0.0289	0.0659	0.0369
	(0.0750)	(0.0188)	(0.0766)	(0.0230)	(0.0789)	(0.0277)
PR	0.918	1.000	0.964	1.000	0.936	1.000
Setting B						
ADE_{α_0}	0.0793	0.0315	0.0556	0.0322	0.0549	0.0323
	(0.0420)	(0.0176)	(0.0642)	(0.0151)	(0.0594)	(0.0180)
ADE_{α_1}	0.0163	0.0074	0.0159	0.0125	0.0168	0.0147
	(0.0190)	(0.0073)	(0.0172)	(0.0099)	(0.0200)	(0.0118)
ADE_{α_6}	0.0178	0.0091	0.0164	0.0133	0.0237	0.0152
	(0.0227)	(0.0081)	(0.0232)	(0.0117)	(0.0258)	(0.0130)
ADE_{β_3}	0.0522	0.0218	0.0509	0.0275	0.0716	0.0387
	(0.0466)	(0.0219)	(0.0611)	(0.0215)	(0.0722)	(0.0291)
PR	0.902	0.998	0.942	1.000	0.898	0.998
Setting C						
ADE_{α_0}	0.0814	0.0343	0.0549	0.0335	0.0505	0.0304
	(0.0734)	(0.0174)	(0.0739)	(0.0166)	(0.0918)	(0.0174)
ADE_{α_1}	0.0148	0.0079	0.0148	0.0119	0.0191	0.0157
	(0.0185)	(0.0081)	(0.0147)	(0.0102)	(0.0170)	(0.0129)
ADE_{α_6}	0.0187	0.0092	0.0181	0.0118	0.0214	0.0166
	(0.0307)	(0.0092)	(0.0244)	(0.0118)	(0.0346)	(0.0158)
ADE_{β_3}	0.0581	0.0245	0.0550	0.0300	0.0672	0.0381
	(0.0659)	(0.0201)	(0.0720)	(0.0248)	(0.0903)	(0.0294)
PR	0.842	0.994	0.926	0.998	0.866	0.994

TABLE 2 | Model selection results under different quantiles levels.

T	$\tau = 0.01$			$\tau = 0.05$			$\tau = 0.10$			$\tau = 0.25$		
	1000	**2000**	**5000**	**1000**	**2000**	**5000**	**1000**	**2000**	**5000**	**1000**	**2000**	**5000**
PR	27.6%	51.0%	87.0%	90.8%	97.4%	100.0%	98.6%	99.2%	100.0%	99.2%	99.4%	100.0%

where $Z_t = 0.2Z_{t-1} - 0.2Z_{t-2} + v_t$ with $v_t \sim N(0, 0.5)$, $\alpha_0 = -0.2$, $\alpha_1 = 0.2$, $\alpha_3 = -0.2$, $\beta_3 = 0.6$, X_t is generated from an i.i.d. Gaussian distribution, X_{t-1} and X_{t-3} are lagged variables of X_t, and $\varepsilon_{t,\tau}$ follows an i.i.d. tick-exponential family of τ as detailed in Example 1. Similarly, the two-step estimation procedures introduced in Section 2 are used to estimate the parameters of the model, and the quantile autoregressive approximation is employed in the first step. The truncation parameter is set to $n_0 = 10\log_{10}(T)$, and the maximum of the order (p, q) is set to $(p = 5, q = 5)$. Table 2 presents the accuracy of our proposed method in correctly selecting the model under various quantile levels $\tau = 0.01, 0.05, 0.10$, and 0.25.

Across all quantile levels, it is observed that the accuracy of our method for identifying the true model improves with an increase in sample size. For instance, at the quantile levels $\tau = 0.05, 0.10,$

TABLE 3 | 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

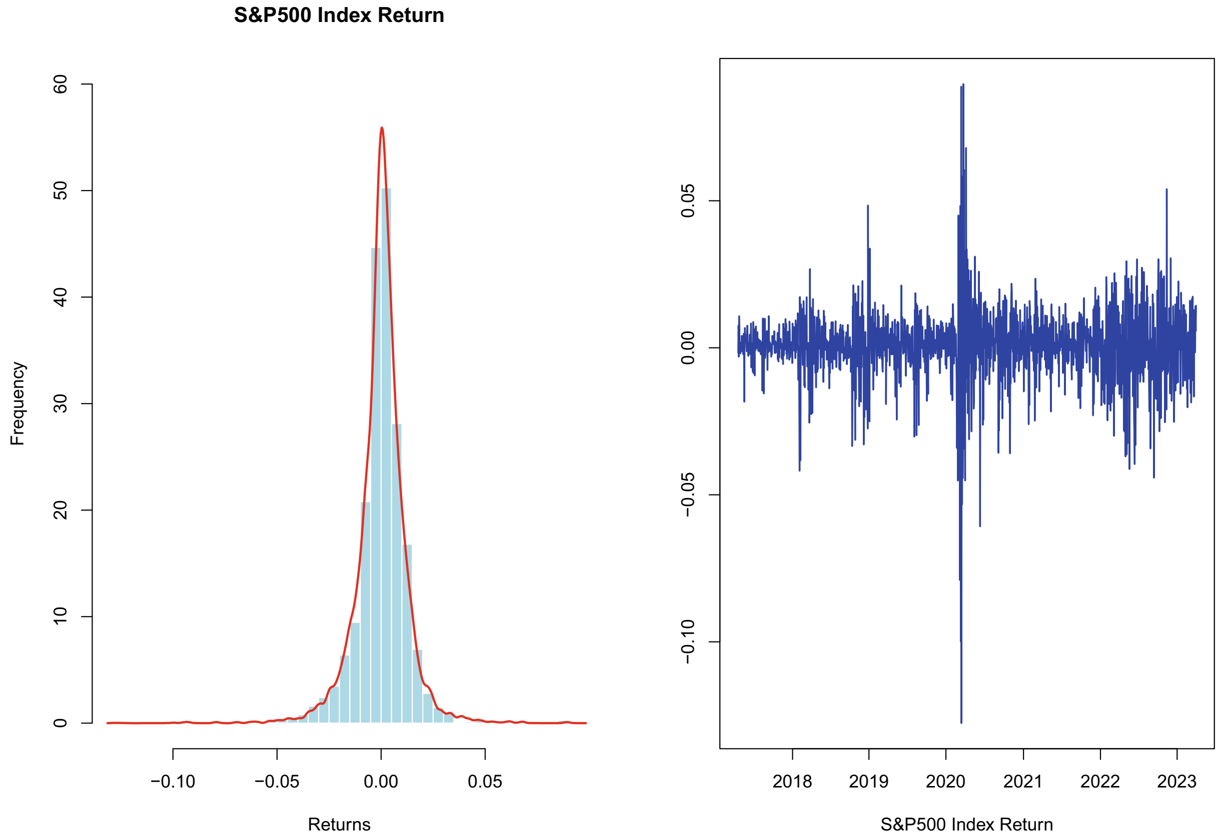

FIGURE 1 | Time series and histogram plot of stock return series: S&P500.

and 0.25, the PRs are 90.8%, 98.6%, and 99.2%, respectively, when the sample size $T = 1,000$. Upon doubling the sample size, the PRs further increase to 97.4%, 99.2%, and 99.4%, respectively, for the same quantile levels. However, it is not surprising that the model selection accuracy may be influenced by the dependent random errors for extremely low quantiles, as evidenced by a PR of 51.0% for the quantile level $\tau = 0.01$ when the sample size is $T = 2,000$. This issue is mitigated when the sample size increases to $T = 5,000$.

4 | An 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 1,500 observations in total. Note that the first 1,000 observations are used for in-sample model fitting, and the remaining 500 observations are for out-of-sample forecasting. The data are downloaded from the 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 3 reports the summary statistics of the return series. It clearly shows that, for the S&P500 return series, the sample mean is close to zero but the distribution is slightly negatively skewed and has a 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 in the period 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), GARCH(1,1) model with Gaussian innovations (GGARCH), GARCH(1,1) model with student's $t(4)$ innovations (TGARCH), 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 (QAR) model as $q_{t,\tau} = \alpha_{0,\tau} + \alpha_{1,\tau} Y_{t-1}$, and the symmetric absolute value CAViaR(1,1) model as $q_{t,\tau} = \alpha_{0,\tau} + \alpha_{1,\tau}|Y_{t-1}| + \beta_{2,\tau} q_{t-1,\tau}$. Finally, we consider using CAViaR(p, q) model with $X_{t,i} = |Y_{t-i}|$, $i = 1, \ldots, p$, that is,

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

TABLE 4 | 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

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 models considered are estimated on a rolling window of length 1,000. 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 of each method is computed, where the coverage ratio assesses the proportion of observations falling below the VaR predictions. Ideally, for estimation of the conditional τth quantile, the coverage ratio should be τ. 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 4 reports the coverage ratios and their p-values (in parentheses) for post-sample predictions of conditional quantiles under three quantile levels $\tau = 1\%$, 5%, and 10%, in which the p-values are computed based on a significance test with perfect coverage ratio as the 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 survive by the binomial test for the 5% significance level. The DQ test further validates the results of the coverage ratio test. The CAViaR(p, q) model is the only model not rejected by the out-of-sample DQ test for the 10% significance level. The empirical results suggest the usefulness of our method for the given dataset.

5 | Conclusion

In this article, 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 simulation data and the prediction of VaR for a real empirical example. The empirical results demonstrate that the CAViaR model with adaptive Lasso outperforms other benchmark models in terms of coverage ratio test and dynamic quantile test.

Acknowledgements

The authors thank the Editor-in-Chief (Professor Robert Taylor), Associate Editor, and the anonymous referees for their great comments and suggestions that improved significantly the quality and presentation of the paper. Also, they gratefully acknowledge the financial supports, in part, from the Natural Science Foundation of China with grant numbers *72103207, 72133002, 72033008 and 71988101*, 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 with grant number *22YJAZH038*.

Disclosure

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.

Conflicts of Interest

The authors declare no conflicts of interest.

Data Availability Statement

The data that support the findings of this study are available from the corresponding author upon reasonable request.

Endnotes

[1] We thank the anonymous referee's suggestion to examine this sensitivity analysis.

References

Creal, D., S. J. Koopman, and A. Lucas. 2013. "Generalized Autoregressive Score Models With Applications." *Journal of Applied Econometrics* 28, no. 5: 777–795.

De Rossi, G., and A. Harvey. 2009. "Quantiles, Expectiles and Splines." *Journal of Econometrics* 152, no. 2: 179–185.

Engle, R. F., and S. Manganelli. 2004. "CAViaR: Conditional Autoregressive Value at Risk by Regression Quantiles." *Journal of Business & Economic Statistics* 22, no. 4: 367–381.

Fissler, T., and J. F. Ziegel. 2016. "Higher Order Elicitability and Osband's Principle." *Annals of Statistics* 44, no. 4: 1680–1707.

Gao, L., W. Ye, and R. Guo. 2022. "Jointly Forecasting the Value-At-Risk and Expected Shortfall of Bitcoin With a Regime-Switching CAViaR Model." *Finance Research Letters* 48: 102826.

Geyer, C. J. 1994. "On the Asymptotics of Constrained M-Estimation." *Annals of Statistics* 22, no. 4: 1993–2010.

Harvey, A. C. 2013. *Dynamic Models for Volatility and Heavy Tails: With Applications to Financial and Economic Time Series*. New York: Cambridge University Press.

Hautsch, N., J. Schaumburg, and M. Schienle. 2015. "Financial Network Systemic Risk Contributions." *Review of Finance* 19, no. 2: 685–738.

Komunjer, I. 2005. "Quasi-Maximum Likelihood Estimation for Conditional Quantiles." *Journal of Econometrics* 128, no. 1: 137–164.

Kuester, K., S. Mittnik, and M. S. Paolella. 2006. "Value-At-Risk Prediction: A Comparison of Alternative Strategies." *Journal of Financial Econometrics* 4, no. 1: 53–89.

Laporta, A. G., L. Merlo, and L. Petrella. 2018. "Selection of Value at Risk Models for Energy Commodities." *Energy Economics* 74: 628–643.

Patton, A. J., J. F. Ziegel, and R. Chen. 2019. "Dynamic Semiparametric Models for Expected Shortfall (and Value-at-Risk)." *Journal of Econometrics* 211, no. 2: 388–413.

Powell, J. L. 1983. "The Asymptotic Normality of Two-Stage Least Absolute Deviations Estimators." *Econometrica* 51, no. 5: 1569–1575.

Taylor, J. W. 2008. "Estimating Value at Risk and Expected Shortfall Using Expectiles." *Journal of Financial Econometrics* 6, no. 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, no. 1: 121–133.

Taylor, S. J. 1986. *Modeling Financial Time Series*. New York: Wiley.

White, H. L. 2001. *Asymptotic Theory for Econometricians*. New York: Academic Press.

White, H. L., T. H. Kim, and S. Manganelli. 2015. "VAR for VaR: Measuring Tail Dependence Using Multivariate Regression Quantiles." *Journal of Econometrics* 187, no. 1: 169–188.

Wu, Y., and Y. Liu. 2009. "Variable Selection in Quantile Regression." *Statistica Sinica* 19: 801–817.

Xiao, Z., and R. Koenker. 2009. "Conditional Quantile Estimation for Generalized Autoregressive Conditional Heteroscedasticity Models." *Journal of the American Statistical Association* 104, no. 488: 1696–1712.

Zou, H. 2006. "The Adaptive Lasso and Its Oracle Properties." *Journal of the American Statistical Association* 101, no. 476: 1418–1429.

Appendix A

Mathematical Proofs

Note that the detailed proofs of main theorems and the necessary lemmas are presented in this appendix.

A: Proofs of Main Theorems

Proof of Theorem 1. For the consistency of $\widetilde{\theta}_\tau$, we have the following lemma with its proof given in the next subsection. □

Lemma 1. *(Consistency) Given Assumptions A1-A7, one has* $\widetilde{\theta}_\tau \xrightarrow{\mathcal{P}} \theta_\tau$, *where* $\widetilde{\theta}_\tau$ *is given by*

$$\widetilde{\theta}_\tau = \arg\min_{\theta\in\Theta} \frac{1}{T}\sum_{t=m}^{T} E\{Q_\tau(Y_t - \widehat{\mathbf{Z}}_{t,\tau}^\top\theta_\tau)|\mathcal{F}_t\}$$

where $\xrightarrow{\mathcal{P}}$ *denotes the convergence in probability.*

Further, by the Bahadur representation of quantile regression, it is easy to derive that

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

where $\widehat{\mathbf{D}}_T \equiv E\left(\frac{1}{T}\sum_{t=m}^{T}h_t(\widehat{\mathbf{Z}}_{t,\tau}^\top\theta_\tau - \mathbf{Z}_{t,\tau}^\top\theta_\tau|\mathcal{F}_t)\widehat{\mathbf{Z}}_{t,\tau}\widehat{\mathbf{Z}}_{t,\tau}^\top\right)$, $\psi_\tau(u) = \tau - I(u \leq 0)$ with $I(\cdot)$ denoting the indicator function. First, we will show that $\widehat{\mathbf{D}}_T \to \mathbf{D}$ using the following lemma with its proof presented in the next subsection.

Lemma 2. *Suppose Assumptions A1-A7 hold. Then one has*

$$\frac{1}{T}\sum_{t=m}^{T}\widehat{\mathbf{Z}}_{t,\tau}\widehat{\mathbf{Z}}_{t,\tau}^\top \to \frac{1}{T}\sum_{t=m}^{T}\mathbf{Z}_{t,\tau}\mathbf{Z}_{t,\tau}^\top$$

With Assumption A2, Assumption A3 and Lemma A.2, it suffices to show that

$$\begin{aligned}
&\left\|\frac{1}{T}\sum_{t=m}^{T}h_t(\widehat{\mathbf{Z}}_{t,\tau}^\top\theta_\tau - \mathbf{Z}_{t,\tau}^\top\theta_\tau|\mathcal{F}_t)\widehat{\mathbf{Z}}_{t,\tau}\widehat{\mathbf{Z}}_{t,\tau}^\top - \frac{1}{T}\sum_{t=m}^{T}h_t(0|\mathcal{F}_t)\mathbf{Z}_{t,\tau}\mathbf{Z}_{t,\tau}^\top\right\| \\
&\leq \left\|\frac{1}{T}\sum_{t=m}^{T}h_t(\widehat{\mathbf{Z}}_{t,\tau}^\top\theta_\tau - \mathbf{Z}_{t,\tau}^\top\theta_\tau|\mathcal{F}_t)\widehat{\mathbf{Z}}_{t,\tau}\widehat{\mathbf{Z}}_{t,\tau}^\top \right. \\
&\quad \left. -\frac{1}{T}\sum_{t=m}^{T}h_t(\widehat{\mathbf{Z}}_{t,\tau}^\top\theta_\tau - \mathbf{Z}_{t,\tau}^\top\theta_\tau|\mathcal{F}_t)\mathbf{Z}_{t,\tau}\mathbf{Z}_{t,\tau}^\top\right\| \\
&\quad + \left\|\frac{1}{T}\sum_{t=m}^{T}h_t(\widehat{\mathbf{Z}}_{t,\tau}^\top\theta_\tau - \mathbf{Z}_{t,\tau}^\top\theta_\tau|\mathcal{F}_t)\mathbf{Z}_{t,\tau}\mathbf{Z}_{t,\tau}^\top - \frac{1}{T}\sum_{t=m}^{T}h_t(0|\mathcal{F}_t)\mathbf{Z}_{t,\tau}\mathbf{Z}_{t,\tau}^\top\right\| \\
&\leq N\left\|\frac{1}{T}\sum_{t=m}^{T}\widehat{\mathbf{Z}}_{t,\tau}\widehat{\mathbf{Z}}_{t,\tau}^\top - \frac{1}{T}\sum_{t=m}^{T}\mathbf{Z}_{t,\tau}\mathbf{Z}_{t,\tau}^\top\right\| \\
&\quad + \left\|\frac{1}{T}\sum_{t=m}^{T}\left[h_t(\widehat{\mathbf{Z}}_{t,\tau}^\top\theta_\tau - \mathbf{Z}_{t,\tau}^\top\theta_\tau|\mathcal{F}_t) - h_t(0|\mathcal{F}_t)\right]\mathbf{Z}_{t,\tau}\mathbf{Z}_{t,\tau}^\top\right\| = o_p(1)
\end{aligned}$$

in which the second term converges to 0 due to

$$\begin{aligned}
h_t(\widehat{\mathbf{Z}}_{t,\tau}^\top\theta_\tau - \mathbf{Z}_{t,\tau}^\top\theta_\tau|\mathcal{F}_t) - h_t(0|\mathcal{F}_t) &\leq L\left|\widehat{\mathbf{Z}}_{t,\tau}^\top\theta_\tau - \mathbf{Z}_{t,\tau}^\top\theta_\tau\right| \\
&\leq L\left(\|\theta_\tau\| \times \left\|\widehat{\mathbf{Z}}_{t,\tau} - \mathbf{Z}_{t,\tau}\right\|\right) = o_p(1)
\end{aligned}$$

For the asymptotic properties of $\frac{1}{\sqrt{T}}\sum_{t=m}^{T}\psi_\tau(Y_t - \widehat{\mathbf{Z}}_{t,\tau}^\top\theta_\tau)\widehat{\mathbf{Z}}_{t,\tau}$, we consider the following decomposition:

$$\begin{aligned}
&\frac{1}{\sqrt{T}}\sum_{t=m}^{T}\psi_\tau(Y_t - \widehat{\mathbf{Z}}_{t,\tau}^\top\theta_\tau)\widehat{\mathbf{Z}}_{t,\tau} \\
&= \frac{1}{\sqrt{T}}\sum_{t=m}^{T}\psi_\tau(Y_t - \mathbf{Z}_{t,\tau}^\top\theta_\tau)\mathbf{Z}_{t,\tau} + \frac{1}{\sqrt{T}}\sum_{t=m}^{T}\psi_\tau(Y_t - \mathbf{Z}_{t,\tau}^\top\theta_\tau)(\widehat{\mathbf{Z}}_{t,\tau} - \mathbf{Z}_{t,\tau}) \\
&\quad - \frac{1}{\sqrt{T}}\sum_{t=m}^{T}(I(\varepsilon_{t,\tau} < \widehat{\mathbf{Z}}_{t,\tau}^\top\theta_\tau - \mathbf{Z}_{t,\tau}^\top\theta_\tau) - I(\varepsilon_{t,\tau} < 0))(\widehat{\mathbf{Z}}_{t,\tau} - \mathbf{Z}_{t,\tau}) \\
&\quad - \frac{1}{\sqrt{T}}\sum_{t=m}^{T}(I(\varepsilon_{t,\tau} < \widehat{\mathbf{Z}}_{t,\tau}^\top\theta_\tau - \mathbf{Z}_{t,\tau}^\top\theta_\tau) - I(\varepsilon_{t,\tau} < 0))\mathbf{Z}_{t,\tau} \qquad \text{(A1)}
\end{aligned}$$

where $\varepsilon_{t,\tau} = Y_t - q_{t,\tau}(\theta_\tau)$. Then, we will show that the second and third terms on the right hand of (A1) are $o_p(1)$ in the following lemma with its proof relegated to the next subsection.

Lemma 3. *Suppose Assumptions A1-A7 hold. Then one has*

(i) $\frac{1}{\sqrt{T}}\sum_{t=m}^{T}(\widehat{\mathbf{Z}}_{t,\tau} - \mathbf{Z}_{t,\tau})(\widehat{\mathbf{Z}}_{t,\tau} - \mathbf{Z}_{t,\tau})^\top = o_p(1)$;

(ii) $\frac{1}{\sqrt{T}}\sum_{t=m}^{T}\psi_\tau(Y_t-\mathbf{Z}_{t,\tau}^\top\theta_\tau)(\widehat{\mathbf{Z}}_{t,\tau}-\mathbf{Z}_{t,\tau})=o_p(1)$;

(iii) $\frac{1}{\sqrt{T}}\sum_{t=m}^{T}(I(\varepsilon_{t,\tau}<\widehat{\mathbf{Z}}_{t,\tau}^\top\theta_\tau-\mathbf{Z}_{t,\tau}^\top\theta_\tau)-I(\varepsilon_{t,\tau}<0))(\widehat{\mathbf{Z}}_{t,\tau}-\mathbf{Z}_{t,\tau})=o_p(1)$.

As for the last term in the right hand side of (A1), one has

$$\begin{aligned}
&\frac{1}{\sqrt{T}}\sum_{t=m}^{T}(I(\varepsilon_{t,\tau}<\widehat{\mathbf{Z}}_{t,\tau}^\top\theta_\tau-\mathbf{Z}_{t,\tau}^\top\theta_\tau)-I(\varepsilon_{t,\tau}<0))\mathbf{Z}_{t,\tau}\\
&=\sqrt{T}E\left[\frac{1}{T}\sum_{t=m}^{T}E(I(\varepsilon_{t,\tau}<\widehat{\mathbf{Z}}_{t,\tau}^\top\theta_\tau-\mathbf{Z}_{t,\tau}^\top\theta_\tau)-I(\varepsilon_{t,\tau}<0)|\mathcal{F}_t)\mathbf{Z}_{t,\tau}\right]\\
&\times(1+o(1))\\
&=\sqrt{T}E\left[\frac{1}{T}\sum_{t=m}^{T}h_t(0|\mathcal{F}_t)\mathbf{Z}_{t,\tau}(\widehat{\mathbf{Z}}_{t,\tau}-\mathbf{Z}_{t,\tau})^\top\theta_\tau\right](1+o(1))\\
&=\sqrt{T}E\left[h_t(0|\mathcal{F}_t)\mathbf{Z}_{t,\tau}\sum_{i=1}^{q}(\widehat{q}_{t-i,\tau}-q_{t-i,\tau})\beta_i\right](1+o(1))\\
&=E\left[h_t(0|\mathcal{F}_t)\mathbf{Z}_{t,\tau}\sum_{i=1}^{q}\overline{\mathbf{M}}_{t-i}^\top\beta_i\right]\sqrt{T}(\widehat{\mathbf{a}}_\tau-\mathbf{a}_\tau)(1+o(1))\\
&=\left(\sum_{i=1}^{q}E(h_t(0|\mathcal{F}_t)\mathbf{Z}_{t,\tau}\overline{\mathbf{M}}_{t-i}^\top\beta_i)\right)\sqrt{T}(\widehat{\mathbf{a}}_\tau-\mathbf{a}_\tau)(1+o(1))\\
&=\Gamma\sqrt{T}(\widehat{\mathbf{a}}_\tau-\mathbf{a}_\tau)+o(1)
\end{aligned}$$

which completes the proof of the first part of the theorem. As for the second part, we know that the Bahadur representation for the estimator $\widehat{\mathbf{a}}_\tau$ in the first step is as follows

$$\sqrt{T}(\widehat{\mathbf{a}}_\tau-\mathbf{a}_\tau)=\mathbf{B}^{-1}\left(\frac{1}{\sqrt{T}}\psi_\tau(\varepsilon_{t,\tau})\overline{\mathbf{M}}_t\right)+o_p(1)$$

Then, the asymptotic representation of $\widetilde{\theta}_\tau$ can be reformulated as

$$\begin{aligned}\sqrt{T}(\widetilde{\theta}_\tau-\theta_\tau)&=\mathbf{D}^{-1}\left(\frac{1}{\sqrt{T}}\sum_{t=m}^{T}\psi_\tau(\varepsilon_{t,\tau})\left(\mathbf{Z}_{t,\tau}-\Gamma\mathbf{B}^{-1}\overline{\mathbf{M}}_t\right)\right)\\
&+o_p(1)=\mathbf{D}^{-1}\mathbf{S}_T+o_p(1)\end{aligned}$$

$\mathbf{S}_T=\frac{1}{\sqrt{T}}\sum_{t=m}^{T}\psi_\tau(\varepsilon_{t,\tau})\left(\mathbf{Z}_{t,\tau}-\Gamma\mathbf{B}^{-1}\overline{\mathbf{M}}_t\right)$. We know $\{\psi_\tau(\varepsilon_{t,\tau})\left(\mathbf{Z}_{t,\tau}-\Gamma\mathbf{B}^{-1}\overline{\mathbf{M}}_t\right)\}$ is α-mixing of size $-r/(r-1)$, $r>1$, for all t, $1\le t\le T$, and by Assumption A6, it satisfies that $E(\|\mathbf{Z}_{t,\tau}-\Gamma\mathbf{B}^{-1}\overline{\mathbf{M}}_t\|^{2r+\epsilon})<\infty$, for some $\epsilon>0$. At the same time, one has $E[\psi_\tau(\varepsilon_{t,\tau})\left(\mathbf{Z}_{t,\tau}-\Gamma\mathbf{B}^{-1}\overline{\mathbf{M}}_t\right)]=E[\left(\mathbf{Z}_{t,\tau}-\Gamma\mathbf{B}^{-1}\overline{\mathbf{M}}_t\right)E(\psi_\tau(\varepsilon_{t,\tau})|\mathcal{F}_t)]=0$ for the first order condition. Together with Assumption A5, the central limit theory for sequences of dependent heterogeneously distributed random variables (see Theorem 5.20 in White (2001).) can be applied, and the second part of the theorem is proved.

Proof of Theorem 2. Let $\widehat{\theta}_\tau=\theta_\tau+\mathbf{u}/\sqrt{T}$,

$$G_T(\theta_\tau)=\frac{1}{T}\sum_{t=m}^{T}\mathrm{Q}_\tau(Y_t-\widehat{\mathbf{Z}}_{t,\tau}^\top\theta_\tau)+\lambda_T\sum_{i=1}^{p+q+1}\widehat{\omega}_i|\theta_{i,\tau}|,$$

and $\widehat{\mathbf{u}}=\operatorname{argmin}G_T(\theta_\tau+\mathbf{u}/\sqrt{T})$. Then $\widehat{\mathbf{u}}=\sqrt{T}(\widehat{\theta}_\tau-\theta_\tau)$. Denote $V_T(\mathbf{u})=G_T(\theta_\tau+\mathbf{u}/\sqrt{T})-G_T(\theta_\tau)$. It is easy to derive that

$$\begin{aligned}V_T(\mathbf{u})&=\mathbf{u}^\top\widehat{\mathbf{D}}_T\mathbf{u}-2\frac{\psi_\tau(\widehat{\varepsilon}_\tau)^\top\widehat{\mathbf{Z}}_\tau}{\sqrt{T}}\mathbf{u}\\
&+\frac{\lambda_T}{\sqrt{T}}\sum_{i=1}^{p+q+1}\widehat{\omega}_i\sqrt{T}\left(\left|\theta_{i,\tau}+\frac{u_i}{\sqrt{T}}\right|-|\theta_{i,\tau}|\right)\end{aligned}$$

where $\widehat{\varepsilon}_\tau=(\widehat{\varepsilon}_{m,\tau},\ldots,\widehat{\varepsilon}_{T,\tau})^\top$ with $\widehat{\varepsilon}_{t,\tau}=Y_t-\widehat{\mathbf{Z}}_{t,\tau}^\top\theta_\tau$, and u_i denotes the i-th element of $\mathbf{u}$. By Theorem 1 and following the proof of the Theorem 3 in Wu and Liu (2009), one has $V_T(\mathbf{u})\xrightarrow{\mathcal{L}}V(\mathbf{u})$ for every $\mathbf{u}$, where

$$V(\mathbf{u})=\begin{cases}\mathbf{u}_{\mathcal{A}}^\top\mathbf{D}_{\mathcal{A}}\mathbf{u}_{\mathcal{A}}-2\mathbf{u}_{\mathcal{A}}^\top\mathbf{S}_{\mathcal{A}}, & \text{if } u_i=0 \text{ for } \forall i\notin\mathcal{A},\\ \infty, & \text{otherwise,}\end{cases}$$

where $\mathbf{D}_{\mathcal{A}}$ is the sub-matrix of $\mathbf{D}$ whose rows and columns are chosen from $\mathbf{D}$ according to the row and column indexes set $\mathcal{A}$, and $\mathbf{S}_{\mathcal{A}}$ is the sub-vector of $\mathbf{S}_T$, whose elements are chosen from $\mathbf{S}_T$ with index set $\mathcal{A}$. Since $V(\mathbf{u})$ is convex in $\mathbf{u}$ and has a unique minimum, the epi-convergence results of Geyer (1994) imply that

$$\widehat{\mathbf{u}}_{\mathcal{A}}\xrightarrow{\mathcal{L}}\mathbf{D}_{\mathcal{A}}^{-1}\mathbf{S}_{\mathcal{A}}\quad\text{and}\quad\widehat{\mathbf{u}}_{\mathcal{A}^c}\xrightarrow{\mathcal{L}}0$$

which establishes the asymptotic normality part.

Next, we show the consistency property of model selection. For any $\widehat{\theta}_{\mathcal{A}}-\theta_{\mathcal{A}}=o(T^{-1/2})$, $0<\|\widehat{\theta}_{\mathcal{A}_c}\|\le CT^{-1/2}$,

$$\begin{aligned}&G_T(((\widehat{\theta}_{\mathcal{A}})^\top,\mathbf{0}^\top)^\top)-G_T(((\widehat{\theta}_{\mathcal{A}})^\top,(\widehat{\theta}_{\mathcal{A}_c})^\top)^\top)\\
&=\left[G_T(((\widehat{\theta}_{\mathcal{A}})^\top,\mathbf{0}^\top)^\top)-G_T(((\theta_{\mathcal{A}}^*)^\top,\mathbf{0}^\top)^\top)\right]\\
&-\left[G_T(((\widehat{\theta}_{\mathcal{A}})^\top,(\widehat{\theta}_{\mathcal{A}_c})^\top)^\top)-G_T(((\theta_{\mathcal{A}}^*)^\top,\mathbf{0}^\top)^\top)\right]\\
&=U_T(\sqrt{T}((\widehat{\theta}_{\mathcal{A}}-\theta_{\mathcal{A}}^*)^\top,\mathbf{0}^\top)^\top)\\
&-U_T(\sqrt{T}((\widehat{\theta}_{\mathcal{A}}-\theta_{\mathcal{A}}^*)^\top,(\widehat{\theta}_{\mathcal{A}_c})^\top)^\top)-\frac{\lambda_T}{\sqrt{T}}\sum_{j\in\mathcal{A}_c}\widehat{\omega}_j|\widehat{\theta}_j|\end{aligned}$$

where $U_T(\mathbf{u})=1/T\sum_{t=m}^{T}\{\mathrm{Q}_\tau(Y_t-\widehat{\mathbf{Z}}_{t,\tau}^\top(\theta_\tau+\mathbf{u}/\sqrt{T}))-\mathrm{Q}_\tau(Y_t-\widehat{\mathbf{Z}}_{t,\tau}^\top\theta_\tau)\}$. The first two terms has been proved to be bounded in the proof of Lemma 1 in Wu and Liu (2009), and for the third term one has

$$\frac{\lambda_T}{\sqrt{T}}\sum_{j\in\mathcal{A}_c}\widehat{\omega}_j|\widehat{\theta}_j|=\left(T^{(\eta-1)/2}\lambda_T\right)\sum_{j\in\mathcal{A}_c}\left|(\sqrt{T}|\widehat{\theta}_j|)^{-\eta}\right||\widehat{\theta}_j|\to\infty$$

due to the condition that $T^{(\eta-1)/2}\lambda_T\to\infty$. Since the third term dominates the other two terms, then, $G_T(((\widehat{\theta}_{\mathcal{A}})^\top,\mathbf{0}^\top)^\top)-G_T(((\widehat{\theta}_{\mathcal{A}})^\top,(\widehat{\theta}_{\mathcal{A}_c})^\top)^\top)<0$ for large T. This proves the consistency part of Theorem 2. □

Appendix B

Proofs of Lemmas

This subsection contains the detailed proof of Lemmas 1–g195 3 presented in the previous section.

Proof of Lemma 1. For the proof of Lemma 1, the first part in the proof of the Theorem 3 in Xiao and Koenker (2009). can be easily applied. To save space, we skip its proof. □

Proof of Lemma 2. Write $\widehat{\mathbf{Z}}_\tau=(\mathbf{X},\widehat{\mathbf{Q}}_\tau)$, where

$$\mathbf{X}=\begin{pmatrix}X_{m,1}&\cdots&X_{m,p}\\ \vdots&\ddots&\vdots\\ X_{T,1}&\cdots&X_{T,p}\end{pmatrix},\quad\widehat{\mathbf{Q}}_\tau=\begin{pmatrix}\widehat{q}_{m-1,\tau}&\cdots&\widehat{q}_{m-q,\tau}\\ \vdots&\ddots&\vdots\\ \widehat{q}_{T-1,\tau}&\cdots&\widehat{q}_{T-q,\tau}\end{pmatrix}$$

Then,

$$\frac{1}{T}\widehat{\mathbf{Z}}_\tau\widehat{\mathbf{Z}}_\tau^\top=\begin{pmatrix}\frac{1}{T}\mathbf{X}^\top\mathbf{X}&\frac{1}{T}\mathbf{X}^\top\widehat{\mathbf{Q}}_\tau\\ \frac{1}{T}\widehat{\mathbf{Q}}_\tau^\top\mathbf{X}&\frac{1}{T}\widehat{\mathbf{Q}}_\tau^\top\widehat{\mathbf{Q}}_\tau\end{pmatrix}$$

For a typical entry of $\frac{1}{T}\mathbf{X}^\top\hat{\mathbf{Q}}_\tau$ or $\frac{1}{T}\hat{\mathbf{Q}}_\tau^\top\mathbf{X}$, one has

$$\begin{aligned}\frac{1}{T}\sum_{t=m}^{T}X_{t,j}\hat{q}_{t-k,\tau} &= \frac{1}{T}\sum_{t=m}^{T}X_{t,j}q_{t-k,\tau}+\frac{1}{T}\sum_{t=m}^{T}X_{t,j}(\hat{q}_{t-k,\tau}-q_{t-k,\tau})\\ &= E(X_{t,j}q_{t-k,\tau})+\frac{1}{T}\sum_{t=m}^{T}X_{t,j}\overline{\mathbf{M}}_{t-k}^\top(\hat{\mathbf{a}}_\tau-\mathbf{a}_\tau)+o(1)\\ &= E(X_{t,j}q_{t-k,\tau})+o(1)\end{aligned}$$

due to Assumption A4 and Assumption A6, for any $j=1,\ldots,p$ and $k=1,\ldots,q$. Now, consider $\frac{1}{T}\hat{\mathbf{Q}}_\tau^\top\hat{\mathbf{Q}}_\tau$, a typical entry of which being is equal to

$$\begin{aligned}\frac{1}{T}\sum_{t=m}^{T}\hat{q}_{t-j,\tau}\hat{q}_{t-k,\tau} &= \sum_{i=0}^{d}\hat{a}_i\left(\frac{1}{T}\sum_{t=m}^{T}\hat{q}_{t-j,\tau}\overline{M}_{t-k,i}\right)\\ &= \sum_{i=0}^{d}(a_i+o(T^{-1/2}))\{E(\overline{M}_{t-k,i}q_{t-j,\tau})+o(1)\}\\ &= E\left[\left(\sum_{i=0}^{d}a_i\overline{M}_{t-k,i}\right)q_{t-j,\tau}\right]+o(1)\\ &= E\left(q_{t-k,\tau}q_{t-j,\tau}\right)+o(1)\end{aligned}$$

by Assumption A6. This completes the proof. □

Proof of Lemma 3. For the proof of (i), it suffices to show that $\frac{1}{\sqrt{T}}\sum_{t=m}^{T}(\hat{q}_{t-i,\tau}-q_{t-i,\tau})(\hat{q}_{t-j,\tau}-q_{t-j,\tau})=o_p(1)$, for any $1\le i,j\le q$. With $\hat{q}_{t-i,\tau}-q_{t-i,\tau}=\overline{\mathbf{M}}_{t-i}^\top(\hat{\mathbf{a}}_\tau-\mathbf{a}_\tau)$, one has

$$\begin{aligned}&\frac{1}{\sqrt{T}}\sum_{t=m}^{T}(\hat{q}_{t-i,\tau}-q_{t-i,\tau})(\hat{q}_{t-j,\tau}-q_{t-j,\tau})\\ &= \sqrt{T}(\hat{\mathbf{a}}_\tau-\mathbf{a}_\tau)^\top\left(\frac{1}{T^{3/2}}\sum_{t=m}^{T}\overline{\mathbf{M}}_{t-i}\overline{\mathbf{M}}_{t-j}^\top\right)\sqrt{T}(\hat{\mathbf{a}}_\tau-\mathbf{a}_\tau)=o_p(1)\end{aligned}$$

due to Assumption A4. For (ii), it is easy to prove that the expectation of this term is 0, and the variance of this term is

$$\begin{aligned}&\mathrm{Var}\left(\frac{1}{\sqrt{T}}\sum_{t=m}^{T}\psi_\tau(Y_t-\mathbf{Z}_{t,\tau}^\top\theta_\tau)(\hat{\mathbf{Z}}_{t,\tau}-\mathbf{Z}_{t,\tau})\Big|\mathcal{F}_t\right)\\ &= \frac{1}{T}\sum_{t=m}^{T}\sum_{s=m}^{T}E\Big(\psi_\tau(Y_t-\mathbf{Z}_{t,\tau}^\top\theta_\tau)\psi_\tau(Y_s-\mathbf{Z}_{s,\tau}^\top\theta_\tau)\\ &\quad\times(\hat{\mathbf{Z}}_{t,\tau}-\mathbf{Z}_{t,\tau})(\hat{\mathbf{Z}}_{s,\tau}-\mathbf{Z}_{s,\tau})^\top\Big|\mathcal{F}_t\Big)\\ &= \frac{2\tau(1-\tau)}{T}\sum_{t=m}^{T}(\hat{\mathbf{Z}}_{t,\tau}-\mathbf{Z}_{t,\tau})(\hat{\mathbf{Z}}_{t,\tau}-\mathbf{Z}_{t,\tau})^\top=o_p(1)\end{aligned}$$

by applying (i). For (iii), one has

$$\begin{aligned}&E\left(\frac{1}{\sqrt{T}}\sum_{t=m}^{T}(I(\varepsilon_{t,\tau}<\hat{\mathbf{Z}}_{t,\tau}^\top\theta_\tau-\mathbf{Z}_{t,\tau}^\top\theta_\tau)-I(\varepsilon_{t,\tau}<0))(\hat{\mathbf{Z}}_{t,\tau}-\mathbf{Z}_{t,\tau})\Big|\mathcal{F}_t\right)\\ &= \frac{1}{\sqrt{T}}\sum_{t=m}^{T}(\hat{\mathbf{Z}}_{t,\tau}-\mathbf{Z}_{t,\tau})E\left(I(\varepsilon_{t,\tau}<\hat{\mathbf{Z}}_{t,\tau}^\top\theta_\tau-\mathbf{Z}_{t,\tau}^\top\theta_\tau)-I(\varepsilon_{t,\tau}<0)\Big|\mathcal{F}_t\right)\\ &= \frac{1}{\sqrt{T}}\sum_{t=m}^{T}h_t(0|\mathcal{F}_t)(\hat{\mathbf{Z}}_{t,\tau}-\mathbf{Z}_{t,\tau})(\hat{\mathbf{Z}}_{t,\tau}-\mathbf{Z}_{t,\tau})^\top\theta_\tau\\ &\le L\left(\frac{1}{\sqrt{T}}\sum_{t=m}^{T}(\hat{\mathbf{Z}}_{t,\tau}-\mathbf{Z}_{t,\tau})(\hat{\mathbf{Z}}_{t,\tau}-\mathbf{Z}_{t,\tau})^\top\right)\theta_\tau=o_p(1)\end{aligned}$$

where ε lies between 0 and $\hat{\mathbf{Z}}_{t,\tau}^\top\theta_\tau-\mathbf{Z}_{t,\tau}^\top\theta_\tau$, and

$$\begin{aligned}&\mathrm{Var}\left(\frac{1}{\sqrt{T}}\sum_{t=m}^{T}(I(\varepsilon_{t,\tau}<\hat{\mathbf{Z}}_{t,\tau}^\top\theta_\tau-\mathbf{Z}_{t,\tau}^\top\theta_\tau)-I(\varepsilon_{t,\tau}<0))\mathbf{Z}_{t,\tau}\Big|\mathcal{F}_t\right)\\ &= \frac{1}{T}\sum_{t=m}^{T}\sum_{s=m}^{T}(\hat{\mathbf{Z}}_{t,\tau}-\mathbf{Z}_{t,\tau})(\hat{\mathbf{Z}}_{s,\tau}-\mathbf{Z}_{s,\tau})^\top\\ &\quad\times\Big\{E\Big[\Big(I(\varepsilon_{s,\tau}<\hat{\mathbf{Z}}_{s,\tau}^\top\theta_\tau-\mathbf{Z}_{s,\tau}^\top\theta_\tau)\\ &\quad-I(\varepsilon_{s,\tau}<0)\Big)\Big(I(\varepsilon_{t,\tau}<\hat{\mathbf{Z}}_{t,\tau}^\top\theta_\tau-\mathbf{Z}_{t,\tau}^\top\theta_\tau)-I(\varepsilon_{t,\tau}<0)\Big)\Big|\mathcal{F}_t\Big]\\ &\quad-E\Big(I(\varepsilon_{s,\tau}<\hat{\mathbf{Z}}_{s,\tau}^\top\theta_\tau-\mathbf{Z}_{s,\tau}^\top\theta_\tau)-I(\varepsilon_{s,\tau}<0)\Big|\mathcal{F}_t\Big)\\ &\quad\times E\Big(I(\varepsilon_{t,\tau}<\hat{\mathbf{Z}}_{t,\tau}^\top\theta_\tau-\mathbf{Z}_{t,\tau}^\top\theta_\tau)-I(\varepsilon_{t,\tau}<0)\Big|\mathcal{F}_t\Big)\Big\}\\ &= \frac{2}{T}\sum_{t=m}^{T}(\hat{\mathbf{Z}}_{t,\tau}-\mathbf{Z}_{t,\tau})(\hat{\mathbf{Z}}_{t,\tau}-\mathbf{Z}_{t,\tau})^\top\\ &\quad\times\Big\{\mathrm{Var}\Big[\Big(I(\varepsilon_{t,\tau}<\hat{\mathbf{Z}}_{t,\tau}^\top\theta_\tau-\mathbf{Z}_{t,\tau}^\top\theta_\tau)-I(\varepsilon_{t,\tau}<0)\Big)\Big|\mathcal{F}_t\Big]\Big\}\\ &\le \frac{C}{T}\sum_{t=m}^{T}(\hat{\mathbf{Z}}_{t,\tau}-\mathbf{Z}_{t,\tau})(\hat{\mathbf{Z}}_{t,\tau}-\mathbf{Z}_{t,\tau})^\top\\ &\quad\times\Big[Pr\Big(\varepsilon_{t,\tau}<\Big|\hat{\mathbf{Z}}_{t,\tau}^\top\theta_\tau-\mathbf{Z}_{t,\tau}^\top\theta_\tau\Big|\Big|\mathcal{F}_t\Big)\\ &\quad-Pr\Big(\varepsilon_{t,\tau}<-\Big|\hat{\mathbf{Z}}_{t,\tau}^\top\theta_\tau-\mathbf{Z}_{t,\tau}^\top\theta_\tau\Big|\Big|\mathcal{F}_t\Big)\Big]\\ &= o_p(1)\end{aligned}$$

by applying (i). □