Journal of Financial Econometrics, 2025, **23(2)**, nbaf010
https://doi.org/10.1093/jjfinec/nbaf010
Article

OXFORD

Assessing Tail Risk via a Generalized Conditional Autoregressive Expectile Model

Zongwu Cai [1], Ying Fang [2,3], and Dingshi Tian [4]

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

Address correspondence to Dingshi Tian, Department of Statistics, School of Statistics and Mathematics, Zhongnan University of Economics and Law, Wuhan, Hubei 430073, China, or e-mail: dstian@zuel.edu.cn

Abstract

This article proposes a generalized conditional autoregressive expectile model, including autoregressive components in assessing tail risk, which can be treated as an infinite version of the conditional autoregressive expectile model proposed by Kuan, Yeh, and Hsu (2009) and can be implemented as a vehicle for estimating the CAViaR model proposed in Engle and Manganelli (2004) and studied by Xiao and Koenker (2009). Due to the unobservable latent components in the proposed model, the quasi-maximum likelihood estimation method is suggested and a HAC covariance matrix estimator is proposed. Furthermore, a dynamic expectile test is proposed for both in-sample model adequacy evaluation and out-of-sample forecasting for comparison purposes. Finally, Monte Carlo simulations and a real example are conducted to illustrate that the proposed methodology is practically useful. Our empirical study demonstrates that the tail risk characterized by the proposed model achieves a better performance in the period of the Covid-19 epidemic.

Keywords: conditional autoregressive expectile model, COVID-19 pandemic, dynamic testing, expectile modeling, quasi-maximum likelihood estimation, tail risk

JEL classifications: C32, C51, C58, G17

Assessing tail risk is one of most important tasks in financial risk management. Due to the recent outbreak of the Covid-19 epidemic, how to capture downside risk exposures is a big demanding task not only for financial institutions and regulators but also for academic research. The commonly adopted risk measure Value-at-Risk (VaR), the quantile of a portfolio loss distribution, which evaluates the potential maximum loss for a specified probability level, has been criticized to provide a too conservative risk measure in the case of the occurrence of catastrophic events, because a quantile based risk measure focuses on the probability of the occurrence of extreme losses but by nature, is insensitive to the size of extreme

Received: October 03, 2023. **Revised:** January 11, 2025. **Editorial decision:** March 02, 2025. **Accepted:** March 06, 2025

losses. By rectifying the undesirable property of VaR, an alternative risk measure expectile has recently received increased attentions in the literature of financial risk management.

Expectile, first introduced by Newey and Powell (1987), is a risk measure based on the minimization of asymmetrically weighted mean square errors. Contrary to a quantile based risk measure using an absolute loss function in the check function, expectile is sensitive to the magnitude of extreme losses owing to the adoption of a quadratic loss function. In addition, expectile has more desirable properties among alternative risk measures. For instance, the expectile is a coherent risk measure as in Artzner et al. (1999), satisfying properties of translation invariance, sub-additivity, positive homogeneity and monotonicity, while the VaR is lack of the sub-additivity property, which violates the principle of diversification in risk management. Compared to another coherent risk measure expected shortfall (ES), Gneiting (2011) showed that the ES does not enjoy the desirable property of elicitability, which is proven to be correlated to backtesting as in Embrechts and Hofert (2014) and Ziegel (2016). The reader is referred to the recent survey paper by Tian, Cai, and Fang (2019) for more details about the properties of these risk measures.

In response to the aforementioned desirable properties of expectiles, an increased number of conditional expectile models have been developed in the literature in recent years. For example, Kuan, Yeh, and Hsu (2009) proposed a class of conditional autoregressive expectile (CARE) models, in which positive and negative lagged returns are allowed to capture asymmetric dynamic effects on tail expectiles. Furthermore, using one-to-one relations among quantile, expected shortfall and expectile, Taylor (2008) considered using CARE models to estimate both VaR and ES. By assuming that the error term follows an asymmetric normal distribution, Gerlach and Chen (2016) and Gerlach, Walpole, and Wang (2017) proposed the CARE-R and CARE-X models by incorporating realized ranges and realized volatility into tail risk forecasting, respectively. To better capturing time-varying features of conditional expectiles, Xu, Mihoci, and Härdle (2018) utilized a local parametric approach to study the parameter instability of tail risk dynamics. Using nonparametric estimation, Xie, Zhou, and Wan (2014) introduced a varying coefficient expectile model, while Cai, Fang, and Tian (2018) proposed a partially varying coefficient expectile model, in which some coefficients are allowed to be constant but others are allowed to vary with other random variables. More recently, Jiang, Hu, and Yu (2022) proposed a single index expectile model to estimate both conditional VaR and ES. Further, Daouia, Girard, and Stupfler (2018, 2020), Daouia, Gijbels, and Stupfler (2019), and Xu, Hou, and Li (2022) investigated the expectile models in the extreme tails. Other related literature on expectile models includes, but not limited to, Bellini and Di Bernardino (2017), Zhang and Li (2017), Mohammedi, Bouzebda, and Laksaci (2021) and the references therein.

In this article, we propose a generalized conditional dynamic expectile model by specifying autoregressive components of lagged expectiles in the model, termed as a generalized conditional autoregressive expectile (GCARE) model. Compared to the model in Taylor (2008) for allowing one autoregressive expectile in the CARE model, our GCARE model is general enough to handle more autoregressive components of expectile and lags of covariates. Following a similar idea of the conditional autoregressive Value-at-Risk (CAViaR) model by Engle and Manganelli (2004), we expect the addition of lagged expectiles to not only better capture the dynamic effects in return distributions but also provide a parsimonious model since the GCARE model can be transformed into a CARE model with infinite dimensions. Clearly, the GCARE model not only allows for dynamic effects of lagged returns on conditional expectiles as the CARE model does, but also allows for the evolution of expectiles over time.

However, the inclusion of the autoregressive expectiles brings new challenge to the model estimation. The estimation approach used in the CARE models cannot be directly employed since the expectile autoregressive components are not observable at time t.

Instead, we propose a quasi-maximum likelihood method to estimate the GCARE model using an iterative estimation procedure. Furthermore, model specification tests for conditional autoregressive expectile models are still less developed. To the best of our knowledge, the paper by Kuan, Yeh, and Hsu (2009) is the only work in the literature to consider an encompassing test to judge model adequacy between two given non-nested expectile models. However, this test cannot be directly applied, when lagged expectiles are included. To fill this gap, we propose dynamic expectile (DE) tests, which can serve as a valuable addition to the traditional toolkit in this area. To be specific, the in-sample test can be implemented as a model specification test for model selection purposes, and the out-of-sample test can be used to compare forecasting performances between competing expectile models. Finally, we apply our GCARE model to a portfolio insurance strategy for the S&P 500 index from March 2018 to December 2022, which includes the period of the Covid-19 epidemic. To guarantee a minimum portfolio value, which is called a floor, at the end of a given time horizon, the strategy needs to determine a proportion of the total portfolio value invested in risky assets, which is called a multiple, and the remaining part in risk-less assets. We apply the CAViaR model, the variants of the CARE models and our GCARE models to estimate the multiple, which represents the leverage of the portfolio risk exposure. The tail risk characterized by the GCARE model achieves a better performance, particularly in the period of the Covid-19 epidemic and the detailed analysis is given in Section 5.

The main contributions of this paper are outlined as follows. First, we propose the GCARE model, which offers enhanced flexibility to accommodate more autoregressive components and lags of covariates. Notably, Taylor's (2008) CARE models are a special case of GCARE(p, q), when both p and q equal 1. Moreover, we also provide an EBIC to facilitate the selection of the parameters p and q within the GCARE models. Furthermore, the asymptotic theory of the GCARE model is derived, which permits the constructing of model specification test as well as the prediction evaluation method. Also, the in-sample DE test and out-of-sample DE test are proposed. The in-sample test serves as a model specification test for model selection purposes, while the out-of-sample test can be utilized to compare forecasting performances between competing expectile models. Finally, empirical studies have demonstrated that GCARE(p, q) models, particularly for $p, q > 1$, exhibit superior performance. For example, in Section 4.2, we have shown that second order expectile lag component is highly significant and that GCARE(p, q) models outperform GCARE (1,1) model for both in-sample DE test and out-of-sample DE test considering the p-value of the test statistics. In Section 4.3, our findings indicate that GCARE(p, q) models outperform alternative models based on the returns and Sharpe ratio of time invariant portfolio protection (TIPP) based portfolios.

The rest of this article is organized as follows: Section 1 introduces the GCARE model and its estimation procedure. Section 2 investigates the asymptotic properties of the proposed QMLEs, the heteroskedasticity and autocorrelation consistent (HAC) covariance matrix, and dynamic expectile tests. Monte Carlo experiments and empirical analysis results of a real data example are reported in Sections 3 and 4, respectively. Section 5 concludes the article. All technical proofs are deferred to Appendix with the detailed proofs of some lemmas given in the Supplementary material.

1 GCARE Model

1.1 Model Setup

Assume that $\{Y_t, \mathbf{X}_t\}_{t\in\mathbb{Z}}$ is a sequence of strictly stationary random variables, where Y_t is a scalar variable of interest and $\mathbf{X}_t$ is a k-dimensional vector of covariates, which is allowed to include lagged values of Y_t, $\mathbb{Z}$ denotes the set of integers. Let $\mathcal{F}_t = \sigma\{(Y_s, \mathbf{X}_s); s \le t-1\}$ be the information set (σ-field) available at time t, $H(\mathcal{F}_t)$ be the Hilbert space consisting of those random variables which are measurable with respect to $\mathcal{F}_t$, and $e_\tau(\mathbf{W}_t, \boldsymbol{\theta}_{0,\tau})$ denote the τth conditional expectile of Y_t given $\mathbf{W}_t$, where $\mathbf{W}_t \in H(\mathcal{F}_t)$ is a vector of random variables and $\boldsymbol{\theta}_{0,\tau}$ is an unknown parameter in the parameter space Θ. Note that $\mathbf{W}_t$ is allowed to include both covariates $\mathbf{X}_t$ and some lags of latent variable $e_{t,\tau}(\boldsymbol{\theta}_{0,\tau})$. The τth conditional expectile of Y_t given $\mathbf{W}_t$ is then defined as

$$e_{t,\tau}(\boldsymbol{\theta}_{0,\tau}) = e_\tau(\mathbf{W}_t, \boldsymbol{\theta}_{0,\tau}) = \arg\min_{u \in \mathcal{R}} E\{\mathrm{Q}_\tau(Y_t - u)|\mathbf{W}_t\},$$

where $\mathrm{Q}_\tau(u) = |\tau - \mathrm{I}(u \leq 0)|u^2$, which, as pointed out by Newey and Powell (1987), is continuously differentiable in u, and $\mathrm{I}(\cdot)$ denotes an indicator function. Note that the minimization problem is well defined for $E(|Y_t|) < \infty$. This paper considers a class of GCARE(p,q) models as follows:

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

where $\boldsymbol{\theta}_{0,\tau} = (\alpha_{0,\tau}, \boldsymbol{\alpha}_{1,\tau}^\top \ldots, \boldsymbol{\alpha}_{p,\tau}^\top, \beta_{1,\tau}, \ldots, \beta_{q,\tau})^\top \in \mathcal{R}^{kp+q+1}$ with $p \geq 0$, $q > 0$, $e_{t-j,\tau}$ is the jth lag of $e_{t,\tau}$, $A(L) = \sum_{i=1}^{p} \boldsymbol{\alpha}_{i,\tau}^\top L^i$, and $B(L) = \sum_{j=1}^{q} \beta_{j,\tau} L^j$ with L denoting a lag operator. For ease of notation, in what follows, τ is dropped from $\boldsymbol{\theta}_{0,\tau}$, $\boldsymbol{\alpha}_{i,\tau}$ for $0 \leq i \leq p$ and $\beta_{j,\tau}$ for $1 \leq j \leq q$.

Note that p in GCARE(p,q) refers to the order of lags of covariates $\mathbf{X}_t$, and q means the order of the expectile lags. If model (1) is stationary, then it can be rewritten as

$$e_{t,\tau}(\boldsymbol{\theta}_0) = \alpha_0(1 - B(1))^{-1} + A(L)(1 - B(L))^{-1}\mathbf{X}_t = \alpha_0 \left(1 - \sum_{j=1}^{q} \beta_j\right)^{-1} + \sum_{i=1}^{\infty} \boldsymbol{\gamma}_i^\top \mathbf{X}_{t-i},$$

which can be considered as an infinite-dimensional CARE model proposed by Kuan, Yeh, and Hsu (2009). Clearly, $\boldsymbol{\gamma}_i$ for $i \geq 1$ in the above equation is obtained from the power series expansion of $A(L)(1 - B(L))^{-1}$.

First, one can see that model (1) is also general enough to include many other existing expectile models as special cases. For instance, when we set $p = q = 1$ and $X_t = |Y_{t-1}|$ in model (1), it becomes a symmetric absolute value CARE model of Taylor (2008),

$$e_{t,\tau} = \alpha_0 + \alpha_1|Y_{t-1}| + \beta_1 e_{t-1,\tau}.$$

If $p = q = 1$ and $\mathbf{X}_t = (Y_{t-1}^+, Y_{t-1}^-)^\top$, model (1) is restricted to the asymmetric slope CARE model of Taylor (2008),

$$e_{t,\tau} = \alpha_0 + \alpha_{1,1} Y_{t-1}^+ + \alpha_{1,2} Y_{t-1}^- + \beta_1 e_{t-1,\tau},$$

where $v^+ = \max(v, 0)$ and $v^- = \max(-v, 0)$. Moreover, the CARE-R model as in Gerlach and Chen (2016) and the CARE-X model as in Gerlach, Walpole, and Wang (2017) are obtained simply by defining $\mathbf{X}_t$ to be the realized range or the realized measures of volatility in model (1), respectively. It is important to acknowledge that the aforementioned CARE models, including the CARE-R and CARE-X variants, extend beyond the scope of the model framework established by Kuan, Yeh, and Hsu (2009).

Next, we establish a relation between the linear GARCH model as in Taylor (1986) and the proposed GCARE model. Note that the linear GARCH(p, q) model is given by

$$Y_t = \sigma_t \varepsilon_t \quad \text{with} \quad \sigma_t = \alpha_0 + \sum_{i=1}^{p} \alpha_i |Y_{t-i}| + \sum_{j=1}^{q} \beta_j \sigma_{t-j}, \tag{2}$$

where ε_t is an i.i.d. sequence of random variables with zero mean and unit variance. Compared to the linear GARCH model, according to Taylor (1986) and Xiao and Koenker (2009), and the references therein, the standard GARCH model is too sensitive to extreme values so that the former has been widely adopted to model financial returns. Following the idea of Xiao and Koenker (2009), by multiplying both sides of Equation (2) by $e_\tau(\varepsilon_t)$, where $e_\tau(\varepsilon_t)$ is the τth expectile of ε_t, the τth conditional expectile of Y_t at time t, $e_{t,\tau}$, has the following representation:

$$e_{t,\tau} = \bar{\alpha}_{0,\tau} + \sum_{i=1}^{p} \bar{\alpha}_{i,\tau} |Y_{t-i}| + \sum_{j=1}^{q} \beta_j e_{t-j,\tau}, \tag{3}$$

where $\bar{\alpha}_{i,\tau} = \alpha_i e_\tau(\varepsilon_t)$ for $0 \le i \le p$. This is exactly the GCARE(p, q) model, which has been introduced in the subsection.

A byproduct of model (3) is an ES(p, q) model. As discussed by Newey and Powell (1987) and Taylor (2008), we can find the $q_{t,\eta}$ satisfying $e_{t,\tau} = q_{t,\eta}$ with $q_{t,\eta}$ denoting the conditional quantile of Y_t at a probability level η corresponding to τ, then a link between expectile and ES can be formulated in the following expression:

$$\text{ES}_{t,\eta} = \left(1 + \frac{\tau}{(1-2\tau)\eta}\right) e_{t,\tau} - \frac{\tau}{(1-2\tau)\eta} E(Y_t),$$

where $\text{ES}_{t,\eta}$ is the ηth conditional ES of Y_t. If $E(Y_t) = 0$, it can be simplified to

$$\text{ES}_{t,\eta} = \left(1 + \frac{\tau}{(1-2\tau)\eta}\right) e_{t,\tau},$$

which clearly provides an easy way to estimate the conditional ES via GCARE.

1.2 Estimation Procedure

Similar to estimating conditional quantile models as in Engle and Manganelli (2004) and White, Kim, and Manganelli (2015), the QMLE approach can be used to estimate model (1) by maximizing the following quasi-likelihood

$$\text{QL}_{T,\tau}(\boldsymbol{\theta}) \equiv T^{-1} \sum_{t=1}^{T} \ln ql\left(Y_t, e_{t,\tau}(\boldsymbol{\theta})\right),$$

where the conditional quasi-likelihood function at time t is

$$ql\left(Y_t, e_{t,\tau}(\boldsymbol{\theta})\right) = \frac{2}{\sigma} \left(\sqrt{\frac{\pi}{1-\tau}} + \sqrt{\frac{\pi}{\tau}} \right)^{-1} \exp\left\{ -|\tau - I\left(Y_t < e_{t,\tau}(\boldsymbol{\theta})\right)| \frac{\left(Y_t - e_{t,\tau}(\boldsymbol{\theta})\right)^2}{\sigma^2} \right\},$$

which is denoted as an asymmetric normal distribution first used in Gerlach and Chen (2016) and Gerlach, Walpole, and Wang (2017). The above maximization problem is equivalent to minimizing the following classical asymmetric least squares (ALS) loss function,

$$\mathrm{L}_{T,\tau}(\boldsymbol{\theta}) \equiv T^{-1}\sum_{t=1}^{T} w_{t,\tau}(\boldsymbol{\theta})\big(Y_t - e_{t,\tau}(\boldsymbol{\theta})\big)^2,$$

where $w_{t,\tau}(\boldsymbol{\theta}) = |\tau - \mathrm{I}\big(Y_t \le e_{t,\tau}(\boldsymbol{\theta})\big)|$.

As shown above, the ALS function is continuously differentiable, and the second derivative of the ALS function is almost surely continuous. To obtain QMLE, an iterative optimization procedure similar to that in Bollerslev (1986) is called. Define ∇ as the first-order partial derivative operators with respect to $\boldsymbol{\theta}$. Let $\boldsymbol{\theta}^{(i)}$ denote the estimated parameter after the ith iteration. Then, $\boldsymbol{\theta}^{(i+1)}$ is calculated as

$$\boldsymbol{\theta}^{(i+1)} = \boldsymbol{\theta}^{(i)} + \lambda_i\left(\frac{1}{T}\sum_{t=1}^{T} w_{t,\tau}\big(\boldsymbol{\theta}^{(i)}\big)\nabla e_{t,\tau}\big(\boldsymbol{\theta}^{(i)}\big)\nabla^{\top} e_{t,\tau}\big(\boldsymbol{\theta}^{(i)}\big)\right)^{-1}\nabla \mathrm{L}_{T,\tau}\big(\boldsymbol{\theta}^{(i)}\big),$$

where $\nabla \mathrm{L}_{T,\tau}\big(\boldsymbol{\theta}^{(i)}\big)$ and $\nabla e_{t,\tau}\big(\boldsymbol{\theta}^{(i)}\big)$ are evaluated at $\boldsymbol{\theta}^{(i)}$, and λ_i is a variable step length used to maximize the likelihood function in the given direction. Finally, $\hat{\boldsymbol{\theta}}_T$, the QMLE of the given GCARE(p, q) model, is obtained.

Remark 1. *As shown above, the estimation of GCARE models is similar to the implementation of GARCH models. Therefore, to ensure estimation accuracy, the main part of our computation codes are adjusted based on the source codes from the fGarch package in R. Specifically, the Nelder–Mead simplex algorithm and quasi-Newton method are used to identify the optimal parameter estimates. See, for example*, Wuertz et al. (2017) *for details.*

1.3 Model Selection

One question of implementing the GCARE models is how to choose the number of lags. For the selection of p and q, we adopt the following extended Bayesian information criterion (EBIC) based on the form of a penalized log-likelihood as in Chen and Chen (2008) for mean regression and Lee, Noh, and Park (2014) for quantile setting:

$$\mathrm{EBIC}(p,q) = \log\left(\sum_{t=1}^{T} Q_\tau\Big(Y_t - e_{t,\tau}\big(\hat{\boldsymbol{\theta}}_T\big)\Big)\right) + (kp+q+1)\frac{\log T}{2T}C_T,$$

where C_T diverges to infinity as T grows. Here, to ensure both the model complexity and computation efficiency, we set a maximum of $p=5$ and $q=2$ in our computations in Sections 3 and 4. The optimal model is selected with the minimal EBIC within the set of p and q.

2 Asymptotic Theory

In this section, the asymptotic properties for the proposed estimators $\hat{\boldsymbol{\theta}}_T$ are derived, and an HAC covariance matrix estimator and a dynamic expectile test are also proposed.

2.1 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.

Assumption A (Consistency):

A1. (Ω, F, P) is a complete probability space, and the process $\{Y_t, \mathbf{X}_t\}_{t\in\mathbb{Z}}$ is (strong) α-mixing with mixing coefficient $\alpha(\cdot)$ of size $-r/(r-1)$, with $r>1$.
A2. The parameter space Θ is compact, and $\boldsymbol{\theta}_0$ is an interior point of Θ.
A3. For all t, $\mathbf{W}_t \equiv h(\mathbf{X}_t, \ldots, \mathbf{X}_{-\infty}) : \Omega \to \mathcal{R}^m$, where $h(\cdot)$ is a measurable function.
A4. For every t, the function $e_\tau(\cdot, \boldsymbol{\theta}) : \mathbf{W}_t \to \mathcal{R}$ is measurable, and $e_\tau(\mathbf{W}_t, \cdot) : \Theta \to \mathcal{R}$ is twice continuously differentiable almost surely (a.s.).
A5. (Correct specification) Given $\tau \in (0,1)$, there exists $\boldsymbol{\theta}_0 \in \Theta$ which minimizes

$$E\left[|\tau - \mathrm{I}(Y_t \le e_{t,\tau}(\boldsymbol{\theta}_0))|(Y_t - e_{t,\tau}(\boldsymbol{\theta}_0))^2|\mathbf{W}_t\right],$$

for every t.
A6. For every $\boldsymbol{\theta} \in \Theta$ and for all t, $E[\nabla e_{t,\tau}(\boldsymbol{\theta})\nabla^\top e_{t,\tau}(\boldsymbol{\theta})]$ is of full rank.
A7. The error term $\varepsilon_{t,\tau} \equiv Y_t - e_{t,\tau}(\boldsymbol{\theta})$ forms a stationary process, and for all t, $E(|\varepsilon_{t,\tau}|^{2r+\varepsilon}) \le A_0 < \infty$ with a given $r>1$, and some positive constants A_0 and ε.
A8. For every $\boldsymbol{\theta} \in \Theta$ and for all t, $E(\|\nabla e_{t,\tau}(\boldsymbol{\theta})\|^{2r+\varepsilon}) \le B_0 < \infty$ with a given $r>1$, and some positive constants B_0 and ε. Furthermore, for every $\boldsymbol{\theta} \in \Theta$ and for all t, it holds that $E(\|\nabla e_{t,\tau}(\boldsymbol{\theta})\varepsilon_{t,\tau}\|) < \infty$.

Assumption B (Asymptotic normality):

B1. For every $\boldsymbol{\theta} \in \Theta$ and for all t, $\|\nabla e_{t,\tau}(\boldsymbol{\theta})\| < B(\mathbf{W}_t)$, $B(\mathbf{W}_t)$ is some stochastic function of variables $\mathbf{W}_t$ satisfying $E(|B(\mathbf{W}_t)|^{4+\varepsilon}) \le B_1 < \infty$, for some positive constants B_1 and ε. Furthermore, the error term $\varepsilon_{t,\tau}$ forms a stationary process, and for every $\boldsymbol{\theta} \in \Theta$ and all t, $|\varepsilon_{t,\tau}| < A(\mathbf{W}_t)$, where $A(\mathbf{W}_t)$ is some stochastic function of variables $\mathbf{W}_t$ such that $E(|A(\mathbf{W}_t)B(\mathbf{W}_t)|^{2r+\varepsilon}) < B_2$ with a given $r>1$, and some positive constants B_2 and ε.
B2. Define two matrixes $\mathbf{V}_T(\boldsymbol{\theta}_0) = T^{-1}\sum_{t=1}^T\sum_{s=1}^T E[h_{t,\tau}(\boldsymbol{\theta}_0)h_{s,\tau}^\top(\boldsymbol{\theta}_0)]$ where $h_{t,\tau}(\boldsymbol{\theta}_0) \equiv w_{t,\tau}(\boldsymbol{\theta}_0)\varepsilon_{t,\tau}^*\nabla e_{t,\tau}(\boldsymbol{\theta}_0)$ with $\varepsilon_{t,\tau}^* = Y_t - e_{t,\tau}(\boldsymbol{\theta}_0)$ and $\mathbf{D}_T(\boldsymbol{\theta}_0) = T^{-1}\sum_{t=1}^T w_{t,\tau}(\boldsymbol{\theta}_0)\nabla e_{t,\tau}(\boldsymbol{\theta}_0)\nabla^\top e_{t,\tau}(\boldsymbol{\theta}_0)$. These two matrixes are invertible and assume that their inverses are uniformly bounded for $\boldsymbol{\theta}_0$.
B3. $\sum_{t=1}^\infty |E[|h_{t,\tau}(\boldsymbol{\theta}_0) - E(h_{t,\tau}(\boldsymbol{\theta}_0)|\mathcal{F}_t)|^{\frac{2r}{2r-1}}]|^{\frac{2r-1}{2r}} < \infty$.

Remark 2. *Assumption A1 ensures that the dataset is the realization of an α-mixing stochastic process on some suitable probability space. Assumptions A2 is a necessary condition for the proof of consistency. With Assumptions A3 and A4, the expectile lag components can be included in our model. Assumptions A5–A8 are used to establish the consistency of QMLEs, which are same as Assumptions C3–C6 in Engle and Manganelli (2004) under the quantile setting. Note that Assumptions AN1-AN4 in Engle and Manganelli (2004) is simplified to Assumptions B1–B3 and the proof of central limit theory for the functional $h_{t,\tau}(\boldsymbol{\theta}_0)$ of α-mixing process is provided. Assumption B3 is a necessary condition to establish the asymptotics of the measurable functions of infinite α-mixing random variables.*

Remark 3. *Note that Assumption A4 is different from that in Newey and Powell (1987) and Kuan et al. (2009), in which both assume a correct specification condition,* $E[|\tau - \mathrm{I}(Y_t \le e_{t,\tau}(\boldsymbol{\theta}_0))|(Y_t - e_{t,\tau}(\boldsymbol{\theta}_0))^2 \mid \mathcal{F}_t]$, *implying that* $\{w_{t,\tau}(\boldsymbol{\theta}_0)\varepsilon^*_{t,\tau}\}$ *is a martingale difference sequence. However, in our setting, the process* $\{Y_t, \mathbf{X}_t\}_{t\in\mathbb{Z}}$ *is instead strong mixing. In Assumption A3,* $\mathbf{W}_t$ *is defined as the measurable function of* $\mathbf{X}_t$ *and its lagged values. In other words, the HAC covariance matrix estimator is needed here.*

2.2 Asymptotic Properties

The consistency and asymptotic normality of the estimator $\hat{\boldsymbol{\theta}}_T$ are provided in this subsection. To simplify the presentation, we only describe the asymptotic results here, with all technical details relegated to Appendix. Next, we present the consistency and asymptotic normality of $\hat{\boldsymbol{\theta}}_T$ in Theorems 1 and 2, respectively, as follows.

Theorem 1. *(Consistency) Given Assumption A, one has* $\hat{\boldsymbol{\theta}}_T \xrightarrow{\mathcal{P}} \boldsymbol{\theta}_0$, *where* $\hat{\boldsymbol{\theta}}_T$ *is given by*

$$\hat{\boldsymbol{\theta}}_T = \arg\min_{\boldsymbol{\theta}\in\Theta} T^{-1}\sum_{t=1}^{T} w_{t,\tau}(\boldsymbol{\theta})\big(Y_t - e_{t,\tau}(\boldsymbol{\theta})\big)^2,$$

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

Note that to establish the above consistency, it needs to employ Theorem 4.3 in Wooldridge (1994), so that it needs to verify Conditions M1–M3 in Wooldridge (1994).

Theorem 2. *(Asymptotic normality) Under Assumptions A1–A6 and B, one has*

$$\sqrt{T}\big(\hat{\boldsymbol{\theta}}_T - \boldsymbol{\theta}_0\big) \xrightarrow{\mathcal{L}} \mathcal{N}(0, \Sigma(\boldsymbol{\theta}_0)),$$

where $\xrightarrow{\mathcal{L}}$ *denotes the convergence in distribution, and the asymptotic variance is given by* $\Sigma(\boldsymbol{\theta}_0) = \mathbf{D}_T^{-1}(\boldsymbol{\theta}_0)\mathbf{V}_T(\boldsymbol{\theta}_0)\mathbf{D}_T^{-1}(\boldsymbol{\theta}_0)$ *with* $\mathbf{V}_T(\boldsymbol{\theta}_0)$ *and* $\mathbf{D}_T(\boldsymbol{\theta}_0)$ *defined in Assumption B2.*

Note that the main idea of establishing the asymptotic normality of the proposed estimator $\hat{\boldsymbol{\theta}}_T$ is to express it as a linear estimator plus a higher-order term, similar to Theorem 3 in Huber (1967). Therefore, it needs to show that Huber's theorem holds and then, the asymptotic normality can be obtained by applying the central limit theorem to the linear term; see, for example, Weiss (1991).

2.3 HAC Covariance Matrix Estimation

To accommodate the case when innovations in the regression model are serially correlated, the long-run variance matrix estimation of expectile estimators is proposed. We first introduce the class of kernels $\mathcal{K}$ as in Andrews (1991). The class of kernels $\mathcal{K}$ is defined as: $k(\cdot)$ is continuous almost everywhere from $\mathcal{R}$ to $[-1,1]$, $k(0)=1$, $k(x)=k(-x)$, and $\int_{-\infty}^{\infty} k^2(x)\,dx < \infty$. Clearly, truncated kernel, Bartlett, Parzen, Tukey–Hanning and quadratic spectrum kernels are included in the class $\mathcal{K}$.

By change of variables, $V_T(\boldsymbol{\theta}_0)$ can be rewritten as

$$V_T(\boldsymbol{\theta}_0) = \sum_{j=-T+1}^{T-1} \mathrm{Cov}\big(h_{t,\tau}(\boldsymbol{\theta}_0), h_{t-j,\tau}(\boldsymbol{\theta}_0)\big),$$

and the jth order auto-covariance $H_{j,\tau}(\boldsymbol{\theta}_0) \equiv \mathrm{Cov}\big(h_{t,\tau}(\boldsymbol{\theta}_0), h_{t-j,\tau}(\boldsymbol{\theta}_0)\big)$ can be estimated by

$$\tilde{H}_{j,\tau}(\hat{\boldsymbol{\theta}}_T) = \frac{1}{T-j}\sum_{t=1}^{T-j} h_{t,\tau}(\hat{\boldsymbol{\theta}}_T) h_{t+j,\tau}^{\top}(\hat{\boldsymbol{\theta}}_T)$$

where $h_{t,\tau}(\boldsymbol{\theta}) = w_{t,\tau}(\boldsymbol{\theta})\big(Y_t - e_{t,\tau}(\boldsymbol{\theta})\big)\nabla e_{t,\tau}(\boldsymbol{\theta})$. Let S_T denote the lag-truncation order or bandwidth value, the long-run variance estimator of $V_T(\boldsymbol{\theta}_0)$ is given by

$$\tilde{V}_T(\hat{\boldsymbol{\theta}}_T) = \sum_{j=-T+1}^{T-1} \frac{T-|j|}{T} k(j/S_T)\tilde{H}_{j,\tau}(\hat{\boldsymbol{\theta}}_T),$$

where $k(\cdot) \in \mathcal{K}$. Furthermore, the term $D_T(\boldsymbol{\theta}_0)$, which captures the effect of the heteroskedasticity, can be estimated by

$$D_T(\hat{\boldsymbol{\theta}}_T) = \frac{1}{T}\sum_{t=1}^{T} \left|\tau - I\Big(Y_t \le e_{t,\tau}(\hat{\boldsymbol{\theta}}_T)\Big)\right| \nabla e_{t,\tau}(\hat{\boldsymbol{\theta}}_T)\nabla^{\top} e_{t,\tau}(\hat{\boldsymbol{\theta}}_T).$$

Finally, the HAC covariance matrix estimator for GCARE models is then given by

$$\Sigma(\hat{\boldsymbol{\theta}}_T) = D_T^{-1}(\hat{\boldsymbol{\theta}}_T)\tilde{V}_T(\hat{\boldsymbol{\theta}}_T)D_T^{-1}(\hat{\boldsymbol{\theta}}_T).$$

To establish consistency of $\Sigma(\hat{\boldsymbol{\theta}}_T)$, some notations which control the temporal dependence of $h_{t,\tau}(\boldsymbol{\theta}_0)$ are introduced. Let

$$\begin{aligned}
&\kappa_{abcd}(t, t+j, t+m, t+n) \\
&= E\Big[\{h_{a,t,\tau} - E(h_{a,t,\tau})\}\{h_{b,t+j,\tau} - E(h_{b,t+j,\tau})\}\{h_{c,t+m,\tau} - E(h_{c,t+m,\tau})\}\{h_{d,t+n,\tau} - E(h_{d,t+n,\tau})\}\Big] \\
&- E\Big[\Big\{\tilde{h}_{a,t,\tau} - E\Big(\tilde{h}_{a,t,\tau}\Big)\Big\}\Big\{\tilde{h}_{b,t+j,\tau} - E\Big(\tilde{h}_{b,t+j,\tau}\Big)\Big\}\Big\{\tilde{h}_{c,t+m,\tau} - E\Big(\tilde{h}_{c,t+m,\tau}\Big)\Big\}\Big\{\tilde{h}_{d,t+n,\tau} - E\Big(\tilde{h}_{d,t+n,\tau}\Big)\Big\}\Big]
\end{aligned}$$

denote the fourth order cumulant of $(h_{a,t,\tau}, h_{b,t+j,\tau}, h_{c,t+m,\tau}, h_{d,t+n,\tau})$, where $h_{a,t,\tau} \equiv h_{a,t,\tau}(\boldsymbol{\theta}_0)$ is the ath element of $h_{t,\tau}(\boldsymbol{\theta}_0)$, and $\{\tilde{h}_{t,\tau}\}$ denotes the Gaussian sequence with the same mean and covariance structure as $\{h_{t,\tau}(\boldsymbol{\theta}_0)\}$. The following assumptions are made to establish the consistency of $\Sigma(\hat{\boldsymbol{\theta}}_T)$.

Assumption C (HAC Covariance Matrix):

C1. For any positive integer $a, b, c, d \le p$, $\sum_{j=-\infty}^{\infty}\sum_{m=-\infty}^{\infty}\sum_{n=-\infty}^{\infty} \kappa_{abcd}(0, j, m, n) < \infty$, in which $h_{t,\tau}(\boldsymbol{\theta}_0)$ is a fourth order stationary sequence of random variables with $\sum_{j=-\infty}^{\infty} \|E(h_{t,\tau}(\boldsymbol{\theta}_0)h_{t-j,\tau}^{\top}(\boldsymbol{\theta}_0))\| < \infty$.

C2. There is a measurable function $\alpha(\cdot)$ satisfying $f(y|\mathbf{W}_t) \le \alpha(\mathbf{W}_t)$, where $f(y|\mathbf{W}_t)$ is the conditional probability density function of Y_t given $\mathbf{W}_t$, and $\alpha(\mathbf{W}_t)$ is integrable with respect to y.

C3. For every $\boldsymbol{\theta} \in \Theta$ and for all t, $\|\nabla^2 e_{t,\tau}(\boldsymbol{\theta})\| < M(\mathbf{W}_t)$, where $M(\mathbf{W}_t)$ is some stochastic function of variables $\mathbf{W}_t$ satisfying $E(|M(\mathbf{W}_t)|^{2+\varepsilon}) < M_1 < \infty$, $E(|A(\mathbf{W}_t)M(\mathbf{W}_t)|^{2+\varepsilon}) \le M_2 < \infty$, and $E(|A(\mathbf{W}_t)B(\mathbf{W}_t)M(\mathbf{W}_t)|^{1+\varepsilon}) \le M_3 < \infty$ for some positive constants M_1, M_2, M_3 and ε.

Remark 4. *Assumptions C1 is a standard cumulant condition in the time series literature, which allows for conditional heteroskedasticity and autocorrelation, but prohibits unconditional heteroskedasticity. Assumptions C2 and C3 are used to obtain the consistency of $\tilde{V}_T(\hat{\boldsymbol{\theta}}_T)$ and $D_T(\hat{\boldsymbol{\theta}}_T)$, in which Assumption C2 is same as Assumption 3 in Newey and Powell (1987) and Assumption C3 is equivalent to Assumption B(ii)–(iii) in Andrews (1991) under mean regression setting.*

Finally, the consistency result of HAC covariance matrix is as follows.

Theorem 3. *Suppose Assumptions A1–A6, B, and C hold, and $S_T^2/T \xrightarrow{\mathcal{P}} 0$. Let $k(\cdot) \in \mathcal{K}$. Then, $\Sigma_T(\hat{\boldsymbol{\theta}}_T)$ is a consistent estimate of $\Sigma_T(\boldsymbol{\theta}_0)$.*

2.4 Dynamic Expectile Test

The correct specification condition for any expectile model should be

$$E\left[w_{t,\tau}(\boldsymbol{\theta}_0)\varepsilon^*_{t,\tau}|\mathbf{W}_t\right] = 0 \tag{4}$$

for every t and $1 \leq t \leq T$. In Kuan, Yeh, and Hsu (2009), an encompassing test is proposed to select an appropriate model between two linear non-nested models. Because the null and alternative models need to first be determined, the encompassing test is not flexible enough to compare the performance of more different expectile models. Furthermore, the existing correct specification tests cannot address cases in which expectile lags are incorporated into the model framework. In this paper, we propose a dynamic expectile test of the form

$$H_0 : E\left[w_{t,\tau}(\boldsymbol{\theta})\varepsilon_{t,\tau}|\mathbf{W}_t\right] = 0 \,\text{versus}\, H_1 : E\left[w_{t,\tau}(\boldsymbol{\theta})\varepsilon_{t,\tau}|\mathbf{W}_t\right] \neq 0, \tag{5}$$

where $w_{t,\tau}(\boldsymbol{\theta}) = |\tau - \mathrm{I}(Y_t \leq e_{t,\tau}(\boldsymbol{\theta}))|$ with $e_{t,\tau}(\boldsymbol{\theta})$ denoting the expectile estimate or forecast of the testing model, $\varepsilon_{t,\tau} = Y_t - e_{t,\tau}(\boldsymbol{\theta})$.

Although a general test of Equation (5) is desirable, however, it is still challenging for such a test because of the high dimensionality of $\mathbf{W}_t$. In light of Engle and Manganelli (2004), we consider an unconditional test of Equation (5) where the intermediate statistics are used for testing specific implications of the general hypothesis such that particular inadequacies of a model can be revealed. From Equation (4), the conditional expectation of $w_{t,\tau}(\boldsymbol{\theta})\varepsilon_{t,\tau}$ given $\mathbf{W}_t$ is zero, which implies that the expected value of $w_{t,\tau}(\boldsymbol{\theta})\varepsilon_{t,\tau}$ is zero too. Furthermore, $w_{t,\tau}(\boldsymbol{\theta})\varepsilon_{t,\tau}$ must be uncorrelated with its own lag values and with $e_{t,\tau}(\boldsymbol{\theta})$ and/or other expectile lags. Therefore, the selection of the instruments is very important for the power of the test. When we choose the instruments, which are not informative to reveal the particular inadequacies of a model, the test would have low power.[1]

To examine whether $w_{t,\tau}(\hat{\boldsymbol{\theta}})\hat{\varepsilon}_{t,\tau}(\hat{\boldsymbol{\theta}})$ satisfies the above conditions, we set up a test to check whether the test statistic $T^{-1/2}\sum_{t=1}^{T}\mathbf{Z}_t(\hat{\boldsymbol{\theta}})w_{t,\tau}(\hat{\boldsymbol{\theta}})\hat{\varepsilon}_{t,\tau}(\hat{\boldsymbol{\theta}})$ is significantly different from zero, in which $\mathbf{Z}_t(\boldsymbol{\theta})$ is the instrument selected. Normally, the lagged $w_{t-i,\tau}(\hat{\boldsymbol{\theta}})\hat{\varepsilon}_{t-i,\tau}(\hat{\boldsymbol{\theta}})$ $(i = 1,\ldots,m)$ are suggested to be included in $\mathbf{Z}_t(\hat{\boldsymbol{\theta}})$, and other functions in the set of $\mathbf{W}_t$ can also be incorporated if they are suspected of being informative. It is worthy to mention that the new test can easily be extended to incorporate various alternatives.

[1] We thank a referee to bring the following issue to our attention. In general, it is challenging to name the alternative hypothesis against which the test has the most/the least power. Nonetheless, by examining the construction of the Dynamic Expectile (DE) test, we can glean some insights. On one hand, since each GCARE(p, q) model can be reformulated as an infinite-dimensional CARE model as proposed by Kuan et al. (2009), the power of a test distinguishing between two GCARE(p, q) with different p and q is likely to be low. To address this, we advocate for the application of the EBIC to select the optimal p and q for GCARE models.

Before deriving the distribution of in-sample and out-of-sample dynamic expectile tests, some required assumptions are introduced.

Assumption D (In-sample dynamic expectile test):

D1. The function $\mathbf{Z}_t(\boldsymbol{\theta})$ is a $q\times 1$ vector of some stochastic functions of variables that belongs to the information set, and for every $\boldsymbol{\theta}\in\Theta$ and all t, $\|\mathbf{Z}_t(\boldsymbol{\theta})\|<C(\mathbf{W}_t)$, where $C(\mathbf{W}_t)$ is some stochastic function of variables $\mathbf{W}_t$ such that $E(|C(\mathbf{W}_t)|^{2+\varepsilon})\le C_1<\infty$ and $E(|B(\mathbf{W}_t)C(\mathbf{W}_t)|^{2r+\varepsilon})\le C_2<\infty$, with a given $r>1$, and some positive constants C_1, C_2 and ε.

D2. For every $\boldsymbol{\theta}\in\Theta$ and for all t, the function $\mathbf{Z}_t(\boldsymbol{\theta})$ is differentiable, and it holds that $\|\nabla\mathbf{Z}_t(\boldsymbol{\theta})\|<D(\mathbf{W}_t)$, where $D(\mathbf{W}_t)$ is some stochastic function of variables $\mathbf{W}_t$ such that $E(|A(\mathbf{W}_t)D(\mathbf{W}_t)|^{2+\varepsilon})\le D_1<\infty$ with some positive constants D_1 and ε.

D3. Let $\Xi_t(\boldsymbol{\theta}_0)=\mathbf{Z}_t(\boldsymbol{\theta}_0)+(T^{-1}\sum_{t=1}^{T}E[\mathbf{Z}_t(\boldsymbol{\theta}_0)w_{t,\tau}(\boldsymbol{\theta}_0)\nabla^\top e_{t,\tau}(\boldsymbol{\theta}_0)])\mathbf{D}_T^{-1}\nabla e_{t,\tau}(\boldsymbol{\theta}_0)$. For every $\boldsymbol{\theta}\in\Theta$ and all t, it holds that $E\|\Xi_t(\boldsymbol{\theta})\varepsilon_{t,\tau}\|^{2r+\varepsilon}<\infty$ with a given $r>1$, and a positive constant ε.

D4. $\sum_{t=1}^{\infty}|E[|g_{t,\tau}(\boldsymbol{\theta}_0)-E(g_{t,\tau}(\boldsymbol{\theta}_0)|\mathcal{F}_t)|^{\frac{2r}{2r-1}}]|^{\frac{2r-1}{2r}}<\infty$, where $g_{t,\tau}(\boldsymbol{\theta}_0)\equiv\Xi_{t,\tau}(\boldsymbol{\theta}_0)w_{t,\tau}(\boldsymbol{\theta}_0)\varepsilon^*_{t,\tau}$.

Assumption E (Out-of-sample dynamic expectile test):

E1. Let T_r denote the number of in-sample observations, and N_r be the number of out-of-sample observations. Then, it holds that $\lim_{r\to\infty}T_r=\infty$, $\lim_{r\to\infty}N_r=\infty$, and $\lim_{r\to\infty}N_r/T_r=0$.

E2. For every $\boldsymbol{\theta}\in\Theta$ and all t, it holds that $E(|A(\mathbf{W}_t)C(\mathbf{W}_t)|^{2r+\varepsilon})\le C_3<\infty$ with a given $r>1$, and some positive constants C_3 and ε.

Remark 5. *Although our dynamic test method is inspired by the idea of dynamic quantile test as in Engle and Manganelli (2004), it is new in the expectile regression literature. Note that Assumptions D1–D4 are different from Assumptions DQ4–6 in Engle and Manganelli (2004). Assumption D4 is a necessary conditions to establish the asymptotics of the measurable functions of infinite α-mixing random variables. Here, we provide primary conditions in time series setting which are sufficient to derive Assumptions DQ4–6 in Engle and Manganelli (2004). Assumptions E1 and E2 are the same as DQ8–9 in Engle and Manganelli (2004).*

The asymptotic distribution of the in-sample dynamic expectile test is considered in Theorem 4, and Theorem 5 is for the out-of-sample case.

Theorem 4. *(In-sample dynamic expectile test) Under Assumptions A1–A6, B, and D, when $T\to\infty$, one has*

$$\left[\sum_{t=1}^{T}\mathbf{Z}_t(\hat{\boldsymbol{\theta}}_T)w_{t,\tau}(\hat{\boldsymbol{\theta}}_T)\hat{\varepsilon}_{t,\tau}(\hat{\boldsymbol{\theta}}_T)\right]\boldsymbol{\Lambda}_I^{-1}(\boldsymbol{\theta}_0)\left[\sum_{t=1}^{T}\mathbf{Z}_t(\hat{\boldsymbol{\theta}}_T)w_{t,\tau}(\hat{\boldsymbol{\theta}}_T)\hat{\varepsilon}_{t,\tau}(\hat{\boldsymbol{\theta}}_T)\right]\xrightarrow{\mathcal{L}}\chi^2_q,$$

where $\Lambda_I(\boldsymbol{\theta}_0)\equiv T^{-1}\sum_{t=1}^{T}\sum_{s=1}^{T}E(\Xi_t(\boldsymbol{\theta}_0)w_{t,\tau}(\boldsymbol{\theta}_0)\varepsilon^*_{t,\tau}\Xi_s^\top(\boldsymbol{\theta}_0)w_{s,\tau}(\boldsymbol{\theta}_0)\varepsilon^*_{s,\tau})$ with its inverse function uniformly bounded, and q denotes the dimension of $\mathbf{Z}_t(\hat{\boldsymbol{\theta}}_T)$.

Remark 6. *The in-sample DE test is a model specification test for the expectile model under study, and it is particularly useful for model selection purposes. Note that* *when $\mathbf{Z}_t(\boldsymbol{\theta})$ is selected to be the regressors of alternative model and the latent variable $e_{t-j,\tau}(\boldsymbol{\theta})$ is not included, our in-sample test is simplified to the encompassing test as in Kuan, Yeh, and Hsu (2009).*

Let $\hat{\boldsymbol{\theta}}_{T_r}$ be the QMLE estimated using in-sample observations. The out-of-sample DE test is as follows.

Theorem 5. *(Out-of-sample dynamic expectile test) Under Assumptions A1–A6, B, D1, and E, when $r \to \infty$, one has*

$$\left[\sum_{t=T_r+1}^{T_r+N_r} \mathbf{Z}_t\left(\hat{\boldsymbol{\theta}}_{T_r}\right) w_{t,\tau}\left(\hat{\boldsymbol{\theta}}_{T_r}\right) \hat{\varepsilon}_{t,\tau}\left(\hat{\boldsymbol{\theta}}_{T_r}\right)\right] \boldsymbol{\Lambda}_O^{-1}(\boldsymbol{\theta}_0)) \left[\sum_{t=T_r+1}^{T_r+N_r} \mathbf{Z}_t\left(\hat{\boldsymbol{\theta}}_{T_r}\right) w_{t,\tau}\left(\hat{\boldsymbol{\theta}}_{T_r}\right) \hat{\varepsilon}_{t,\tau}\left(\hat{\boldsymbol{\theta}}_{T_r}\right)\right] \xrightarrow{\mathcal{L}} \chi_q^2,$$

where $\Lambda_O(\boldsymbol{\theta}_0) \equiv \sum_{t=T_r+1}^{T_r+N_r} \sum_{s=T_r+1}^{T_r+N_r} E(\mathbf{Z}_t(\boldsymbol{\theta}_0) w_{t,\tau}(\boldsymbol{\theta}_0) \varepsilon_{t,\tau}^* \mathbf{Z}_s^\top(\boldsymbol{\theta}_0) w_{s,\tau}(\boldsymbol{\theta}_0) \varepsilon_{s,\tau}^*)$ with its inverse function uniformly bounded, and q denotes the dimension of $\mathbf{Z}_t(\hat{\boldsymbol{\theta}}_{T_r})$.

The out-of-sample DE test can be used to compare the relative performances of different expectile models in terms of predictive ability. Note that the out-of-sample DE test has a nice feature that it does not depend on the estimation procedure. To implement the test, only the sequences of expectile forecasts and return series are needed.

Remark 7. *In the out-of-sample DE test, the constant and expectile forecast are excluded from the instruments. According to Engle and Manganelli (2004), for some particular models, there was collinearity with the matrix of derivatives when the constant and expectile forecast are included. Furthermore, it is worth mentioning that the selection of $\mathbf{Z}_t(\boldsymbol{\theta})$ is very important for the power of in-sample DE test and out-of-sample DE test. As reminded by one of our anonymous referees, there is a trade-off between the number of instruments used and the estimation uncertainty that arises. From the simulation results in Example 3, one can see that the adding of an informative instrument would effectively increase the power of the test, while the size of test deteriorate with a higher number of instruments.*

3 Simulation Studies

In this section, we consider three simulation examples to examine the finite sample performances of our proposed estimators, model selection criterion EBIC and test statistics, respectively. In all of the examples, 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 of the initial values.

Example 1. The data generating process (DGP) follows a linear GARCH(1,1) model given by

$$Y_t = \sigma_t \varepsilon_t \quad \text{with} \quad \sigma_t = \alpha_0 + \alpha_1 |Y_{t-1}| + \beta_1 \sigma_{t-1}, \tag{6}$$

where $\alpha_0 = 0.1$, $\alpha_1 = 0.3$, and $\beta_1 = 0.5$, and ε_t is generated from an i.i.d. Gaussian. The corresponding GCARE(1,1) model is then given by

$$e_{t,\tau} = \bar{\alpha}_{0,\tau} + \bar{\alpha}_{1,\tau}|Y_{t-1}| + \beta_1 e_{t-1,\tau},$$

and the parameters $\bar{\alpha}_{0,\tau} = \alpha_0 e_\tau(\varepsilon_t)$, $\bar{\alpha}_{1,\tau} = \alpha_1 e_\tau(\varepsilon_t)$, where $e_\tau(\varepsilon_t)$ is the τ-th expectile of ε_t. We then apply the estimation procedure developed in the Section 1.2 to estimate the parameters.

Two probability levels $\tau = 0.01$ and 0.05 are considered and simulations are repeated 500 times with three different sample sizes $T = 500$, 1000 and 2000, respectively. To measure estimation performance, the median and the standard deviation (SD) of the absolute deviation of errors (ADE) are reported, where $\text{ADE}_{\theta_j}^{(k)} \equiv |\hat{\theta}_j^{(k)} - \theta_j|$ for $0 \leq j \leq (p+q)$, and $\hat{\theta}_j^{(k)}$ is the estimator in the k-th simulation replication.

The median and SD (in parentheses) of the ADE values of the corresponding QMLE estimators $\hat{\alpha}_0$ (ADE_{α_0}), $\hat{\alpha}_1$ (ADE_{α_1}), and $\hat{\beta}_1$ (ADE_{β_1}) in all cases are reported in Table 1. It is easy to find that the median and SD of ADE values for all the QMLE estimators decrease with the sample sizes. For example, when the sample size $T = 500$, the median and SD of the ADE_{α_1} are 0.1283 and 0.1190 under the probability level $\tau = 0.01$, and they decrease respectively to 0.0937 and 0.0940 when the sample size increases to $T = 1000$. Furthermore, we can observe a same pattern for ADE_{α_0} as well. Specifically, when the sample size is 500 and the probability level is 0.05, the median and SD are 0.0506 and 0.0691, respectively. When the sample size increases to 1000, the corresponding median and SD decrease to 0.0308 and 0.0375, respectively. Finally, the performance for ADE_{β_1} improves when the sample size increases. The median and SD of ADE_{β_1} are 0.0927 and 0.1161 under the sample size $T = 1000$ and probability level $\tau = 0.01$, and they decrease to 0.0654 and 0.0698 when the sample size is doubled.

Example 2. Let $Y_t = e_{t,\tau} + \varepsilon_{t,\tau}$, where $\varepsilon_{t,\tau}$ follows an i.i.d. asymmetric normal distribution AND $(0, \sigma^2_{\varepsilon_{t,\tau}}, \tau)$ with density function

$$f(\varepsilon_{t,\tau}) = \frac{2}{\sigma_{\varepsilon_{t,\tau}}} \left(\sqrt{\frac{\pi}{1-\tau}} + \sqrt{\frac{\pi}{\tau}} \right)^{-1} \exp\left\{ -\text{Q}_\tau\left(\frac{\varepsilon_{t,\tau}}{\sigma_{\varepsilon_{t,\tau}}} \right) \right\},$$

and $\sigma_{\varepsilon_{t,\tau}} = 0.5$, which was also used in Gerlach and Chen (2016), Gerlach, Walpole, and Wang (2017), and Xu, Mihoci, and Härdle (2018). Note that the τth expectile of $\varepsilon_{t,\tau}$ equals 0, which satisfies the model identification condition. We consider the following three settings:

A: $e_{t,\tau} = \alpha_0 + \alpha_1|Y_{t-1}| + \beta_1 e_{t-1,\tau}$ with $\alpha_0 = -0.4$, $\alpha_1 = -0.2$, and $\beta_1 = 0.6$,
B: $e_{t,\tau} = \alpha_0 + \alpha_1|Y_{t-1}| + \alpha_2|Y_{t-2}| + \alpha_3|Y_{t-3}| + \beta_1 e_{t-1,\tau}$ with $\alpha_0 = -0.6$, $\alpha_1 = -0.1$, $\alpha_2 = -0.2$, $\alpha_3 = 0.4$, and $\beta_1 = 0.6$,
C: $e_{t,\tau} = \alpha_0 + \alpha_1|Y_{t-1}| + \alpha_2|Y_{t-2}| + \alpha_3|Y_{t-3}| + \beta_1 e_{t-1,\tau} + \beta_2 e_{t-2,\tau}$ with $\alpha_0 = -0.6$, $\alpha_1 = -0.1$, $\alpha_2 = -0.2$, $\alpha_3 = 0.4$, $\beta_1 = 0.2$, and $\beta_2 = 0.6$.

Two probability levels $\tau = 0.01$ and 0.05 are considered and simulations are repeated 500 times with three different sample sizes $T = 500$, 1000 and 2000, respectively. We then use the estimation method introduced in Section 1.2 to estimate the parameters of the model. To measure estimation performance, the median and the SD of the ADE are

Table 1. Median and SD (in parentheses) of the ADE values in Example 1.

	$\tau=0.01$			$\tau=0.05$		
T	500	1000	2000	500	1000	2000
ADE_{α_0}	0.0778	0.0483	0.0361	0.0506	0.0308	0.0229
	(0.0963)	(0.0628)	(0.0360)	(0.0691)	(0.0375)	(0.0258)
ADE_{α_1}	0.1283	0.0937	0.0621	0.0924	0.0545	0.0437
	(0.1190)	(0.0940)	(0.0664)	(0.1001)	(0.0711)	(0.0566)
ADE_{β_1}	0.1504	0.0927	0.0654	0.1534	0.0883	0.0686
	(0.1678)	(0.1161)	(0.0698)	(0.1936)	(0.1121)	(0.0853)

reported. Moreover, the accuracy of EBIC criterion for all the experimental settings A, B, and C is reported as well.

The median and SD (in parentheses) of the ADE values of the corresponding QMLE estimators $\hat{\alpha}_0$ (ADE_{α_0}), $\hat{\alpha}_1$ (ADE_{α_1}), $\hat{\alpha}_2$ (ADE_{α_2}), $\hat{\alpha}_3$ (ADE_{α_3}), $\hat{\beta}_1$ (ADE_{β_1}) and $\hat{\beta}_2$ (ADE_{β_2}) in all cases are reported in Table 2.

First, we find that the median and SD of ADE values for all the QMLE estimators decrease as the sample size increases. For example, when $T=500$ in Setting A, the median and SD of the ADE_{β_1} are 0.0907 and 0.1193 under the probability level $\tau=0.05$, and they decrease to 0.0409 and 0.0393, respectively, when the sample size increases to 2000. Clearly, the same pattern for ADE_{α_0} can also be observed. Indeed, when $\tau=0.01$ and the sample size is 500, the median and the corresponding SD are 0.0496 and 0.0502 in Setting B. When the sample size increases to 1000, the median and the corresponding SD decrease to 0.0379 and 0.0355, respectively. Finally, the performance for ADE_{α_1}, ADE_{α_2} and ADE_{α_3} is better than that for ADE_{α_0}, ADE_{β_1} and ADE_{β_2}. We find that when the sample size is 500 in Setting C, the median and SD of ADE_{α_1} are 0.0316 and 0.0286 under the probability level $\tau=0.05$, and they decrease to 0.0193 and 0.0209 as the sample size doubled.

Table 3 reports the accuracy of EBIC criterion for selecting the correct p and q for all the experimental settings A, B, and C. Under three settings and all probability levels, we find that the accuracy increases as the sample size increases. For example, when $T=500$, the accuracy of EBIC for selecting the correct order of Setting A are 98.6% and 93.6% under the probability level $\tau=0.01$ and 0.05, and they increase to 98.8% and 97.8%, respectively, when the sample size increases to 1000. Similar results for Setting B can also be observed. For Setting C, the accuracy of EBIC for selecting the correct order when $T=500$ are 87.0% and 69.0% under the probability level $\tau=0.01$ and 0.05, and when the sample size is 2000, they increase to 99.8% and 99.4%, respectively.

Example 3. In this example, we examine the finite-sample properties of the proposed test. To this end, let $Y_t=e_{t,\tau}+\varepsilon_{t,\tau}$, where $\varepsilon_{t,\tau}$ follows an i.i.d. asymmetric normal distribution $\mathrm{AND}(0,\sigma^2_{\varepsilon_{t,\tau}},\tau)$ with $\sigma_{\varepsilon_{t,\tau}}=0.5$. Here we consider the DGP based on the following set up:

$$e_{t,\tau}=(1-\gamma)(\alpha_0+\alpha_1X_t-\beta_1e_{t-1,\tau})-\gamma\sqrt{a_0+a_1X_t^2+b_1e_{t-1,\tau}^2},$$

with $\alpha_0=-0.1$, $\alpha_1=-0.05$, $\beta_1=0.5$, $a_0=0.1$, $a_1=0.2$, and $b_1=0.95$, where $X_t=0.3X_{t-1}-0.5X_{t-2}+v_t$ in which v_t follows a normal distribution $N(0,2^2)$. To permit the stationarity of the series, γ is a selected between 0 and 1 with a series of grid points. The model is referred to as the null model when $\gamma=0$, with all other values of γ indicating alternative models.

Table 2. Median and SD (in parentheses) of the ADE values under three settings.

	$\tau = 0.01$			$\tau = 0.05$		
T	500	1000	2000	500	1000	2000
			Setting A			
ADE_{α_0}	0.1009	0.0639	0.0500	0.1126	0.0869	0.0530
	(0.1155)	(0.0639)	(0.0477)	(0.1717)	(0.0836)	(0.0549)
ADE_{α_1}	0.0205	0.0153	0.0105	0.0329	0.0223	0.0165
	(0.0201)	(0.0137)	(0.0099)	(0.0296)	(0.0212)	(0.0144)
ADE_{β_1}	0.0622	0.0303	0.0294	0.0907	0.0650	0.0409
	(0.0673)	(0.0379)	(0.0280)	(0.1193)	(0.0620)	(0.0393)
			Setting B			
ADE_{α_0}	0.0496	0.0379	0.0254	0.0975	0.0591	0.0459
	(0.0502)	(0.0355)	(0.0255)	(0.0916)	(0.0605)	(0.0404)
ADE_{α_1}	0.0198	0.0134	0.0092	0.0419	0.0255	0.0181
	(0.0178)	(0.0126)	(0.0085)	(0.0313)	(0.0239)	(0.0166)
ADE_{α_2}	0.0242	0.0147	0.0112	0.0459	0.0386	0.0238
	(0.0216)	(0.0143)	(0.0093)	(0.0395)	(0.0292)	(0.0201)
ADE_{α_3}	0.0191	0.0131	0.0096	0.0382	0.0264	0.0190
	(0.0183)	(0.0120)	(0.0079)	(0.0330)	(0.0237)	(0.0163)
ADE_{β_1}	0.0417	0.0278	0.0197	0.0835	0.0531	0.0366
	(0.0374)	(0.0275)	(0.0185)	(0.0789)	(0.0510)	(0.0341)
			Setting C			
ADE_{α_0}	0.0971	0.0725	0.0526	0.1489	0.1046	0.0698
	(0.1522)	(0.0785)	(0.0518)	(0.1715)	(0.1221)	(0.0790)
ADE_{α_1}	0.0243	0.0171	0.0117	0.0316	0.0193	0.0132
	(0.0217)	(0.0153)	(0.0105)	(0.0286)	(0.0209)	(0.0124)
ADE_{α_2}	0.0228	0.0160	0.0103	0.0262	0.0188	0.0135
	(0.0218)	(0.0140)	(0.0096)	(0.0245)	(0.0178)	(0.0117)
ADE_{α_3}	0.0260	0.0178	0.0116	0.0313	0.0222	0.0158
	(0.0257)	(0.0160)	(0.0110)	(0.0270)	(0.0205)	(0.0150)
ADE_{β_1}	0.0396	0.0268	0.0180	0.0502	0.0319	0.0227
	(0.0428)	(0.0264)	(0.0172)	(0.0531)	(0.0356)	(0.0211)
ADE_{β_2}	0.0388	0.0248	0.0188	0.0473	0.0324	0.0245
	(0.0443)	(0.0252)	(0.0165)	(0.0500)	(0.0325)	(0.0211)

Four different probability levels $\tau = 0.01$, 0.05, 0.10, and 0.25 are considered. Simulations are repeated 10,000 times for each of the given sample sizes $T = 1500$ and 3000. We then examine the finite-sample performance of the proposed in-sample DE test and out-of-sample DE test by rejection rates under the null and alternative models, respectively. To investigate how the number of instruments may affect the finite sample properties of the test, we have explored scenarios incorporating one, two, three, and four lags of $w_{t,\tau}(\hat{\boldsymbol{\beta}})\varepsilon_{t,\tau}(\hat{\boldsymbol{\beta}})$ together with one and expectile forecast within the instrument set $\mathbf{Z}_t(\hat{\boldsymbol{\theta}})$, denoted as instruments Setting A, Setting B, Setting C, and Setting D, respectively.

First, we demonstrate the variation in rejection rates for both in-sample and out-of-sample tests across varying γ values. Four distinct probability levels: $\tau = 0.01$, 0.05, 0.10, and 0.25, with fixed sample sizes of $T = 1500$ and $T = 3000$, are considered. As depicted in Figures 1 and 2, the rejection rates for the in-sample and out-of-sample tests are

Table 3. Model selection results under three settings.

T	$\tau = 0.01$			$\tau = 0.05$		
	500	1000	2000	500	1000	2000
Setting A	98.6%	98.8%	99.8%	93.6%	97.8%	100.0%
Setting B	99.0%	99.4%	99.8%	94.2%	99.2%	100.0%
Setting C	87.0%	99.2%	99.8%	69.0%	95.0%	99.4%

presented for different γ values under instrument Setting C. These figures clearly illustrate that the rejection rates for both tests increase with γ and are particularly pronounced for probability levels $\tau = 0.05$, 0.10, and 0.25. It is easy to find that the DE tests suffers from substantial size distortion for expectiles in the extremal tails ($\tau = 0.01$) when the sample size $T = 1500$. The similar situation has also been observed in Horvath et al. (2022) for quantiles. However, the size distortion immensely disappears when the sample size increases to $T = 3000$.

Table 4 reports the finite-sample rejection rates for null model ($\gamma = 0$) and alternative model ($\gamma = 1$) of in-sample DE test. In general, the performance of the in-sample DE test is quite good under the probability level of $\tau = 0.25$, 0.10 and 0.05. For example, when the sample size is $T = 3000$ and the instrument is Setting C, the rejection rate of the in-sample DE test under null model is 0.0527, 0.0521, and 0.0602 with the probability level $\tau = 0.25$, 0.10 and 0.05, respectively. Meanwhile, the power of the in-sample DE test under alternative model ($\gamma = 1$) is 1.0000 for the probability level $\tau = 0.25$, 0.10, and 0.05, and 0.9554 for $\tau = 0.01$, respectively, when the sample size is $T = 1500$.

Our analysis has identified a trade-off between the number of instruments utilized and the resultant estimation uncertainty, within the framework of the in-sample DE test. This trade-off is pronounced at probability levels of $\tau = 0.05$ and $\tau = 0.01$. For instance, when the sample size is $T = 3000$, the rejection rate of the in-sample DE test under the null hypothesis for instrument Setting B is 0.0525, 0.0513, 0.0573, and 0.1041 for the probability level $\tau = 0.25$, 0.10, 0.05 and 0.01, respectively. Under the same sample size and for $\tau = 0.25$, 0.10, 0.05, and 0.01, the rejection rate increase to 0.0527, 0.0521, 0.0602, and 0.1202, respectively, when instrument Setting C is employed. Moreover, the power of the in-sample DE test, given $T = 1500$ and $\tau = 0.01$, is 0.7946 for instrument Setting A. This power increases to 0.9705 upon switching to instrument Setting D, further illustrating the impact of the choice of instruments on the test's efficacy.

Table 5 reports the finite-sample rejection rates for the null model ($\gamma = 0$) and alternative model ($\gamma = 1$) of out-of-sample DE test. It is evident that size distortion still exits when the probability $\tau = 0.01$ and sample size $T = 1500$, however, the situation has been improved when the sample size increases to $T = 3000$. Generally, the out-of-sample DE test performs well under probability levels of $\tau = 0.25$, 0.10 and 0.05. For instance, when the sample size is $T = 3000$ and the instrument is Setting B, the rejection rate of the out-of-sample DE test under null model is 0.0530, 0.0550 and 0.0659 for $\tau = 0.25$, 0.10 and 0.05, respectively. Meanwhile, the power of the out-of-sample DE test under alternative hypothesis is 1.0000 for the probability level $\tau = 0.25$, 0.10, and 0.05, and 0.9978 for $\tau = 0.01$, respectively, under the same sample size. Furthermore, it is observed that the power of the out-of-sample DE test increase with the number of the instruments. Conversely, the situation is reversed when considering the size of the test. This observation highlights the trade-off between the number of instruments used and the estimation uncertainty that arises.

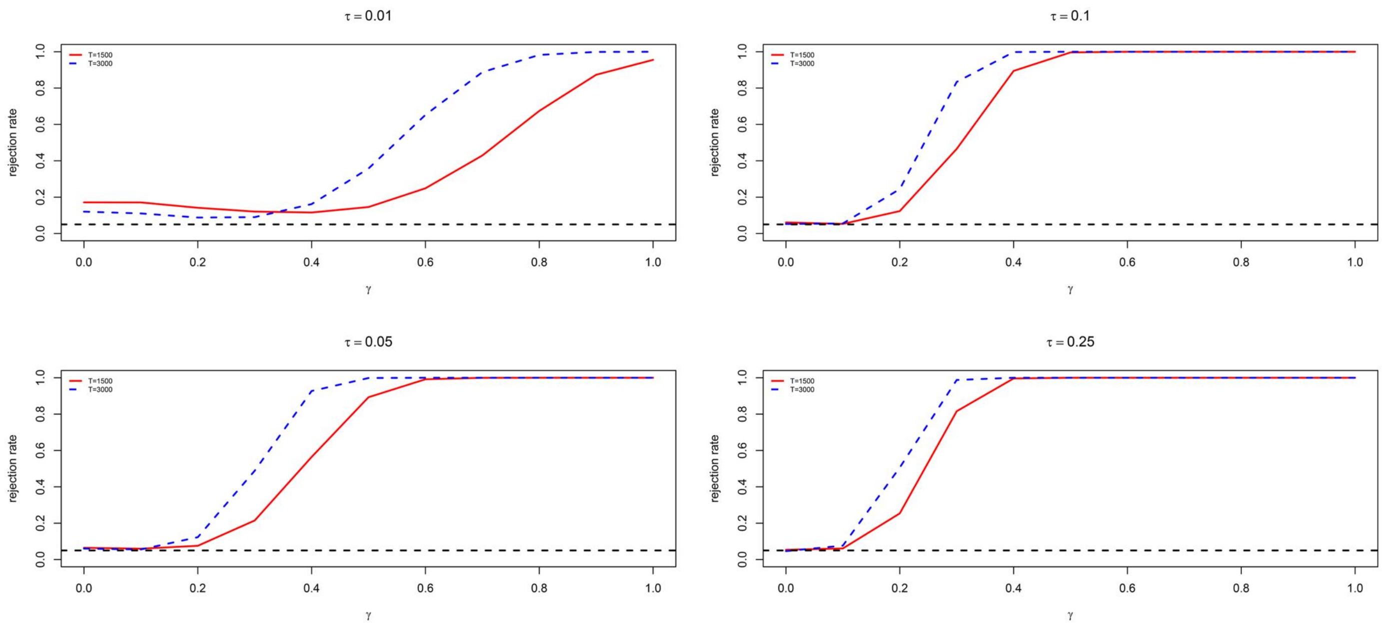

Figure 1. Rejection rate of in-sample DE test for different γs under instrument Setting C.

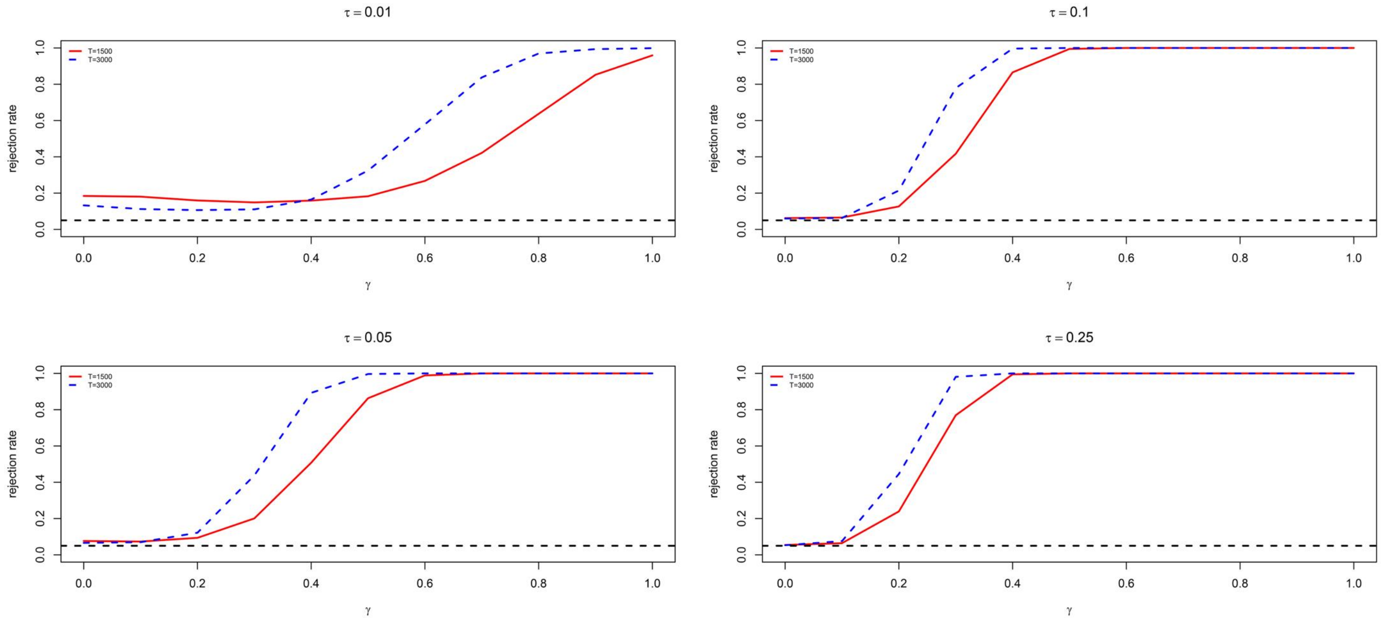

Figure 2. Rejection rate of out-of-sample DE test for different γs under instrument Setting C.

Table 4. Finite-sample rejection rates for null and alternative model: in-sample test.

	$\gamma=0$				$\gamma=1$			
τ	**0.25**	**0.10**	**0.05**	**0.01**	**0.25**	**0.10**	**0.05**	**0.01**
				Setting A				
$T=1500$	0.0506	0.0554	0.0631	0.1021	1.0000	1.0000	1.0000	0.7946
$T=3000$	0.0496	0.0510	0.0573	0.0813	1.0000	1.0000	1.0000	0.9919
				Setting B				
$T=1500$	0.0516	0.0572	0.0697	0.1468	1.0000	1.0000	1.0000	0.9127
$T=3000$	0.0525	0.0513	0.0573	0.1041	1.0000	1.0000	1.0000	0.9994
				Setting C				
$T=1500$	0.0536	0.0608	0.0646	0.1714	1.0000	1.0000	1.0000	0.9554
$T=3000$	0.0527	0.0521	0.0602	0.1202	1.0000	1.0000	1.0000	0.9999
				Setting D				
$T=1500$	0.0518	0.0585	0.0771	0.2047	1.0000	1.0000	1.0000	0.9705
$T=3000$	0.0462	0.0546	0.0614	0.1374	1.0000	1.0000	1.0000	1.0000

Note: Setting A: $\mathbf{Z}_t(\hat{\boldsymbol{\theta}}) = (1, \hat{e}_{t,\tau}, w_{t-1,\tau}(\hat{\boldsymbol{\theta}})\varepsilon_{t-1,\tau}(\hat{\boldsymbol{\theta}}))^\top$.
Setting B: $\mathbf{Z}_t(\hat{\boldsymbol{\theta}}) = (1, \hat{e}_{t,\tau}, w_{t-1,\tau}(\hat{\boldsymbol{\theta}})\varepsilon_{t-1,\tau}(\hat{\boldsymbol{\theta}}), w_{t-2,\tau}(\hat{\boldsymbol{\theta}})\varepsilon_{t-2,\tau}(\hat{\boldsymbol{\theta}}))^\top$.
Setting C: $\mathbf{Z}_t(\hat{\boldsymbol{\theta}}) = (1, \hat{e}_{t,\tau}, w_{t-1,\tau}(\hat{\boldsymbol{\theta}})\varepsilon_{t-1,\tau}(\hat{\boldsymbol{\theta}}), w_{t-2,\tau}(\hat{\boldsymbol{\theta}})\varepsilon_{t-2,\tau}(\hat{\boldsymbol{\theta}}), w_{t-3,\tau}(\hat{\boldsymbol{\theta}})\varepsilon_{t-3,\tau}(\hat{\boldsymbol{\theta}}))^\top$.
Setting D: $\mathbf{Z}_t(\hat{\boldsymbol{\theta}}) = (1, \hat{e}_{t,\tau}, w_{t-1,\tau}(\hat{\boldsymbol{\theta}})\varepsilon_{t-1,\tau}(\hat{\boldsymbol{\theta}}), w_{t-2,\tau}(\hat{\boldsymbol{\theta}})\varepsilon_{t-2,\tau}(\hat{\boldsymbol{\theta}}), w_{t-3,\tau}(\hat{\boldsymbol{\theta}})\varepsilon_{t-3,\tau}(\hat{\boldsymbol{\theta}}), w_{t-4,\tau}(\hat{\boldsymbol{\theta}})\varepsilon_{t-4,\tau}(\hat{\boldsymbol{\theta}}))^\top$.

Table 5. Finite-sample rejection rates for null and alternative model: out-of-sample test.

	$\gamma=0$				$\gamma=1$			
τ	**0.25**	**0.10**	**0.05**	**0.01**	**0.25**	**0.10**	**0.05**	**0.01**
				Setting A				
$T=1500$	0.0592	0.0624	0.0655	0.1133	1.0000	1.0000	1.0000	0.8065
$T=3000$	0.0535	0.0570	0.0652	0.0921	1.0000	1.0000	1.0000	0.9754
				Setting B				
$T=1500$	0.0528	0.0647	0.0732	0.1482	1.0000	1.0000	1.0000	0.9169
$T=3000$	0.0530	0.0550	0.0659	0.1119	1.0000	1.0000	1.0000	0.9978
				Setting C				
$T=1500$	0.0543	0.0624	0.0764	0.1845	1.0000	1.0000	1.0000	0.9591
$T=3000$	0.0554	0.0599	0.0667	0.1325	1.0000	1.0000	1.0000	0.9988
				Setting D				
$T=1500$	0.0566	0.0629	0.0851	0.2165	1.0000	1.0000	1.0000	0.9710
$T=3000$	0.0558	0.0582	0.0687	0.1429	1.0000	1.0000	1.0000	0.9993

Note: Settings A–D are defined as same as in Table 4.

4 Assessing Tail Risk for the S&P500 Index Using GCARE Models

4.1 Data

To illustrate the practical usefulness of our proposed model and its estimation procedure, we apply the GCARE model to estimate tail risks for the S&P500 index daily data from March 29, 2018 to December 31, 2022, with 1,200 observations in total. The data are downloaded from CSMAR Database, and the daily returns are computed as the difference of the log transformation of the index multiplying 100; that is, $Y_t = \log(p_t/p_{t-1}) * 100$, where p_t is the daily index. Table 6 presents the summary statistics of the return series. Although the average return is positive, the S&P500 return series is negatively skewed with fat tails, which provides motivation to use expectile to model tail risks. Figure 3 presents the histogram (left panel) and time series plot (right panel) of the return series. The histogram clearly shows that S&P500 return series has a high peak and fat tail.

Moreover, the time series plot shows that it is relatively stable in the period from 2018 to 2019 but becomes more volatile at the beginning of 2020, the period of the outbreak of the COVID-19 epidemic.

4.2 Empirical Results

To model the aforementioned financial data, Kuan, Yeh, and Hsu (2009) proposed the ABS(p) model

$$e_{t,\tau} = a_{0,\tau} + \sum_{i=1}^{p} \delta_{i,\tau} Y_{t-i}^{+} + \sum_{i=1}^{p} \lambda_{i,\tau} Y_{t-i}^{-},$$

and the SQ(p) model,

$$e_{t,\tau} = a_{0,\tau} + a_{1,\tau} Y_{t-1} + \sum_{i=1}^{p} b_{i,\tau} \left(Y_{t-i}^{+}\right)^2 + \sum_{i=1}^{p} \gamma_{i,\tau} (Y_{t-i}^{-})^2,$$

which have the ability to capture asymmetric properties in the tail risk for financial data. For ABS(p) model and SQ(p) model, we set a maximum value of $p = 5$ for the application of the EBIC in model selection. To compare the relative performance of in-sample fitting and out-of-sample predictability with ABS(p) model and SQ(p) model, two corresponding GCARE type models are proposed. To be specific, the GABS(p, q) model is given by

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

and the GSQ(p, q) model,

$$e_{t,\tau} = \alpha_{0,\tau} + \alpha_{1,\tau} Y_{t-1} + \sum_{i=1}^{p} b_{i,\tau} \left(Y_{t-i}^{+}\right)^2 + \sum_{i=1}^{p} \gamma_{i,\tau} (Y_{t-i}^{-})^2 + \sum_{j=1}^{q} \beta_{j,\tau} e_{t-j,\tau}.$$

Note that both GABS(p, q) model and the GSQ(p, q) model are special cases of GCARE(p, q) by selecting particular p, q and $\mathbf{X}_t$.

Before implementing the aforementioned models and related tests, the construction of instruments $\mathbf{Z}_t(\hat{\boldsymbol{\theta}})$ in the DE test should be firstly addressed. We follow the empirical strategy as in Engle and Manganelli (2004) and Taylor (2008). To be specific, three lags of $w_{t,\tau}(\hat{\boldsymbol{\theta}})\varepsilon_{t,\tau}(\hat{\boldsymbol{\theta}})$ are included in $\mathbf{Z}_t(\hat{\boldsymbol{\theta}})$ for the in-sample DE test, while a constant and expectile

Table 6. Summary statistics of return series.

Mean	Min	Median	Max	SD	Skew.	Kurt.
0.0306	– 12.7652	0.0992	8.9683	1.3907	– 0.7735	13.0873

Note: Sample period: March 29, 2018 to December 31, 2022.

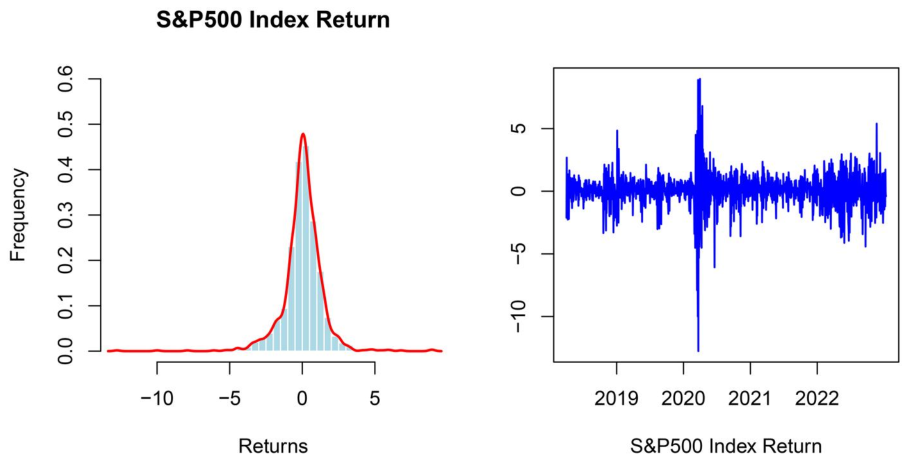

Figure 3. Time series and histogram plot of stock return series: S&P500.

forecast are added into the $\mathbf{Z}_t(\hat{\boldsymbol{\theta}})$ in the out-of-sample DE test. Note that the in-sample data is from March 29, 2018 to March 16, 2022, with 1000 observations in total. The subsequent out-of-sample period consists of 200 data points, which is from March 17, 2022 to December 31, 2022.

Table 7 reports estimation and testing results of four models under different values of τ. Generally speaking, compared to the CARE type models, the GCARE type models are more parsimonious in terms of the number of included parameters and with better out-of-sample performances. For instance, when $\tau = 0.05$, the EBIC leads to the ABS(4) and SQ(5) models which need to estimate 9 and 12 parameters, respectively. However, the EBIC chooses more parsimonious GCARE models. Although the GCARE(1,1) model is adequate to capture the evolution over time at a probability level of $\tau = 0.05$, under $\tau = 0.01$, the EBIC leads to the GABS(2,2) model and GSQ(3,2) model can also be observed in Table 7. The in-sample DE test cannot reject any of the four models under different τ values. When $\tau = 0.05$, all GCARE type models pass the out-of-sample DE test, whereas the ABS(4) model is strongly rejected. However, at the extreme case that the value of τ is 0.01, none of the models meet the criteria of the out-of-sample DE test, a result attributable to the size distortion, as illustrated in simulation Example 3.

Now, we look into the estimation of these coefficients in all these dynamic expectile models. For the GCARE type models, the coefficients of all lagged $e_{t,\tau}$ are highly significant which demonstrate a strong correlation over time in the tail return distribution. In the GABS(1,1) and GSQ(1,1) models, the estimates of the first-order lagged expectiles, $e_{t-1,\tau}$, are both positive, with values of 0.65 and 0.57, respectively. When two lagged expectiles are included, such as in GABS(2,2) and GSQ(3,2), the coefficients of first-order and second-order lagged expectiles are all significant. Indeed, the first-order lagged expectiles are positively significant for GABS(2,2) and GSQ(3,2) model, while the second-order

Table 7. Estimates and relevant statistics for different expectile models.

	$\tau=0.01$			$\tau=0.05$			$\tau=0.01$			$\tau=0.05$	
	ABS(4)	GABS(2,2)	GABS(1,1)	ABS(4)	GABS(1,1)		SQ(5)	GSQ(3,2)	GSQ(1,1)	SQ(5)	GSQ(1,1)
Cons	−2.2647***	−0.7960***	−0.3512***	−1.3357***	−0.1376**	Cons	−2.2672***	−1.0097***	−0.2560	−1.3558***	−0.1980
(SD)	(0.2399)	(0.2821)	(0.1099)	(0.1166)	(0.0666)	(SD)	(0.1843)	(0.2213)	(0.1575)	(0.1196)	(0.1463)
Y^+_{t-1}	0.2331	0.2096	−0.3335**	−0.0260	−0.3405	Y_{t-1}	0.4790**	−0.3931**	−0.6491***	0.3851**	−0.4566***
(SD)	(0.4695)	(0.2231)	(0.1304)	(0.3771)	(0.2412)	(SD)	(0.2229)	(0.1938)	(0.2117)	(0.1603)	(0.1766)
Y^-_{t-1}	−0.1388	−0.1108	−0.6544***	−0.0977	−0.4705***	$(Y^+_{t-1})^2$	−0.0658	0.0927***	0.0089	−0.0740	−0.0407
(SD)	(0.1649)	(0.1590)	(0.1021)	(0.2212)	(0.1473)	(SD)	(0.0942)	(0.0305)	(0.0318)	(0.0759)	(0.0487)
Y^+_{t-2}	−0.1182	−0.4702***		−0.2041		$(Y^-_{t-1})^2$	0.1253**	0.0736	0.0225	0.1142**	0.0239
(SD)	(0.3102)	(0.1717)		(0.2731)		(SD)	(0.0624)	(0.0545)	(0.0493)	(0.0558)	(0.0365)
Y^-_{t-2}	−0.7301	−0.8464***		−0.6517		$(Y^+_{t-2})^2$	0.0479	0.0521		0.0258	
(SD)	(0.4653)	(0.2083)		(0.4305)		(SD)	(0.0722)	(0.0580)		(0.0588)	
Y^+_{t-3}	−0.4027			−0.2831		$(Y^-_{t-2})^2$	−0.0652	0.1417***		−0.0626	
(SD)	(0.6104)			(0.2549)		(SD)	(0.0563)	(0.0461)		(0.0433)	
Y^-_{t-3}	−0.7392			−0.4969		$(Y^+_{t-3})^2$	−0.1096	−0.0051		−0.0617	
(SD)	(1.0437)			(0.5285)		(SD)	(0.0800)	(0.0483)		(0.0700)	
Y^+_{t-4}	−0.4614			−0.1759		$(Y^-_{t-3})^2$	−0.1488	0.0188		−0.1154	
(SD)	(0.6056)			(0.3013)		(SD)	(0.1369)	(0.0596)		(0.1338)	
Y^-_{t-4}	−0.1747			0.0589		$(Y^+_{t-4})^2$	−0.0608			−0.0232	
(SD)	(0.2340)			(0.2136)		(SD)	(0.1001)			(0.0726)	
Y^+_{t-5}						$(Y^-_{t-4})^2$	0.0392			0.0258	

(continued)

Table 7. (continued)

	$\tau=0.01$			$\tau=0.05$			$\tau=0.01$			$\tau=0.05$	
	ABS(4)	GABS(2,2)	GABS(1,1)	ABS(4)	GABS(1,1)		SQ(5)	GSQ(3,2)	GSQ(1,1)	SQ(5)	GSQ(1,1)
(SD)						(SD)	(0.0252)			(0.0282)	
Y^-_{t-5}						$(Y^+_{t-5})^2$	−0.0757			−0.0459	
(SD)						(SD)	(0.1195)			(0.0486)	
						$(Y^-_{t-5})^2$	0.0315			0.0128	
						(SD)	(0.0754)			(0.0673)	
e_{t-1}		0.8330***	0.6677***		0.6477***	e_{t-1}		0.7949***	0.6686***		0.5689***
(SD)		(0.1769)	(0.0547)		(0.0852)	(SD)		(0.1547)	(0.0420)		(0.0693)
e_{t-2}		−0.5259***				e_{t-2}		−0.5309***			
(SD)		(0.1537)				(SD)		(0.0993)			
DE in-sample						DE in-sample					
P-value	0.9714	0.9935	0.1207	0.9841	0.2580	*P*-value	0.9966	0.9022	0.2051	0.8443	0.4492
DE out-of-sample						DE out-of-sample					
P-value	0.0000***	0.0031***	0.0001***	0.0000***	0.5957	*P*-value	0.0035***	0.0138**	0.0002***	0.1139	0.7681

Note: ** and *** denote the significance at 5% and 1%, respectively.

In-sample period: March 29, 2018–March 16, 2022. Out-of-sample period: March 17, 2022–December 31, 2022.

ABS(p): $e_{t,\tau}=a_{0,\tau}+\sum_{i=1}^{p}\delta_{i,\tau}Y^+_{t-i}+\sum_{i=1}^{p}\lambda_{i,\tau}Y^-_{t-i}$.

SQ(p): $e_{t,\tau}=a_{0,\tau}+a_{1,\tau}Y_{t-1}+\sum_{i=1}^{p}b_{i,\tau}(Y^+_{t-i})^2+\sum_{i=1}^{p}\gamma_{i,\tau}(Y^-_{t-i})^2$.

GABS(p, q): $e_{t,\tau}=\alpha_{0,\tau}+\sum_{i=1}^{p}\delta_{i,\tau}Y^+_{t-i}+\sum_{i=1}^{p}\lambda_{i,\tau}Y^-_{t-i}+\sum_{j=1}^{q}\beta_{j,\tau}e_{t-j,\tau}$.

$GSQ(p,q)$: $e_{t,\tau}=\alpha_{0,\tau}+\alpha_{1,\tau}Y_{t-1}+\sum_{i=1}^{p}b_{i,\tau}(Y^+_{t-i})^2+\sum_{i=1}^{p}\gamma_{i,\tau}(Y^-_{t-i})^2+\sum_{j=1}^{q}\beta_{j,\tau}e_{t-j,\tau}$.

Null hypothesis: $E[w_{t,\tau}(\boldsymbol{\theta})\varepsilon_{t,\tau}|\mathbf{W}_t]=0$, Alternative hypothesis: $E[w_{t,\tau}(\boldsymbol{\theta})\varepsilon_{t,\tau}|\mathbf{W}_t]\neq 0$.

lagged expectiles are negatively significant for these two models. For SQ and GSQ models, the coefficients of Y_{t-1} are all significant, but positive in SQ models while negative in GSQ models. The underlying rationale is that the incorporation of lagged expectiles may exert an influence on the magnitude of the coefficients associated with Y_{t-1}. The coefficients of $(Y_{t-1}^{-})^2$ in all SQ models are positively significant. In the GSQ(3,2) model, the coefficients of $(Y_{t-1}^{+})^2$ is positively significant when $\tau = 0.01$. In the GABS(2,2) model, the coefficients of Y_{t-2}^{+} and Y_{t-2}^{-} are negatively significant when $\tau = 0.01$, with an asymmetric effect in the magnitude. In the GABS(1,1) model, the coefficient of Y_{t-1}^{-} is negatively significant when $\tau = 0.05$, indicating that recent negative returns are likely to depress the conditional expectile. However, most coefficients, except the coefficients of the constant term, in ABS and SQ models are insignificant maybe due to the fact of including too many irrelevant lagged terms. Generally speaking, the estimation results demonstrate a relatively strong impact from lagged negative returns on tail risks.

4.3 Time Invariant Portfolio Protection Strategy

Next, we apply the proposed GCARE model to the popular proportion portfolio insurance strategies, and compare its performances to other competing risk measure models. We consider a portfolio invested into the S&P 500 index and into the well-recognized risk-less asset, the daily treasury bill rates of one year.[2] The sample data are split into two subsamples: the in-sample data is from April 9, 2014 to March 28, 2018, with 1000 observations in total; the out-of-sample period consists of 1200 data, which is from March 29, 2018 to December 31, 2022. The proportion portfolio strategies need to determine a so called multiple, a proportion invested into risky assets, and the remaining is invested into risk-less assets to guarantee a minimum value of the portfolio over a given investment horizon, which is called a floor in the literature. Obviously, the multiple represents the portfolio's risk exposure. Many conditional tail risk models can be used to determine the multiple in a portfolio. Among various proportion portfolio strategies, we consider the recently developed TIPP strategy.

Let us first introduce the TIPP strategy. Denote V_t the value of the covered portfolio at time $t \in (0, T]$, and suppose it's initial value $V_0 = 100$. The minimum value of the portfolio that is acceptable for an investor at any time t, the floor, is defined as

$$F_t \equiv \max\left(\nu F \cdot e^{-rf_t \cdot (T-t)}, \nu \sup_{s \le t} V_s\right) \le V_t,$$

where F is the guarantee at the maturity T and we set $F = 105.77$ considering the asset appreciation with expected annualized risk-free rate of 1.4533%,[3] rf_t is the risk-free rate, $\nu \in (0, 1)$ is defined as the floor percentage. In TIPP strategy, an 80% floor means that the trading program will be constructed so that the portfolio can never decline below 80% of the highest value it ever reached. As suggested by Hamidi, Maillet, and Prigent (2014), ν is usually defined as equal to 90%. The excess of the portfolio value above the floor, $C_t = V_t - F_t$, is defined as the cushion value. In the strategy, the amount of $G_t = mC_t$ with m denoting the multiple invested into the risky asset with return R_t, and the remaining amount of the portfolio value is invested into the risk-less asset. The performance of the strategy is completely determined by the multiple.

Assume a discrete-time trading between two rebalancing times t and $t+1$. According to the strategy introduced, the portfolio value evolves as follows: $V_{t+1} = V_t + G_t R_{t+1} + (V_t - G_t) rf_{t+1}$ and the cushion value satisfies $C_{t+1} = C_t\{1 + mR_{t+1} + (1-m) rf_{t+1}\}$. The

[2] It is downloaded from https://home.treasury.gov/policy-issues/financing-the-government/interest-rate-statistics.

[3] It is calculated using the average return of daily treasury bill rates of one year from April 10, 2013 to April 9, 2014.

above equation proves that, if the guarantee holds at the time t, it will also hold at time $t+1$ provided that the term $(1+mR_{t+1}+(1-m)rf_{t+1})$ is positive. Assume that $(R_{t+1}-rf_{t+1})$ may be non-positive and rf_t is negligible in a very short period, the upper bound of multiple m is easily derived with

$$m \leq (-R_{t+1}^{-})^{-1},$$

where $R_{t+1}^{-} = \min(0, R_{t+1})$, which indicates how the multiple reacts to the market conditions. Using a quantile hedging method, Jiang, Ma, and An (2009) proposed to compute the updated multiple using the tail risk measure VaR, namely,

$$m_t = |VaR_{\eta,t}|^{-1},$$

where $VaR_{\eta,t}$ is the ηth conditional VaR of the risk asset R_t. Many dynamic quantile models can be used in the quantile hedging approach, such as the CAViaR model. However, since the VaR has recently been criticized for its insensitiveness to the magnitude of extreme losses at tails, the ES was introduced by Hamidi, Maillet, and Prigent (2014) to compute the conditional multiple such as

$$m_t = |ES_{\eta,t}|^{-1},$$

where $ES_{\eta,t}$ is the ηth conditional ES of the risk asset R_t. Then using the one-to-one link between expectiles and ES illustrated in Section 1.1, we can apply the CARE and GCARE type models to compute the conditional multiples.

To be specific, the four dynamic expectile models considered in section 4.2 are employed to establish the ES based TIPP strategy. The probability level η of $VaR_{\eta,t}$ and $ES_{\eta,t}$ is selected as 0.01 as in Hamidi, Maillet, and Prigent (2014). To compute a corresponding τ for the given η, a linear interpolation technique as in Taylor (2008) is conducted. Further, the EBIC leads to ABS(5) model, SQ(5) model, GABS(1,2) model, and GSQ(3,1) model, respectively. For the aim of comparison, the historical simulation method using moving windows of length 250, 500, and 1000 days together with the RiskMetrics are implemented as the benchmark models. Further, we also consider a quantile based TIPP strategy based on the CAViaR(p, q) model, which has the form

$$q_{t,\eta} = \alpha_{0,\eta} + \sum_{i=1}^{p} \alpha_{i,\eta} Y_{t-i}^{+} + \sum_{i=1}^{p} \gamma_{i,\eta} Y_{t-i}^{+} + \sum_{j=1}^{q} \beta_j q_{t-j,\eta},$$

and the EBIC leads to a CAViaR(2,2) model. To overcome the potential look-ahead bias, m is estimated using the day-ahead post-sample predictions of the aforementioned models on a rolling window of length 1000. In practice, we follow Xu, Mihoci, and Härdle (2018) to truncate the multiple at $m = 12$ and $m = 1$ for all TIPP strategies. To elaborate, the value of m is adjusted to $m = 12$ or $m = 1$ when it exceeds 12 or falls below 1, respectively.

As suggested by one of our anonymous referees, we have also introduced a mean-variance (M-V afterwards) strategy for comparative analysis with the TIPP strategy. In this comparative framework, the S&P 500 index and the treasury rate are designated as the assets under consideration. This re-estimation occurs on a rolling window with a length of 1000 observations. Specifically, the efficient portfolio targeting an expected annualized return of 1.4533% is selected. The portfolio weights for the mean-variance strategy are determined based on the last 1000 observations from the dataset.

Figure 4 depicts the path of portfolio values based on different TIPP strategies and M-V strategy during the investment horizon. We use blue, aquamarine, yellow, pink, black, red, green, lightcyan, orange, lightblue and tan lines to denote the time series of portfolio values based on the S&P500 index, M-V strategy, the ABS(5) model, the SQ(5) model, the GABS (1,2) model, the GSQ(3,1) model, the GABS(1,1) model, the GSQ(1,1) model, historical simulation method with window size 1000, RiskMetrics and the CAViaR(2,2) model, respectively. Note that the performance of each portfolio mainly depends on its performance under extreme market conditions. The occurrence of extreme losses may destroy the portfolio when a model cannot capture the risk precisely and therefore obtains an inappropriate multiple. From Figure 4, one can also see that all of the dynamic multiple-based cushioned portfolios guarantee the target floor at every trading day during the investment horizon even when large market downturn occurs at the beginning of 2020, which is more clearly shown in Figure 5. Figure 5 depicts the heat map of the multiple values during the investment horizon, and we can see that the multiple values of all the portfolios drop to a safe value (near 1) quickly during the large downturn in the investment horizon (January 2020–March 2020).

In general, the risk-sensitive portfolio can take the chance to make a fortune in market jumps. However, this approach inherently involves a trade-off. To curtail exposure to downside risk, such portfolios are often calibrated to favor lower multiples. As a result, they may not track the index closely, particularly during periods of bullish trends, thus potentially forgoing some of the gains associated with bull markets. To see how the proposed model performs for the Covid-19 pandemic period,

Figure 6 gives the time series plots of the portfolio values during the Covid-19 pandemic period (January 2020.01–June 2020) and Figure 7 displays heatmap of the multiple values (January 2020.01– June 2020). From both Figures 6 and 7, one can see clearly that the ABS(5) model, GABS(1,2) model, GSQ(3,1) model and CAViaR(2,2) model have higher multiples when the market is recovering from the large downturn, which leads to higher returns for these models. It can also be observed from Figure 6 that the portfolio based on SQ(5) model drop to the guarantee during the large downturn at the beginning of March, 2020. Note that the portfolios based on HS1000 and RiskMetrics have relatively stable multiple values in the specific period.

Finally, Table 8 summarizes the descriptive statistics of the portfolio returns of different TIPP strategies together with the M-V strategy. First, we use the difference of the log transformation of the portfolio values to compute the daily returns.

Then, the annualized return, annualized volatility and annualized Sharpe ratio are reported in Table 8, from which, one can see that all the cushioned portfolios have lower volatility than S&P500 series. Among all the portfolio insurance strategies, the GABS(1,2) model exhibits the best performance. The annualized return of GABS(1,2) model is 3.99% with the annualized Sharpe ratio 0.3982. It is also observed that the GABS(1,2) model and GSQ(3,1) model based strategies perform relatively better than the corresponding GABS (1,1) model and GSQ(1,1) model. Meanwhile, the M-V strategy consistently favor holding the treasury rate, so that it has the lowest volatility. Note that the SQ(5) model based strategy performs worst. The annualized return and annualized Sharpe ratio of the SQ(5) model are 0.78% and -0.1318, respectively. The empirical analysis results suggest the usefulness of our GCARE model for the given dataset.

5 Conclusion

In this article, we introduce a class of generalized conditional autoregressive expectile (GCARE) models which include the expectile autoregressive components. An EBIC is proposed to select the number of lagged expectiles and covariates. The QML estimation

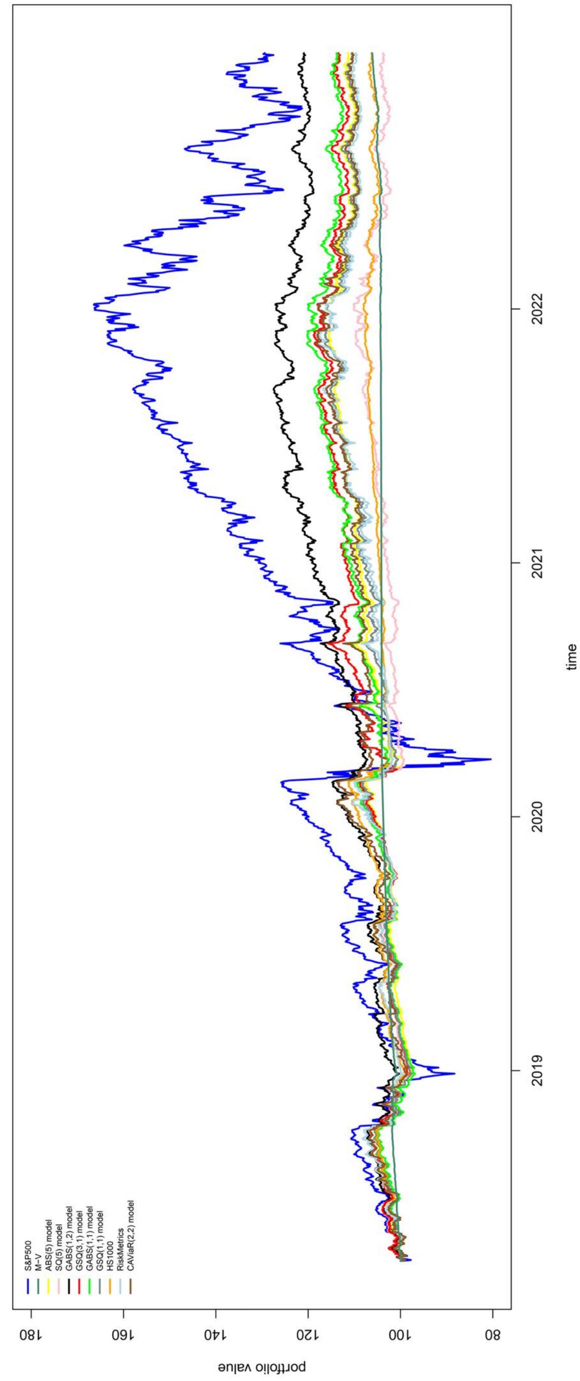

Figure 4. Path of the portfolio values during the investment horizon.

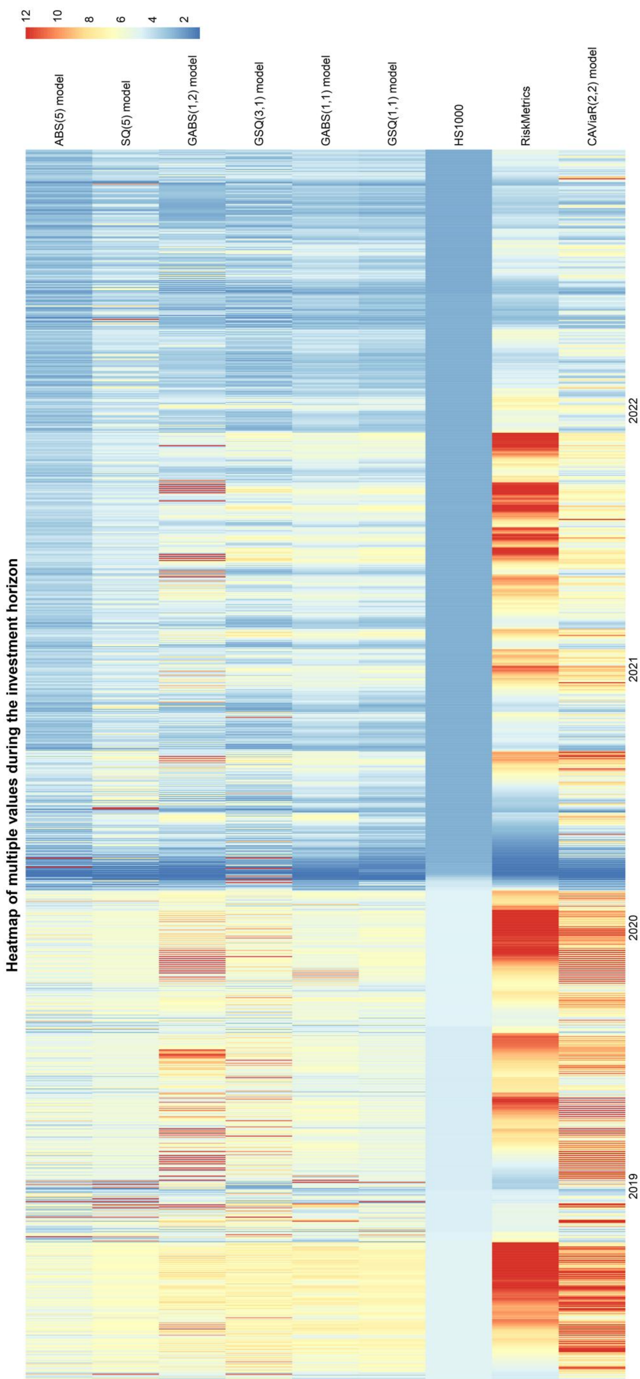

Figure 5. Heatmap of the multiple values during the investment horizon.

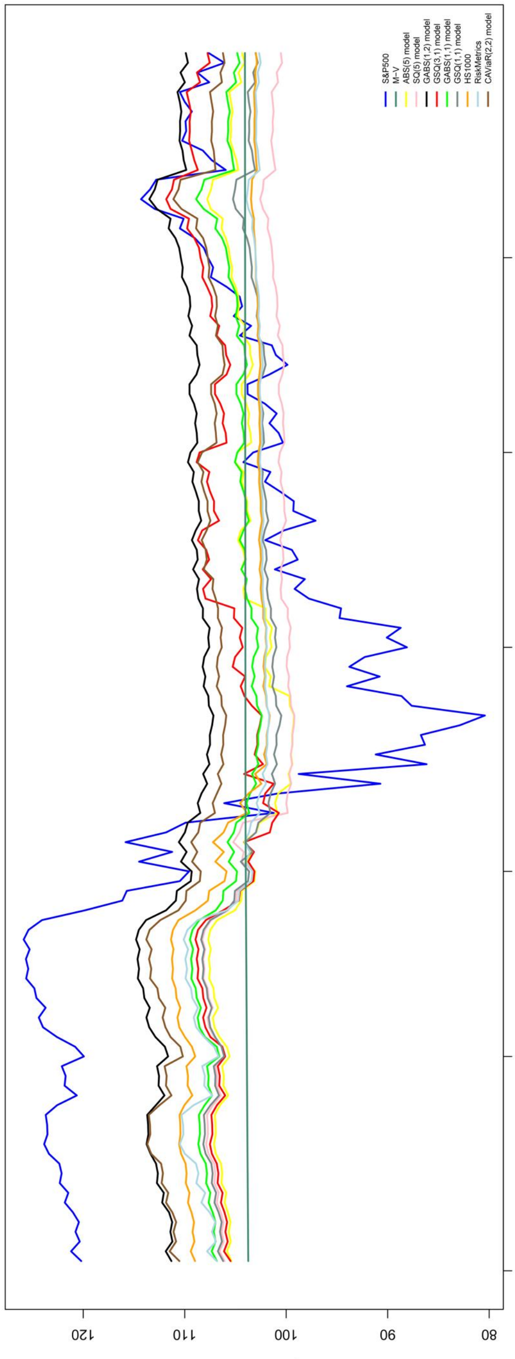

Figure 6. Time series plot of the portfolio values during the COVID-19 pandemic period (January 2020–June 2020).

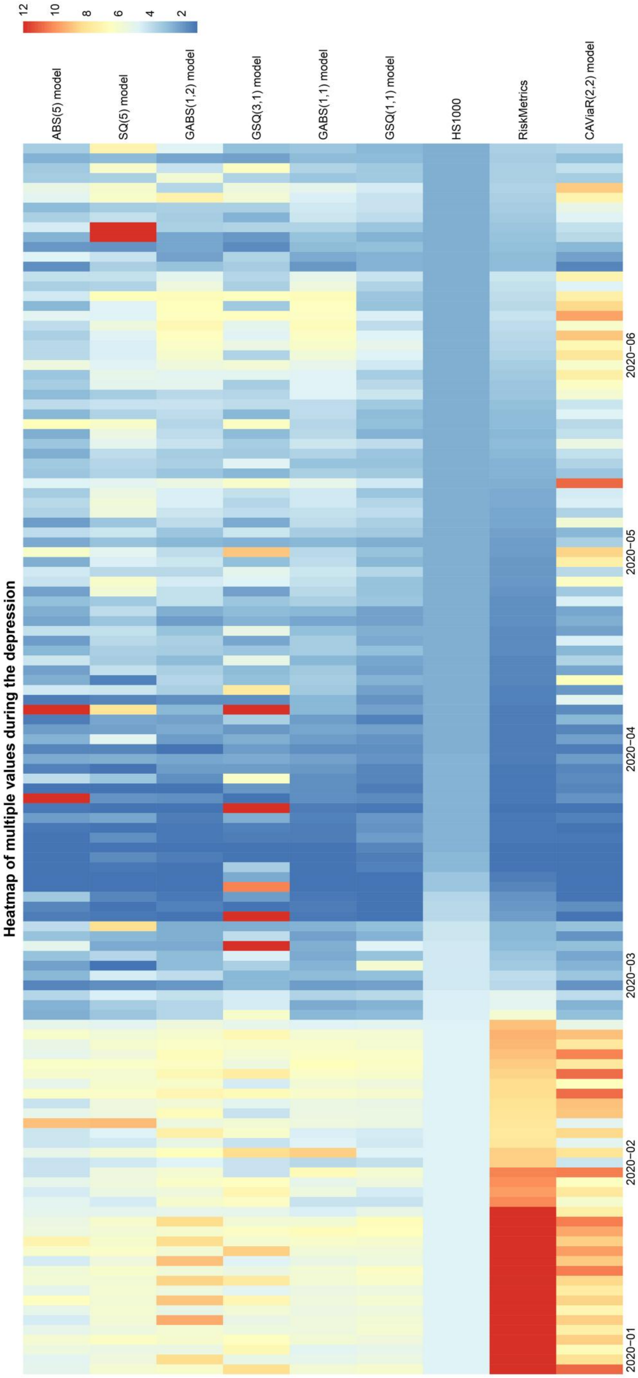

Figure 7. Heatmap of the multiple values during the COVID-19 pandemic period (January 2020–June 2020).

Table 8. Portfolio performance based on the TIPP strategy.

	Return (%)	Volatility (%)	Sharpe
S&P500	5.4147	22.1974	0.1701
M-V	1.2484	0.6771	−0.3845
HS250	1.0957	3.9387	−0.1049
HS500	0.8602	3.9855	−0.1627
HS1000	1.2784	4.1460	−0.0555
RiskMetrics	1.9480	6.7491	0.0651
ABS(5)	2.2684	5.3956	0.1408
SQ(5)	0.7795	5.5310	−0.1318
GABS(1,2)	3.9900	6.2305	0.3982
GABS(1,1)	2.7226	5.7899	0.2097
GSQ(3,1)	2.6482	6.4427	0.1769
GSQ(1,1)	2.1117	5.5242	0.1092
CAViaR(2,2)	2.0739	6.7604	0.0836

Note: In-sample period: April 9, 2014 to March 28, 2018. Out-of-sample period: March 29, 2018 to December 31, 2022.
M-V: Mean-Variance strategy with expected annualized return 1.4533%.
HS250, HS500, HS1000: Historical simulation utilizing moving windows of lengths 250, 500, and 1000 days.
RiskMetrics: The conventional RiskMetrics method based on Gaussian IGARCH(1,1) model.
CAViaR(p, q): $q_{t,\eta} = \alpha_{0,\eta} + \sum_{i=1}^{p} \alpha_{i,\eta} Y_{t-i}^{+} + \sum_{i=1}^{p} \gamma_{i,\eta} Y_{t-i}^{+} + \sum_{j=1}^{q} \beta_j q_{t-j,\eta}$.
ABS(p), SQ(p), GABS(p, q), and GSQ(p, q) are same as in Table 6.

method is employed to estimate the proposed model and the consistency and asymptotic normality of the QMLEs are established. In addition, dynamic expectile tests are constructed to evaluate in-sample model goodness of fit and out-of-sample predictive ability. Finally, the proposed GCARE models are applied to the TIPP strategy by determining the portfolio's risk exposure. The empirical results demonstrate that the GCARE model outperforms other existing models with the highest Sharpe values.

Supplemental Material

Supplemental material is available at *Journal of Financial Econometrics* online.

Funding

The authors gratefully acknowledge the financial supports, in part, from the National Science Fund of China (NSFC) with grants #72103207 and the key project grants #72033008 and #72133002, Basic Science Center Program of NSFC with grant #71988101, and "the Fundamental Research Funds for the Central Universities", Zhongnan University of Economics and Law with grant #2722023BY018.

Conflict of interest

We confirm that this work is original and has not been published elsewhere, nor is it currently under consideration for publication elsewhere. Also, we declare that we have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper. We have no conflicts of interest to disclose. Finally, we declare that we do not use any generative AI and AI-assisted technologies in the writing process.

Appendix: Mathematical Proofs

Note that the detailed proofs of main theorems are presented in this appendix with the detailed proofs of some lemmas given in the Supplementary material.

Proof of Theorem 1: Let $\mathrm{L}_{0,\tau}(\boldsymbol{\theta}) \equiv E\big(\mathrm{L}_{T,\tau}(\boldsymbol{\theta})\big) = T^{-1}\sum_{t=1}^{T} E\big(l_{t,\tau}(\boldsymbol{\theta})\big)$, and $R_{t,\tau}(\boldsymbol{\theta}) = E\big(l_{t,\tau}(\boldsymbol{\theta})|\mathbf{W}_t\big)$ with $l_{t,\tau}(\boldsymbol{\theta}) \equiv w_{t,\tau}(\boldsymbol{\theta})\varepsilon_{t,\tau}^2$. Then we have the following lemma.

Lemma A1. *Let $\boldsymbol{\theta}_0$ be a interior point in Θ. For $\boldsymbol{\theta} \in \Theta$ and a large enough T, if*

i) $\sup_{\boldsymbol{\theta}\in\Theta} |\mathrm{L}_{T,\tau}(\boldsymbol{\theta}) - \mathrm{L}_{0,\tau}(\boldsymbol{\theta})| \xrightarrow{\mathcal{P}} 0$;
ii) $\mathrm{L}_{0,\tau}(\boldsymbol{\theta})$ *has a unique minimum at* $\boldsymbol{\theta}_0$;
iii) $\mathrm{L}_{T,\tau}(\boldsymbol{\theta})$ *is convex in* $\boldsymbol{\theta}$,
then $\hat{\boldsymbol{\theta}}_T = \mathrm{argmin}_{\boldsymbol{\theta}\in\Theta}\mathrm{L}_{T,\tau}(\boldsymbol{\theta})$ *exists with probability approaching one and* $\hat{\boldsymbol{\theta}}_T \xrightarrow{\mathcal{P}} \boldsymbol{\theta}_0$.

In this proof, the conditions (i)–(iii) imposed in Lemma A1 are verified. For condition (i), the uniform convergence of $\mathrm{L}_{T,\tau}(\boldsymbol{\theta})$ to $\mathrm{L}_{0,\tau}(\boldsymbol{\theta})$ is checked here. We know from Assumption A1 that $\{Y_t, \mathbf{X}_t\}_{t\in\mathbb{Z}}$ is α-mixing process, then it is also ergodic by Proposition 3.44 in White (2001). Since the sequence $l_{t,\tau}(\boldsymbol{\theta}) \in H(\mathcal{F}_t)$ is measurable with respect to $\mathcal{F}_t$, by Theorem 3.35 in White (2001), it is easy to show that $l_{t,\tau}(\boldsymbol{\theta})$ is also an ergodic process. Now, for a given $r > 1$ and $\varepsilon > 0$, one has $E[|l_{t,\tau}(\boldsymbol{\theta})|^{r+\varepsilon/2}] \le CE[|\varepsilon_{t,\tau}|^{2r+\varepsilon}] < \infty$ according to Assumption A7. Here and throughout the appendix, C denotes a generic constant, which takes different values at different places. Then, by the weak law of large numbers for ergodic sequences of Theorem 3.34 in White (2001), one has

$$\mathrm{L}_{T,\tau}(\boldsymbol{\theta}) = T^{-1}\sum_{t=1}^{T} l_{t,\tau}(\boldsymbol{\theta}) \xrightarrow{\mathcal{P}} T^{-1}\sum_{t=1}^{T} E\big[l_{t,\tau}(\boldsymbol{\theta})\big] = \mathrm{L}_{0,\tau}(\boldsymbol{\theta}).$$

Further, as $l_{t,\tau}(\boldsymbol{\theta})$ is a convex function for every t, $1 \le t \le T$, which leads $\{\mathrm{L}_{T,\tau}(\boldsymbol{\theta})\}$ to be a sequence of random convex functions. For a compact set Θ, one has

$$\sup_{\boldsymbol{\theta}\in\Theta} |\mathrm{L}_{T,\tau}(\boldsymbol{\theta}) - \mathrm{L}_{0,\tau}(\boldsymbol{\theta})| \xrightarrow{\mathcal{P}} 0,$$

by Convexity Lemma in Pollard (1991), therefore condition (i) is verified.

To prove condition (ii) that $\mathrm{L}_{0,\tau}(\boldsymbol{\theta})$ is uniquely minimized at $\boldsymbol{\theta}_0$, it is easy to show that $\|\nabla l_{t,\tau}(\boldsymbol{\theta})\| \le 2\|\nabla e_{t,\tau}(\boldsymbol{\theta})\big(Y_t - e_{t,\tau}(\boldsymbol{\theta})\big)\|$, and it is dominated by an integrable function by Assumptions A8. Therefore, it follows that

$$\frac{\partial R_{t,\tau}(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}} = -2E\left[\nabla e_{t,\tau}(\boldsymbol{\theta})\left\{\tau\int_{e_{t,\tau}(\boldsymbol{\theta})}^{\infty}\big(y - e_{t,\tau}(\boldsymbol{\theta})\big)f(y|\mathbf{W}_t)dy + (1-\tau)\int_{-\infty}^{e_{t,\tau}(\boldsymbol{\theta})}\big(y - e_{t,\tau}(\boldsymbol{\theta})\big)f(y|\mathbf{W}_t)dy\right\}\right],$$

where $f(y|\mathbf{W}_t)$ is the conditional probability function of Y_t given $\mathbf{W}_t$. Given the fact that $\int_{\infty}^{z}(y-z)f(y|\mathbf{W}_t)dy$ is continuously differentiable in z, and its derivative $-\int_{-\infty}^{z} f(y|\mathbf{W}_t)dy$

is uniformly bounded by 1, we know that $\frac{\partial R_{t,\tau}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}$ is continuously differentiable. Therefore, one has

$$\begin{aligned}
&\frac{\partial^2 R_{t,\tau}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^\top} \\
&= 2E\left[\nabla e_{t,\tau}(\boldsymbol{\theta}) \nabla^\top e_{t,\tau}(\boldsymbol{\theta}) \left\{ \tau \int_{e_{t,\tau}(\boldsymbol{\theta})}^{\infty} f(y|\mathbf{W}_t)dy + (1-\tau) \int_{-\infty}^{e_{t,\tau}(\boldsymbol{\theta})} f(y|\mathbf{W}_t)dy \right\}\right] \\
&= 2E\left[w_{t,\tau}(\boldsymbol{\theta}) \nabla e_{t,\tau}(\boldsymbol{\theta}) \nabla^\top e_{t,\tau}(\boldsymbol{\theta})\right].
\end{aligned}$$

Let $\zeta = \min(\tau, 1-\tau)$, it follows that $\frac{\partial^2 R_{t,\tau}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^\top} - 2\zeta E\left[\nabla e_{t,\tau}(\boldsymbol{\theta}) \nabla^\top e_{t,\tau}(\boldsymbol{\theta})\right]$ is positive semi-definite by Assumption A6. For any $\boldsymbol{\theta} \in \Theta$, by using the second order mean value expansion of $R(\boldsymbol{\theta})$ around a fixed $\tilde{\boldsymbol{\theta}}$, one has

$$\begin{aligned}
R_{t,\tau}(\boldsymbol{\theta}) - R_{t,\tau}(\tilde{\boldsymbol{\theta}}) &= \frac{\partial R_{t,\tau}(\tilde{\boldsymbol{\theta}})}{\partial \boldsymbol{\theta}^\top}(\boldsymbol{\theta} - \tilde{\boldsymbol{\theta}}) + (\boldsymbol{\theta} - \tilde{\boldsymbol{\theta}})^\top \frac{\partial^2 R_{t,\tau}(\bar{\boldsymbol{\theta}})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^\top}(\boldsymbol{\theta} - \tilde{\boldsymbol{\theta}}) \\
&= \frac{\partial R_{t,\tau}(\tilde{\boldsymbol{\theta}})}{\partial \boldsymbol{\theta}^\top}(\boldsymbol{\theta} - \tilde{\boldsymbol{\theta}}) + 2(\boldsymbol{\theta} - \tilde{\boldsymbol{\theta}})^\top E\left[\nabla e_{t,\tau}(\bar{\boldsymbol{\theta}}) \nabla^\top e_{t,\tau}(\bar{\boldsymbol{\theta}}) w_{t,\tau}(\bar{\boldsymbol{\theta}})\right](\boldsymbol{\theta} - \tilde{\boldsymbol{\theta}}) \qquad \text{(A1)} \\
&\geq \frac{\partial R_{t,\tau}(\tilde{\boldsymbol{\theta}})}{\partial \boldsymbol{\theta}^\top}(\boldsymbol{\theta} - \tilde{\boldsymbol{\theta}}) + 2\zeta M_e \|\boldsymbol{\theta} - \tilde{\boldsymbol{\theta}}\|^2,
\end{aligned}$$

where $\bar{\boldsymbol{\theta}}$ lies between $\tilde{\boldsymbol{\theta}}$ and $\boldsymbol{\theta}$, and $M_e > 0$ is the minimum eigenvalue of $E[\nabla e_{t,\tau}(\bar{\boldsymbol{\theta}}) \nabla^\top e_{t,\tau}(\bar{\boldsymbol{\theta}}) w_{t,\tau}(\bar{\boldsymbol{\theta}})]$. By the above inequality, it is readily seen that $R_{t,\tau}(\boldsymbol{\theta}) > R_{t,\tau}(\tilde{\boldsymbol{\theta}})$ for $\boldsymbol{\theta}$ outside some closed ball centered at $\tilde{\boldsymbol{\theta}}$. By the continuity of $R_{t,\tau}(\boldsymbol{\theta})$, it has a minimum $\boldsymbol{\theta}_0$ inside this ball, then $\boldsymbol{\theta}_0$ is a global minimum given that $R_{t,\tau}(\boldsymbol{\theta}_0) \leq R_{t,\tau}(\tilde{\boldsymbol{\theta}})$. Moreover, with Assumption A5 one has $\frac{\partial R_{t,\tau}(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}} = 0$. Then, by the convexity of $R_{t,\tau}(\boldsymbol{\theta})$, the fact that $\boldsymbol{\theta}_0$ is a unique minimum of $R_{t,\tau}(\boldsymbol{\theta})$ follows from Equation (A1) with $\tilde{\boldsymbol{\theta}} = \boldsymbol{\theta}_0$. Finally, Condition (iii) is automatically satisfied given the convexity of $l_{t,\tau}(\boldsymbol{\theta})$ for every t, $1 \leq t \leq T$, and the consistency of $\hat{\boldsymbol{\theta}}_T$ to $\boldsymbol{\theta}_0$ is, therefore, established.

Proof of Theorem 2: Define $\hat{\xi} = \sqrt{T}(\hat{\boldsymbol{\theta}}_T - \boldsymbol{\theta}_0)$, then $\hat{\xi}$ minimizes the function

$$G_T(\xi) = \sum_{t=1}^{T} \left\{ Q_\tau\left(\varepsilon_{t,\tau}^* - T^{-1/2} \xi^\top \nabla e_{t,\tau}(\boldsymbol{\theta}_0)\right) - Q_\tau\left(\varepsilon_{t,\tau}^*\right) \right\},$$

where $\varepsilon_{t,\tau}^* = Y_t - e_{t,\tau}(\boldsymbol{\theta}_0)$, $G_T(\xi)$ is a convex function in ξ.

Lemma A2. *Under Assumptions A1-A6 and B, when $T \to \infty$, one has*

$$G_T(\xi) = \xi^\top \mathbf{D}_T(\boldsymbol{\theta}_0)\xi + \frac{2}{\sqrt{T}} \sum_{t=1}^{T} h_{t,\tau}^\top(\boldsymbol{\theta}_0)\xi + R_T(\xi),$$

where $\sup_{\xi \in \mathcal{R}} |R_T(\xi)| = o_p(1)$ *for any compact set* $\mathcal{R}$.

From Lemma A2 and the convexity lemma in Pollard (1991), $\hat{\xi}$ can be explicitly expressed as

$$\hat{\xi} = -T^{-1/2}\mathbf{D}_T^{-1}(\boldsymbol{\theta}_0)\sum_{t=1}^{T} h_{t,\tau}(\boldsymbol{\theta}_0) + o_p(1),$$

which is equivalent to

$$\sqrt{T}\left(\hat{\boldsymbol{\theta}}_T - \boldsymbol{\theta}_0\right) = -T^{-1/2}\mathbf{D}_T^{-1}(\boldsymbol{\theta}_0)\sum_{t=1}^{T} h_{t,\tau}(\boldsymbol{\theta}_0) + o_p(1). \tag{A2}$$

It is easy to show that $h_{t,\tau}(\boldsymbol{\theta}_0) \in H(\mathcal{F}_t)$ is a vector of random variables which is measurable with respect to $\mathcal{F}_t$, and by Assumption B1, it satisfies that $E(\|w_{t,\tau}(\boldsymbol{\theta}_0)\varepsilon_{t,\tau}^{*}\nabla e_{t,\tau}(\boldsymbol{\theta}_0)\|^{2r+\varepsilon}) < \infty$, for some $\varepsilon > 0$. At the same time, one has $E(h_{t,\tau}(\boldsymbol{\theta}_0)) = E[\nabla e_{t,\tau}(\boldsymbol{\theta}_0)E(w_{t,\tau}(\boldsymbol{\theta}_0)\varepsilon_{t,\tau}^{*}|\mathbf{W}_t)] = 0$ for the first order condition. Together with Assumptions B2 and B3, the central limit theory for functionals of mixing sequences (see Theorem 18.6.2 in Ibragimov and Linnik (1971)) can be applied, and one has

$$T^{-1/2}\sum_{t=1}^{T} h_{t,\tau}(\boldsymbol{\theta}_0) \xrightarrow{\mathcal{L}} N(0, \mathbf{V}_T(\boldsymbol{\theta}_0)),$$

where

$$\mathbf{V}_T(\boldsymbol{\theta}_0) = \mathrm{Var}\left(T^{-1/2}\sum_{t=1}^{T} h_{t,\tau}(\boldsymbol{\theta}_0)\right) = T^{-1}\sum_{t=1}^{T}\sum_{s=1}^{T} E\left[h_{t,\tau}(\boldsymbol{\theta}_0)h_{s,\tau}^{\top}(\boldsymbol{\theta}_0)\right],$$

which, together with Equation (A2), concludes that the proof is completed.

Proof of Theorem 3: We will split the proof into two parts. For the consistency of $\mathbf{D}_T$, one has the following lemma.

Lemma A3. *Suppose Assumptions A1–A6, B, and C hold. Then one has*

$$\mathbf{D}_T\left(\hat{\boldsymbol{\theta}}_T\right) \to \mathbf{D}_T(\boldsymbol{\theta}_0).$$

For the analysis of $\tilde{\mathbf{V}}\left(\hat{\boldsymbol{\theta}}_T\right)$, consider the identity

$$\tilde{\mathbf{V}}_T\left(\hat{\boldsymbol{\theta}}_T\right) - \mathbf{V}_T(\boldsymbol{\theta}_0) = \tilde{\mathbf{V}}_T\left(\hat{\boldsymbol{\theta}}_T\right) - \tilde{\mathbf{V}}_T(\boldsymbol{\theta}_0) + \tilde{\mathbf{V}}_T(\boldsymbol{\theta}_0) - \mathbf{V}_T(\boldsymbol{\theta}_0),$$

where $\tilde{\mathbf{V}}_T(\boldsymbol{\theta}_0) = \sum_{j=-T+1}^{T-1}\frac{T-|j|}{T}k(j/S_T)\tilde{\mathbf{H}}_{j,\tau}(\boldsymbol{\theta}_0)$, with

$$\mathbf{H}_{j,\tau}(\boldsymbol{\theta}_0) = \frac{1}{T-j}\sum_{t=1}^{T-j} h_{t,\tau}(\boldsymbol{\theta}_0)h_{t+j,\tau}^{\top}(\boldsymbol{\theta}_0).$$

Here we first prove that $\tilde{\mathbf{V}}_T\left(\hat{\boldsymbol{\theta}}_T\right) - \tilde{\mathbf{V}}_T(\boldsymbol{\theta}_0) = o_p(1)$. To this end, a mean value expansion of $\tilde{\mathbf{V}}_T\left(\hat{\boldsymbol{\theta}}_T\right)$ about $\boldsymbol{\theta}_0$ yields

$$\begin{aligned}\frac{\sqrt{T}}{S_T}\left(\tilde{\mathbf{V}}_T\left(\hat{\boldsymbol{\theta}}_T\right) - \tilde{\mathbf{V}}_T(\boldsymbol{\theta}_0)\right) &= S_T^{-1}\nabla^{\top}\tilde{\mathbf{V}}_T(\tilde{\boldsymbol{\theta}})\sqrt{T}\left(\hat{\boldsymbol{\theta}}_T - \boldsymbol{\theta}_0\right)\\ &= S_T^{-1}\sum_{j=-T+1}^{T-1}\frac{T-|j|}{T}k(j/S_T)\nabla^{\top}\tilde{\mathbf{H}}_{j,\tau}(\tilde{\boldsymbol{\theta}})\sqrt{T}\left(\hat{\boldsymbol{\theta}}_T - \boldsymbol{\theta}_0\right),\end{aligned}$$

where $\tilde{\boldsymbol{\theta}}$ lies between $\hat{\boldsymbol{\theta}}_T$ and $\boldsymbol{\theta}_0$. In addition, one has

$$\sup_{j\geq 1}\|\nabla\tilde{\mathbf{H}}_{j,\tau}(\tilde{\boldsymbol{\theta}})\| = \sup_{j\geq 1}\left\|T^{-1}\sum_{t=|j|+1}^{T}\left(h_{t,\tau}(\tilde{\boldsymbol{\theta}})\nabla h_{t-|j|,\tau}^{\top}(\tilde{\boldsymbol{\theta}})+h_{t-|j|,\tau}(\tilde{\boldsymbol{\theta}})\nabla h_{t,\tau}^{\top}(\tilde{\boldsymbol{\theta}})\right)\right\|$$

$$\leq C\left(T^{-1}\sum_{t=1}^{T}\sup_{\boldsymbol{\theta}\in\Theta}\|h_{t,\tau}(\boldsymbol{\theta})\|^2\right)^{1/2}\left(T^{-1}\sum_{t=1}^{T}\sup_{\boldsymbol{\theta}\in\Theta}\|\nabla h_{t,\tau}^{\top}(\boldsymbol{\theta})\|^2\right)^{1/2},$$

in which

$$T^{-1}\sum_{t=1}^{T}\sup_{\boldsymbol{\theta}\in\Theta}\|h_{t,\tau}(\boldsymbol{\theta})\|^2 = T^{-1}\sum_{t=1}^{T}\|h_{t,\tau}(\boldsymbol{\theta}_0)\|^2 + T^{-1}\sum_{t=1}^{T}\sup_{\boldsymbol{\theta}\in\Theta}\|\nabla h_{t,\tau}^{\top}(\boldsymbol{\theta})\|^2\left(\tilde{\boldsymbol{\theta}}-\boldsymbol{\theta}_0\right),$$

where $\tilde{\boldsymbol{\theta}}$ lies between $\boldsymbol{\theta}$ and $\boldsymbol{\theta}_0$. Together with the following two facts

$$P\left(T^{-1}\sum_{t=1}^{T}\|h_{t,\tau}(\boldsymbol{\theta}_0)\|^2 \geq \sup_{t\geq 1}E\left(\|h_{t,\tau}(\boldsymbol{\theta}_0)\|^2\right)\right) \leq \frac{E\left(T^{-1}\sum_{t=1}^{T}\|h_{t,\tau}(\boldsymbol{\theta}_0)\|^2\right)}{\sup_{t\geq 1}E\left(\|h_{t,\tau}(\boldsymbol{\theta}_0)\|^2\right)} < 1,$$

and

$$P\left(T^{-1}\sum_{t=1}^{T}\sup_{\boldsymbol{\theta}\in\Theta}\|\nabla h_{t,\tau}^{\top}(\boldsymbol{\theta})\|^2 \geq \sup_{t\geq 1}E\sup_{\boldsymbol{\theta}\in\Theta}\left(\|h_{t,\tau}(\boldsymbol{\theta}_0)\|^2\right)\right)$$

$$\leq \frac{E\left(T^{-1}\sum_{t=1}^{T}\sup_{\boldsymbol{\theta}\in\Theta}\|\nabla h_{t,\tau}^{\top}(\boldsymbol{\theta})\|^2\right)}{\sup_{t\geq 1}E\sup_{\boldsymbol{\theta}\in\Theta}\left(\|h_{t,\tau}(\boldsymbol{\theta}_0)\|^2\right)} < 1,$$

as well as Assumption C3, $\sup_{j\geq 1}\|\nabla\tilde{\mathbf{H}}_{j,\tau}(\tilde{\boldsymbol{\theta}})\| = O_p(1)$ is proved. This results Theorem 2 and the fact that $S_T^{-1}\sum_{j=-T+1}^{T-1}|k(j/S_T)| \to \int_{-\infty}^{\infty}|k(x)|dx < \infty$ prove the claim. Further, by Theorem 1 in Andrews (1991), it is easy to show that $\tilde{\mathbf{V}}_T(\boldsymbol{\theta}_0) - \mathbf{V}_T(\boldsymbol{\theta}_0) = o_p(1)$. Therefore, the consistency of $\tilde{\mathbf{V}}_T(\hat{\boldsymbol{\theta}}_T)$ to $\mathbf{V}_T(\boldsymbol{\theta}_0)$ is proved, and the proof of Theorem 3 is complete.

Proof of Theorem 4: To establish the asymptotic result of test statistic $T^{-1/2}\sum_{t=1}^{T}\mathbf{Z}_t(\hat{\boldsymbol{\theta}}_T)w_{t,\tau}(\hat{\boldsymbol{\theta}}_T)\hat{\varepsilon}_{t,\tau}$, where $\hat{\varepsilon}_{t,\tau} = \varepsilon_{t,\tau}(\hat{\boldsymbol{\theta}})$, we first employ a continuously differentiable function to approximate the discontinuous function $w_{t,\tau}(\hat{\boldsymbol{\theta}}_T)$. Then, the mean value theorem is applied around $\boldsymbol{\theta}_0$, and the new test statistics is proved to converge in distribution to the normal distribution stated in Theorem 4. Finally, the approximation of the test statistic to the new test statistic is proved.

Define $w^*_{t,\tau}(\hat{\boldsymbol{\theta}}_T) = \tau + (1-2\tau)\{1+\exp(c_T^{-1}\hat{\varepsilon}_{t,\tau})\}^{-1} \equiv \tau + (1-2\tau)I^*_T(\hat{\varepsilon}_{t,\tau})$, where c_T is a non-stochastic sequence such that $\lim_{T\to\infty} c_T T^{1/2} = 0$. Then, one has

$$\begin{aligned}\nabla w^*_{t,\tau}(\hat{\boldsymbol{\theta}}_T) &= (1-2\tau)c_T^{-1}\exp\{c_T^{-1}\hat{\varepsilon}_{t,\tau}\}\left[1+\exp\{c_T^{-1}\hat{\varepsilon}_{t,\tau}\}\right]^{-2}\nabla e_{t,\tau}(\hat{\boldsymbol{\theta}}_T)\\ &\equiv (1-2\tau)K_{c_T}(\hat{\varepsilon}_{t,\tau})\nabla e_{t,\tau}(\hat{\boldsymbol{\theta}}_T).\end{aligned}$$

Applying the mean value theorem gives

$$\begin{aligned}T^{-1/2}\sum_{t=1}^{T}\mathbf{Z}_t(\hat{\boldsymbol{\theta}}_T)w^*_{t,\tau}(\hat{\boldsymbol{\theta}}_T)\hat{\varepsilon}_{t,\tau} = T^{-1/2}\sum_{t=1}^{T}\Big[&\mathbf{Z}_t(\boldsymbol{\theta}_0)w^*_{t,\tau}(\boldsymbol{\theta}_0)\varepsilon^*_{t,\tau} + \{\nabla^\top\mathbf{Z}_t(\tilde{\boldsymbol{\theta}})w^*_{t,\tau}(\tilde{\boldsymbol{\theta}})\tilde{\varepsilon}_{t,\tau}\\ &+\mathbf{Z}_t(\tilde{\boldsymbol{\theta}})(1-2\tau)K_{c_T}(\tilde{\varepsilon}_{t,\tau})\nabla^\top e_{t,\tau}(\tilde{\boldsymbol{\theta}})\tilde{\varepsilon}_{t,\tau}\\ &-\mathbf{Z}_t(\tilde{\boldsymbol{\theta}})w^*_{t,\tau}(\tilde{\boldsymbol{\theta}})\nabla^\top e_{t,\tau}(\tilde{\boldsymbol{\theta}})\}(\hat{\boldsymbol{\theta}}_T-\boldsymbol{\theta}_0)\Big],\end{aligned} \tag{A3}$$

where $\tilde{\boldsymbol{\theta}}$ lies between $\hat{\boldsymbol{\theta}}_T$ and $\boldsymbol{\theta}_0$, $\tilde{\varepsilon}_{t,\tau} = Y_t - e_{t,\tau}(\tilde{\boldsymbol{\theta}})$. Reformulate the above equation by adding and subtracting appropriate terms, one has

$$\begin{aligned}&T^{-1/2}\sum_{t=1}^{T}\mathbf{Z}_t(\hat{\boldsymbol{\theta}}_T)w^*_{t,\tau}(\hat{\boldsymbol{\theta}}_T)\hat{\varepsilon}_{t,\tau}\\ &= T^{-1/2}\sum_{t=1}^{T}\Big[\mathbf{Z}_t(\boldsymbol{\theta}_0)w_{t,\tau}(\boldsymbol{\theta}_0)\varepsilon^*_{t,\tau} + (\mathbf{Z}_t(\boldsymbol{\theta}_0)w^*_{t,\tau}(\boldsymbol{\theta}_0)\varepsilon^*_{t,\tau} - \mathbf{Z}_t(\boldsymbol{\theta}_0)w_{t,\tau}(\boldsymbol{\theta}_0)\varepsilon^*_{t,\tau})\\ &+\{\nabla^\top\mathbf{Z}_t(\tilde{\boldsymbol{\theta}})w^*_{t,\tau}(\tilde{\boldsymbol{\theta}})\tilde{\varepsilon}_{t,\tau} + \mathbf{Z}_t(\tilde{\boldsymbol{\theta}})(1-2\tau)K_{c_T}(\tilde{\varepsilon}_{t,\tau})\nabla^\top e_{t,\tau}(\tilde{\boldsymbol{\theta}})\tilde{\varepsilon}_{t,\tau} - \mathbf{Z}_t(\tilde{\boldsymbol{\theta}})w^*_{t,\tau}(\tilde{\boldsymbol{\theta}})\nabla^\top e_{t,\tau}(\tilde{\boldsymbol{\theta}})\}(\hat{\boldsymbol{\theta}}_T-\boldsymbol{\theta}_0)\\ &-\mathbf{Z}_t(\boldsymbol{\theta}_0)w_{t,\tau}(\boldsymbol{\theta}_0)\nabla^\top e_{t,\tau}(\boldsymbol{\theta}_0)\mathbf{D}_T^{-1}(\boldsymbol{\theta}_0)T^{-1}\sum_{t=1}^{T}h_{t,\tau}(\boldsymbol{\theta}_0) + \mathbf{Z}_t(\boldsymbol{\theta}_0)w_{t,\tau}(\boldsymbol{\theta}_0)\nabla^\top e_{t,\tau}(\boldsymbol{\theta}_0)\mathbf{D}_T^{-1}(\boldsymbol{\theta}_0)\\ &T^{-1}\sum_{t=1}^{T}h_{t,\tau}(\boldsymbol{\theta}_0) - E[\mathbf{Z}_t(\boldsymbol{\theta}_0)w_{t,\tau}(\boldsymbol{\theta}_0)\nabla^\top e_{t,\tau}(\boldsymbol{\theta}_0)]\mathbf{D}_T^{-1}(\boldsymbol{\theta}_0)T^{-1}\sum_{t=1}^{T}h_{t,\tau}(\boldsymbol{\theta}_0)\\ &+E[\mathbf{Z}_t(\boldsymbol{\theta}_0)w_{t,\tau}(\boldsymbol{\theta}_0)\nabla^\top e_{t,\tau}(\boldsymbol{\theta}_0)]\mathbf{D}_T^{-1}(\boldsymbol{\theta}_0)T^{-1}\sum_{t=1}^{T}h_{t,\tau}(\boldsymbol{\theta}_0)\Big]\end{aligned}$$

In the following lemma, some of the terms in the above equation are proved to be $o_p(1)$.

Lemma A4. *Under the same assumptions as in Theorem 4, when $T\to\infty$, one has*

i) $\|T^{-1/2}\sum_{t=1}^{T}(\mathbf{Z}_t(\boldsymbol{\theta}_0)w^*_{t,\tau}(\boldsymbol{\theta}_0)\varepsilon^*_{t,\tau} - \mathbf{Z}_t(\boldsymbol{\theta}_0)w_{t,\tau}(\boldsymbol{\theta}_0)\varepsilon^*_{t,\tau})\| = o_p(1)$;

ii) $\|T^{-1/2}\sum_{t=1}^{T}\Big\{\Big(\nabla^\top\mathbf{Z}_t(\tilde{\boldsymbol{\theta}})w^*_{t,\tau}(\tilde{\boldsymbol{\theta}})\tilde{\varepsilon}_{t,\tau} + \mathbf{Z}_t(\tilde{\boldsymbol{\theta}})(1-2\tau)K_{c_T}(\tilde{\varepsilon}_{t,\tau})\nabla^\top e_{t,\tau}(\tilde{\boldsymbol{\theta}})\tilde{\varepsilon}_{t,\tau} - \mathbf{Z}_t(\tilde{\boldsymbol{\theta}})w^*_{t,\tau}(\tilde{\boldsymbol{\theta}})$ $\nabla^\top e_{t,\tau}(\tilde{\boldsymbol{\theta}})\Big)(\hat{\boldsymbol{\theta}}_T-\boldsymbol{\theta}_0) - \mathbf{Z}_t(\boldsymbol{\theta}_0)w_{t,\tau}(\boldsymbol{\theta}_0)\nabla^\top e_{t,\tau}(\boldsymbol{\theta}_0)\mathbf{D}_T^{-1}(\boldsymbol{\theta}_0)T^{-1}\sum_{t=1}^{T}h_{t,\tau}(\boldsymbol{\theta}_0)\}\| = o_p(1)$;

iii) $\|T^{-1}\sum_{t=1}^{T}\Big\{\mathbf{Z}_t(\boldsymbol{\theta}_0)w_{t,\tau}(\boldsymbol{\theta}_0)\nabla^\top e_{t,\tau}(\boldsymbol{\theta}_0)\mathbf{D}_T^{-1}(\boldsymbol{\theta}_0)T^{-1/2}\sum_{t=1}^{T}h_{t,\tau}(\boldsymbol{\theta}_0) - E\big[\mathbf{Z}_t(\boldsymbol{\theta}_0)w_{t,\tau}(\boldsymbol{\theta}_0)$ $\nabla^\top e_{t,\tau}(\boldsymbol{\theta}_0)]\mathbf{D}_T^{-1}(\boldsymbol{\theta}_0)T^{-1/2}\sum_{t=1}^{T}h_{t,\tau}(\boldsymbol{\theta}_0)\}\| = o_p(1)$.

Then, it follows from Lemma A4 that

$$
\begin{aligned}
&T^{-1/2}\sum_{t=1}^{T}\mathbf{Z}_t(\hat{\boldsymbol{\theta}}_T)w^*_{t,\tau}(\hat{\boldsymbol{\theta}}_T)\hat{\varepsilon}_{t,\tau}\\
&=T^{-1/2}\sum_{t=1}^{T}\left\{\mathbf{Z}_t(\boldsymbol{\theta}_0)w_{t,\tau}(\boldsymbol{\theta}_0)\varepsilon^*_{t,\tau}+E\big[\mathbf{Z}_t(\boldsymbol{\theta}_0)w_{t,\tau}(\boldsymbol{\theta}_0)\nabla^{\top}e_{t,\tau}(\boldsymbol{\theta}_0)\big]\mathrm{D}_T^{-1}(\boldsymbol{\theta}_0)T^{-1}\sum_{t=1}^{T}h_{t,\tau}(\boldsymbol{\theta}_0)\right\}\\
&+o_p(1)\\
&=T^{-1/2}\sum_{t=1}^{T}\left\{\mathbf{Z}_t(\boldsymbol{\theta}_0)+\left(T^{-1}\sum_{t=1}^{T}E\big[\mathbf{Z}_t(\boldsymbol{\theta}_0)w_{t,\tau}(\boldsymbol{\theta}_0)\nabla^{\top}e_{t,\tau}(\boldsymbol{\theta}_0)\big]\right)\mathrm{D}_T^{-1}(\boldsymbol{\theta}_0)\nabla e_{t,\tau}(\boldsymbol{\theta}_0)\right\}\\
&w_{t,\tau}(\boldsymbol{\theta}_0)\varepsilon^*_{t,\tau}+o_p(1)\\
&\equiv T^{-1/2}\sum_{t=1}^{T}\boldsymbol{\Xi}_t(\boldsymbol{\theta}_0)w_{t,\tau}(\boldsymbol{\theta}_0)\varepsilon^*_{t,\tau},
\end{aligned}
$$

and it is easy to show that $g_{t,\tau}(\boldsymbol{\theta}_0)\equiv\Xi_{t,\tau}(\boldsymbol{\theta}_0)w_{t,\tau}(\boldsymbol{\theta}_0)\varepsilon^*_{t,\tau}\in H(\mathcal{F}_t)$ is a vector of random variables which is measurable with respect to $\mathcal{F}_t$, and by Assumption D3, it satisfies that $E(\|g_{t,\tau}(\boldsymbol{\theta}_0)\|^{2r+\varepsilon})<\infty$, for some $\varepsilon>0$. At the same time, one has

$$
E\big(g_{t,\tau}(\boldsymbol{\theta}_0)\big)=E\big[g_{t,\tau}(\boldsymbol{\theta}_0)|\mathbf{W}_t)\big]=0
$$

for the first order condition. Then, with Assumption D4, the central limit theory for functionals of mixing sequences (see Theorem 18.6.2 in Ibragimov and Linnik 1971) can be applied, and the proof is complete.

Proof of Theorem 5: Following the same manner as in Eqution (A3), one has

$$
\begin{aligned}
&N_r^{-1/2}\sum_{t=T_r+1}^{T_r+N_r}\mathbf{Z}_t(\hat{\boldsymbol{\theta}}_{T_r})w^*_{t,\tau}(\hat{\boldsymbol{\theta}}_{T_r})\hat{\varepsilon}_{t,\tau}(\hat{\boldsymbol{\theta}}_{T_r})\\
&\quad=N_r^{-1/2}\sum_{t=T_r+1}^{T_r+N_r}[\mathbf{Z}_t(\boldsymbol{\theta}_0)w^*_{t,\tau}(\boldsymbol{\theta}_0)\varepsilon^*_{t,\tau}+\{\nabla^{\top}\mathbf{Z}_t(\tilde{\boldsymbol{\theta}})w^*_{t,\tau}(\tilde{\boldsymbol{\theta}})\tilde{\varepsilon}_{t,\tau}\\
&\quad+\mathbf{Z}_t(\tilde{\boldsymbol{\theta}})(1-2\tau)K_{c_T}(\tilde{\varepsilon}_{t,\tau})\nabla^{\top}e_{t,\tau}(\tilde{\boldsymbol{\theta}})\tilde{\varepsilon}_{t,\tau}-\mathbf{Z}_t(\tilde{\boldsymbol{\theta}})w^*_{t,\tau}(\tilde{\boldsymbol{\theta}})\nabla^{\top}e_{t,\tau}(\tilde{\boldsymbol{\theta}})\}(\hat{\boldsymbol{\theta}}_{T_r}-\boldsymbol{\theta}_0)],
\end{aligned}
$$

where $\tilde{\boldsymbol{\theta}}$ lies between $\hat{\boldsymbol{\theta}}_{T_r}$ and $\boldsymbol{\theta}_0$, and $\tilde{\varepsilon}_{t,\tau}=Y_t-e_{t,\tau}(\tilde{\boldsymbol{\theta}})$. Assumption E1, consistency of $\hat{\boldsymbol{\theta}}_{T_r}$, and Slutsky's theorem yield

$$
\begin{aligned}
&\lim_{r\to\infty}N_r^{-1/2}\sum_{t=T_r+1}^{T_r+N_r}\mathbf{Z}_t(\hat{\boldsymbol{\theta}}_{T_r})w^*_{t,\tau}(\hat{\boldsymbol{\theta}}_{T_r})\hat{\varepsilon}_{t,\tau}(\hat{\boldsymbol{\theta}}_{T_r})\\
&=\lim_{r\to\infty}[N_r^{-1/2}\sum_{t=T_r+1}^{T_r+N_r}\mathbf{Z}_t(\boldsymbol{\theta}_0)w^*_{t,\tau}(\boldsymbol{\theta}_0)\varepsilon^*_{t,\tau}+\left(\frac{N_r}{T_r}\right)^{1/2}\frac{1}{N_r}\{\nabla^{\top}\mathbf{Z}_t(\tilde{\boldsymbol{\theta}})w^*_{t,\tau}(\tilde{\boldsymbol{\theta}})\tilde{\varepsilon}_{t,\tau}\\
&+\mathbf{Z}_t(\tilde{\boldsymbol{\theta}})(1-2\tau)K_{c_T}(\tilde{\varepsilon}_{t,\tau})\nabla^{\top}e_{t,\tau}(\tilde{\boldsymbol{\theta}})\tilde{\varepsilon}_{t,\tau}-\mathbf{Z}_t(\tilde{\boldsymbol{\theta}})w^*_{t,\tau}(\tilde{\boldsymbol{\theta}})\nabla^{\top}e_{t,\tau}(\tilde{\boldsymbol{\theta}})\}T_r^{1/2}(\hat{\boldsymbol{\theta}}_{T_r}-\boldsymbol{\theta}_0)]\\
&=\lim_{r\to\infty}N_r^{-1/2}\sum_{t=T_r+1}^{T_r+N_r}\mathbf{Z}_t(\boldsymbol{\theta}_0)w^*_{t,\tau}(\boldsymbol{\theta}_0)\varepsilon^*_{t,\tau}.
\end{aligned}
$$

With Assumption E2, the proof is completed by the same procedure as in the proof of Theorem 4.

References

Andrews, D. W. 1991. Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimation. *Econometrica* 59: 817–858.

Artzner, P., F. Delbaen, J.-M. Eber, and D. Heath. 1999. Coherent Measures of Risk. *Mathematical Finance* 9: 203–228.

Bellini, F., and E. Di Bernardino. 2017. Risk Management with Expectiles. *The European Journal of Finance* 23: 487–506.

Bollerslev, T. 1986. Generalized Autoregressive Conditional Heteroskedasticity. *Journal of Econometrics* 31: 307–327.

Cai, Z., Y. Fang, and D. Tian. 2018. Assessing Tail Risk Using Expectile Regressions with Partially Varying Coefficients. *Journal of Management Science and Engineering* 3: 183–213.

Chen, J., and Z. Chen. 2008. Extended Bayesian Information Criteria for Model Selection with Large Model Spaces. *Biometrika* 95: 759–771.

Daouia, A., I. Gijbels, and G. Stupfler. 2019. Extremiles: A New Perspective on Asymmetric Least Squares. *Journal of the American Statistical Association* 114: 1366–1381.

Daouia, A., S. Girard, and G. Stupfler. 2018. Estimation of Tail Risk Based on Extreme Expectiles. *Journal of the Royal Statistical Society Series B: Statistical Methodology* 80: 263–292.

Daouia, A., S. Girard, and G. Stupfler. 2020. Tail Expectile Process and Risk Assessment. *Bernoulli* 26: 531–556.

Embrechts, P., and M. Hofert. 2014. Statistics and Quantitative Risk Management for Banking and Insurance. *Annual Review of Statistics and Its Application* 1: 493–514.

Engle, R. F., and S. Manganelli. 2004. CAViaR: Conditional Autoregressive Value at Risk by Regression Quantiles. *Journal of Business & Economic Statistics* 22: 367–381.

Gerlach, R., and C. W. Chen. 2016. Bayesian Expected Shortfall Forecasting Incorporating the Intraday Range. *Journal of Financial Econometrics* 14(1): 128–158.

Gerlach, R., D. Walpole, and C. Wang. 2017. Semi-Parametric Bayesian Tail Risk Forecasting Incorporating Realized Measures of Volatility. *Quantitative Finance* 17: 199–215.

Gneiting, T. 2011. Making and Evaluating Point Forecasts. *Journal of the American Statistical Association* 106: 746–762.

Hamidi, B., B. Maillet, and J.-L. Prigent. 2014. A Dynamic Autoregressive Expectile for Time-Invariant Portfolio Protection Strategies. *Journal of Economic Dynamics and Control* 46: 1–29.

Horvath, P., J. Li, Z. Liao, and A. J. Patton. 2022. A Consistent Specification Test for Dynamic Quantile Models. *Quantitative Economics* 13: 125–151.

Huber, P. J. 1967. "The Behavior of Maximum Likelihood Estimates Under Nonstandard Conditions." In *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability*, Vol. 1, pp. 221–233. California: University of California Press.

Ibragimov, I. A., and Y. V. Linnik. 1971. *Independent and Stationary Sequences of Random Variables*. Groningen: Wolters-Noordhoff.

Jiang, C., Y. Ma, and Y. An. 2009. The Effectiveness of the VaR-Based Portfolio Insurance Strategy: An Empirical Analysis. *International Review of Financial Analysis* 18: 185–197.

Jiang, R., X. Hu, and K. Yu. 2022. Single-Index Expectile Models for Estimating Conditional Value at Risk and Expected Shortfall. *Journal of Financial Econometrics* 20: 345–366.

Kuan, C.-M., J.-H. Yeh, and Y.-C. Hsu. 2009. Assessing Value at Risk with CARE, the Conditional Autoregressive Expectile Models. *Journal of Econometrics* 150: 261–270.

Lee, E. R., H. Noh, and B. U. Park. 2014. Model Selection via Bayesian Information Criterion for Quantile Regression Models. *Journal of the American Statistical Association* 109: 216–229.

Mohammedi, M., S. Bouzebda, and A. Laksaci. 2021. The Consistency and Asymptotic Normality of the Kernel Type Expectile Regression Estimator for Functional Data. *Journal of Multivariate Analysis* 181: 104673.

Newey, W. K., and J. L. Powell. 1987. Asymmetric Least Squares Estimation and Testing. *Econometrica* 55: 819–847.

Pollard, D. 1991. Asymptotics for Least Absolute Deviation Regression Estimators. *Econometric Theory* 7: 186–199.

Taylor, J. W. 2008. Estimating Value at Risk and Expected Shortfall Using Expectiles. *Journal of Financial Econometrics* 6(2): 231–252.

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

Tian, D., Z. Cai, and Y. Fang. 2019. Econometric Modeling of Risk Measures: A Selective Review of the Recent Literature. *Applied Mathematics-A Journal of Chinese Universities* 34: 205–228.

Weiss, A. A. 1991. Estimating Nonlinear Dynamic Models Using Least Absolute Error Estimation. *Econometric Theory* 7: 46–68.

White, H. 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: 169–188.

Wooldridge, J. M. 1994. Estimation and Inference for Dependent Processes. *Handbook of Econometrics* 4: 2639–2738.

Wuertz, D., T. Setz, Y. Chalabi, C. Boudt, P. Chausse, and M. Miklovac. 2017. *fGarch: Rmetrics - Autoregressive Conditional Heteroskedastic Modelling*, R package version 3042.83.

Xiao, Z., and R. Koenker. 2009. Conditional Quantile Estimation for Generalized Autoregressive Conditional Heteroscedasticity Models. *Journal of the American Statistical Association* 104: 1696–1712.

Xie, S., Y. Zhou, and A. T. Wan. 2014. A Varying-Coefficient Expectile Model for Estimating Value-at-Risk. *Journal of Business & Economic Statistics* 32: 576–592.

Xu, W., Y. Hou, and D. Li. 2022. Prediction of Extremal Expectile Based on Regression Models with Heteroscedastic Extremes. *Journal of Business & Economic Statistics* 40: 522–536.

Xu, X., A. Mihoci, and W. K. Härdle. 2018. lCARE-Localizing Conditional Autoregressive Expectiles. *Journal of Empirical Finance* 48: 198–220.

Zhang, F., and Q. Li. 2017. A Continuous Threshold Expectile Model. *Computational Statistics & Data Analysis* 116: 49–66.

Ziegel, J. F. 2016. Coherence and Elicitability. *Mathematical Finance* 26: 901–918.

Journal of Financial Econometrics, 2025, 23, 1–39
https://doi.org/10.1093/jjfinec/nbaf010
Article