ESTIMATION OF HIGH CONDITIONAL TAIL RISK BASED ON EXPECTILE REGRESSION

BY

JIE HU, YU CHEN AND KEQI TAN

ABSTRACT

Assessing conditional tail risk at very high or low levels is of great interest in numerous applications. Due to data sparsity in high tails, the widely used quantile regression method can suffer from high variability at the tails, especially for heavy-tailed distributions. As an alternative to quantile regression, expectile regression, which relies on the minimization of the asymmetric l_2-norm and is more sensitive to the magnitudes of extreme losses than quantile regression, is considered. In this article, we develop a new estimation method for high conditional tail risk by first estimating the intermediate conditional expectiles in regression framework, and then estimating the underlying tail index via weighted combinations of the top order conditional expectiles. The resulting conditional tail index estimators are then used as the basis for extrapolating these intermediate conditional expectiles to high tails based on reasonable assumptions on tail behaviors. Finally, we use these high conditional tail expectiles to estimate alternative risk measures such as the Value at Risk (VaR) and Expected Shortfall (ES), both in high tails. The asymptotic properties of the proposed estimators are investigated. Simulation studies and real data analysis show that the proposed method outperforms alternative approaches.

KEYWORDS

Quantile regression, expectile regression, asymptotic, heavy-tailed distribution, tail index.

1. INTRODUCTION

A noteworthy problem in many fields involving statistical applications is modeling and predicting of extreme events. Extreme events usually refer to the events that happen rarely but lead to significant consequences, for example, large natural disasters, large financial loss, high medical costs, low birth

Astin Bulletin, 1-32. doi:10.1017/asb.2021.3

weights. For such events, it is particularly interesting to model and estimate the tail events of the underlying distribution rather than the averages. In other words, we concentrate on the concept of "tail risk".

Notably, finding a proper risk measure is one of the most important and challenging tasks in financial risk management. Among numerous works in literature, Value at Risk (VaR) is arguably the most common risk measure used in practice. VaR denotes the loss that is likely to be exceeded at a specified probability level, which is actually the quantile of a portfolio loss distribution. However, VaR fails to fulfill the subadditivity property in general (Acerbi, 2002), and hence it is not a coherent risk measure according to the axiomatic foundations of Artzner *et al.* (1999). Moreover, in some extreme cases, for example, occurrences of catastrophic events, VaR becomes a conservative tail risk measure because a quantile-based risk measure depends only on the probability of the occurrence of an extreme loss rather than the magnitude of the extreme loss. It is, therefore, easy to construct two return distributions that have different tail behaviors but the same VaR. Compensating for these two weaknesses of VaR, expectiles, first introduced by Newey and Powell (1987), are reasonable alternative to quantiles as they depend on both the tail realizations and their probabilities (Kuan *et al.*, 2009), and they define a kind of coherent risk measure. This is mainly due to their conception as least squares analogues of quantiles. In addition, expectiles are attractive in applications because they are more tail sensitive than VaR and ES. Finally, since there exists a one-to-one mapping between quantiles and expectiles, as argued in Yao and Tong (1996), and there is a link between VaR and ES, as addressed in Taylor (2008), expectiles can be used to calculate VaR and ES simultaneously. Further theoretical and numerical results obtained by Bellini and Di Bernardino (2017) indicate that expectiles are perfectly reasonable alternatives to both classical quantile-based VaR and ES.

Extreme value theory (EVT) provides another elegant mathematical tool for analyzing extreme events. However, to our knowledge, few related works have been performed with regard to the estimation of high conditional expectiles. The literature of EVT mainly focuses on tail quantiles with independent and identically distributed random variables (see Weissman, 1978; de Haan and Ferreira, 2006; Li *et al.*, 2010). In contrast, tail expectiles, as well as their relative statistical problems, are rarely mentioned in extreme value theory, unlike VaR estimation and ES estimation. Fortunately, Daouia *et al.* (2018) and Daouia *et al.* (2020b) proposed some intermediate and extreme expectile estimators and developed their asymptotic properties. Daouia *et al.* (2020b) first constructed asymmetric least squares estimator for the tail index and derived its asymptotic normality theorem. In many applications, however, the tail quantiles or tail expectiles of the variable Y of interest depend on some covariate X, and thus, it is important to incorporate the covariate information into a given analysis. For instance, risk managers in finance often seek to forecast the low conditional quantiles of a portfolios future returns, or the conditional expectiles on information from the past or assumptions on future interest rate changes (Cai *et al.*, 2018). Therefore, our proposed work on high conditional expectiles is meaningful. Without loss of generality, in this article, we focus on the estimation of high conditional expectiles, since a low expectile of Y can be viewed as a high expectile of $-Y$.

Different from typical EVT methods that estimate the tail index using the upper quantiles (Hill, 1975), upper expectiles (Daouia *et al.*, 2020b), or link it to some covariates (Wang and Tsai, 2009), our proposal integrates expectile regression and EVT to estimate the tail index. By replacing asymmetric l_1-norm with asymmetric l_2-norm, expectile regression has gained increasing attention in recent studies. Sobotka *et al.* (2013) investigated the relationship between womens education and fertility in Botswana via semi-parametric expectile regression. Kuan *et al.* (2009) proposed conditional autoregressive expectile (CARE) models to assess VaR. Cai *et al.* (2018) applied expectile regressions with partially varying coefficients to assess tail risk. Although expectile regression has found applications in various fields, to our knowledge, few works are concerned with the estimation of high conditional expectiles. In this article, we propose a novel method to extrapolate conditional expectiles with an extremely high level by using expectile-based tail index.

To estimate the conditional expectiles in a very far tail, where few observations are available, we need some additional assumptions on the tail behavior. First, we use expectile regression to estimate the "intermediate" conditional expectiles at levels τ_n, where τ_n is close to one but still in the "intermediate" range (mathematically, $\tau_n \to 1$ and $n(1-\tau_n) \to \infty$ as $n \to \infty$, where n is the sample size). Then, we estimate the underlying tail index via a weighted combination of the top order conditional expectiles. Finally, we extrapolate these intermediate conditional expectile estimates to a very high level τ'_n (mathematically, this is a high level in the sense that $\tau'_n \to 1$ and $n(1-\tau'_n) \to c$, where c is a positive constant) through the estimated tail index. These extreme conditional expectiles are then used to calculate expectile-based VaR and expectile-based ES, both in high tails.

Our proposed method enriches the literature in four ways. First, we propose a method to estimate the tail index by virtue of conditional expectiles instead of conditional quantiles as done by Wang *et al.* (2012). Second, we develop a parametric expectile regression model by borrowing information across covariates, which has a more intuitive interpretation than existing methods. Third, we provide another method to estimate the high-tailed conditional VaR and ES based on adapted extreme expectile-based tools. Finally, expectile regression is more sensitive to extreme values than quantile regression, so it can play an early warning role in the detection of heteroscedasticity in financial data applications. In addition, based on the asymmetric least squares loss, the computation of expectile regression can be straightforward and simple, and the theoretical development of expectile regression is more manageable than that of quantile regression and the estimation procedure is more efficient as it uses the entire of the conditional distribution information.

The rest of the article is organized as follows. Section 2 presents the preliminaries of quantiles, expectiles and extreme value theory (EVT). Section 3 illustrates our proposed estimation method and its asymptotic properties. In Section 4, we conduct simulation studies to assess the finite sample performance of the proposed method. The results indicate that the proposed method is more efficient and accurate than quantile-based method when estimating extremely high level of VaR and ES for zero-mean heavy-tailed distributions. In Section 5, we apply the proposed method to two financial datasets of the Chinese stock market. Finally, Section 6 concludes the article. All the technical details about our proposed method are provided in the Appendix.

2. BACKGROUND

2.1. Quantiles and expectiles

Let Y denote a random variable with the distribution function F_Y. Given $\tau \in (0, 1)$, the VaR at a probability level τ is defined by the τth quantile $q_\tau := F_Y^{\leftarrow}(\tau) = \inf\{y \in \mathbb{R} : F_Y(y) \geq \tau\}$. Koenker and Bassett (1978) proposed a method to estimate the τth quantile by minimizing the following asymmetrically absolute deviation problem:

$$q_\tau := \arg\min_{m\in\mathbb{R}} E\Big[|\tau - \mathbb{I}(Y \leq m)| \cdot |Y - m|\Big],$$

where $\mathbb{I}(\cdot)$ is the indicator function. By replacing the asymmetric l_1-norm with the asymmetric l_2-norm, the τth expectile (see Newey and Powell, 1987) is defined as follows:

$$\xi_\tau := \arg\min_{m\in\mathbb{R}} E\Big[|\tau - \mathbb{I}(Y \leq m)| \cdot (Y - m)^2\Big].$$

In terms of interpretability, the τth quantile indicates the point where $100\tau\%$ of Y have values less than this number, while the τth expectile specifies the position of ξ_τ such that the average distance from the data below ξ_τ to ξ_τ itself is $100\tau\%$ of the average distance between ξ_τ and all the data, that is,

$$\tau = \frac{E[|Y - \xi_\tau|\mathbb{I}(Y \leq \xi_\tau)]}{E|Y - \xi_\tau|}. \tag{2.1}$$

Thus, the ξ_τ shares an intuitive interpretation similar to that of q_τ, replacing the number of observations by distance.

2.2. Extreme value theory

Throughout the paper, the notation $a_n \sim b_n$ represents that $a_n/b_n \to 1$ as $n \to \infty$. We focus on high expectiles in the right tail and restrict ourselves in heavy-tailed distributions that are attracted to the maximum domain of Pareto-type distributions with tail index $0 < \gamma < 1$.

We say that $F_0(\cdot)$ is a Pareto-type distribution if its survival function satisfies

$$\overline{F}_0(y) := 1 - F_0(y) = y^{-1/\gamma}\ell(y), \tag{2.2}$$

for $y > 0$ large enough, where ℓ is a slowly varying function at infinity, that is, a positive function on $(0, \infty)$ satisfying $\ell(ty)/\ell(t) \to 1$, as $t \to \infty$, for any $y > 0$. The index γ tunes the tail heaviness of $\overline{F}_0$, that is, the larger the index is, the heavier the right tail is. The assumption $\gamma < 1$ ensures the existence of the first moment of $F_0(\cdot)$, and hence ensures the existence of expectiles. By Corollary 1.2.10 in de Haan and Ferreira (2006), the model assumption (2.2) is equivalent to

$$\lim_{t\to\infty} \frac{U_0(tz)}{U_0(t)} = z^{\gamma} \quad \text{for all} \quad z > 0, \tag{2.3}$$

where $U_0(t) = \inf\{y \in \mathbb{R} : F_0(y) \geq 1 - 1/t\} = F_0^{-1}(1 - 1/t)$ for $t \geq 1$, representing the $(1 - 1/t)$th quantile of the random variable. Under (2.2), Bellini and Di Bernardino (2017) pointed out that

$$\frac{\xi_{0,\tau}}{q_{0,\tau}} \sim (\gamma^{-1} - 1)^{\gamma} \quad \text{as} \quad \tau \to 1, \tag{2.4}$$

where $q_{0,\tau}$ and $\xi_{0,\tau}$ are the τth quantile and expectile of $F_0(\cdot)$, respectively. The asymptotic equality of (2.4) establishes a connection between $q_{0,\tau}$ and $\xi_{0,\tau}$ under the heavy-tailed distribution assumption, and this suggests us to estimate VaR in a high level based on the high-tailed expectile.

To obtain the asymptotic normality of the tail index estimation, the following second-order regular variation condition, which is denoted by $\mathcal{C}_2(\gamma, \rho, \mathbf{A})$, is assumed as follows:

Condition $\mathcal{C}_2(\gamma, \rho, \mathbf{A})$: for all $z > 0$,

$$\lim_{t\to\infty} \frac{1}{A(t)} \left[\frac{U_0(tz)}{U_0(t)} - z^{\gamma} \right] = z^{\gamma} \frac{z^{\rho} - 1}{\rho}, \quad \rho \leq 0, \tag{2.5}$$

where $A(t) \in RV(\rho)$ means that $A(t)$ is a regularly varying function with index ρ, which satisfies $\lim_{t\to\infty} A(tz)/A(t) = z^{\rho}$ for all $z \in \mathbb{R}^+$.

Hereafter, $(z^{\rho} - 1)/\rho$ is represented as $\log z$ when $\rho = 0$. The condition $\mathcal{C}_2(\gamma, \rho, \mathbf{A})$, which controls the rate of convergence in (2.3), is a standard condition to obtain the asymptotic properties in extreme value theory. For more details on the second-order condition, we refer to de Haan and Ferreira (2006), which gives abundant examples of commonly used continuous distributions that satisfy the condition (2.5), along with thorough discussions on the interpretation and the rationale behind this second-order condition. For instance, the t distribution with the degree of freedom ν satisfies (2.5) where $\gamma = 1/\nu$ and $\rho = -2/\nu$.

3. Proposed method

Quantiles and expectiles are readily extended to conditional quantiles and conditional expectiles. In practice, quantile regression is more robust to outliers than expectile regression. In other words, quantile regression is not sensitive to outliers, and this makes it inefficient and underestimated for heavy-tailed distributions; this motivates us to consider the method of estimating the tail index γ based on expectile regression.

Suppose we observe a random sample $\{(\mathbf{x}_i, \mathrm{y}_i), i = 1, \ldots, n\}$ from the random vector $(\mathbf{X}, Y)$, where $\mathbf{x}_i$ is the p-dimensional design vector and y_i is the univariate response variable. Denote $F_Y(\cdot|\mathbf{x})$ the conditional distribution of Y given $\mathbf{x}$ and $\xi_Y(\tau|\mathbf{x})$ the τth conditional expectile of Y given $\mathbf{x}$. Then $\xi_Y(\tau|\mathbf{x})$ is expressed as the following linear expectile regression model:

$$\xi_Y(\tau|\mathbf{x}) = \alpha(\tau) + \mathbf{x}^T\boldsymbol{\beta}(\tau), \quad \text{for all } \tau \in [\tau_l, 1], \tag{3.6}$$

where the expectile coefficients $(\alpha(\tau), \boldsymbol{\beta}(\tau)^T)^T$ may vary across $\tau \in [\tau_l, 1]$, and τ_l is a defined expectile level that can be close to 1. The estimator can be obtained in the form of a vector that minimizes the following asymmetric least squares loss function:

$$\boldsymbol{Q}_n((\alpha, \boldsymbol{\beta}); \tau) = \sum_{i=1}^{n} \varrho_\tau(y_i - \alpha - \mathbf{x}_i^T\boldsymbol{\beta}), \tag{3.7}$$

where $\varrho_\tau(u) = |\tau - \mathbb{I}(u < 0)| \cdot u^2$ is the expectile loss function. It should be pointed out that the computation required for expectile regression is more straightforward than that of quantile regression since it is based on the squared loss function as in (3.7) and can be easily solved by an iterated weighted least squares algorithm. Additionally, the sensitivity of expectile regression to extreme values can be beneficial when detecting heteroscedasticity in a given dataset, as this is one of the main issues in financial applications.

Our main objective in this work is to estimate the conditional expectile $\xi_Y(\tau_n'|\mathbf{x})$, as well as the expectile-based VaR and ES at an extremely high level. Here, τ_n' may approach one at any rate. Throughout this article, we also assume that $F_Y(\cdot|\mathbf{x})$ is in the maximum domain of attraction of a Pareto-type distribution, that is, for a given random sample $Y_1, \ldots, Y_n$ from $F_Y(\cdot|\mathbf{x})$, there are sequences $a_n > 0$ and $b_n \in R$ such that

$$\mathbb{P}\left(\frac{\max_{1\le i\le n} Y_i - b_n}{a_n} \le y\right) \to \exp\{-(1+\gamma y)^{-1/\gamma}\},$$

as $n \to \infty$ for $1 + \gamma y \ge 0$, where γ is the tail index. The key idea of the proposed method is to estimate a sequence of intermediate expectiles at levels $n(1-\tau_n) \to \infty$, and extrapolate those intermediate expectiles to an extreme level $n(1-\tau_n') \to c$, where c is a positive constant.

First, define a sequence of expectile levels $\tau_l < \tau_{n-k} < \tau_{n-k+1} < \cdots < \tau_m$, where $m = n - [n^\eta]$ with $[a]$ denoting the integer part of a and $\eta > 0$ is some small constant that make sure $[n^\eta] < k$. Here, assume that $k = k_n \to \infty$ and $k/n \to 0$ as $n \to \infty$. For each $j = n-k, n-k+1, \ldots, m$, define $\tau_j = j/(n+1)$ and estimate $(\widehat{\alpha}(\tau_j), \widehat{\boldsymbol{\beta}}(\tau_j)^T)^T$ by

$$(\widehat{\alpha}(\tau_j), \widehat{\boldsymbol{\beta}}(\tau_j)^T)^T = \arg\min_{\alpha,\boldsymbol{\beta}} \boldsymbol{Q}_n((\alpha, \boldsymbol{\beta}); \tau_j). \tag{3.8}$$

Then, for a given $\mathbf{x}$, we propose a novel method to estimate the tail index γ based on $\widehat{\xi}_j := \widehat{\xi}_Y(\tau_j|\mathbf{x}) = \widehat{\alpha}(\tau_j) + \mathbf{x}^T\widehat{\boldsymbol{\beta}}(\tau_j)$, for $j = n-k, n-k+1, \ldots, m$. The expectiles $\widehat{\xi}_j$ can be roughly regarded as the upper order statistics of a sample from $F_Y(\cdot|\mathbf{x})$. Recalling that $F_Y(\cdot|\mathbf{x})$ is in the maximum domain of attraction of a Pareto-type distribution, therefore, similarly to Daouia *et al.* (2020b), we can estimate γ by

$$\widehat{\gamma} = \frac{1}{k - [n^\eta]} \sum_{j=[n^\eta]}^{k} \log \frac{\widehat{\xi}_{n-j}}{\widehat{\xi}_{n-k}}. \tag{3.9}$$

In contrast, the estimator $\widehat{\gamma}$ can be seen as $\widehat{\gamma}(\mathbf{x})$, which leads to a nonlinear combination of the covariates. Daouia *et al.* (2018) extrapolated any consistent intermediate expectile estimator to estimate an expectile at an extremely high level, that is, estimating $\widehat{\xi}_{n-k}$ to a very high level τ_n' by using the classical Weissman extrapolation formula (Weissman, 1978). Consequently, $\xi_Y(\tau_n'|\mathbf{x})$ can be estimated by

$$\widehat{\xi}_{\tau_n'} := \left(\frac{1-\tau_{n-k}}{1-\tau_n'}\right)^{\widehat{\gamma}} \widehat{\xi}_{n-k}. \tag{3.10}$$

The tail index γ estimated by (3.9) differs from that estimated by the upper $k - [n^\eta]$ conditional quantiles in Wang *et al.* (2012). In this article, we are mostly concerned about financial data applications where extreme events and heteroscedasticity exist. Therefore, we propose a new method to estimate the tail index based on the $k - [n^\eta]$ upper expectiles $\widehat{\xi}_{n-k}, \ldots, \widehat{\xi}_m = \widehat{\xi}_{n-[n^\eta]}$ for the reason that expectile regression is more sensitive to those extreme values than quantile regression. The asymptotic results hold for any choice of m such that $m \le n - [n^\eta]$, in our simulation studies, we choose $\eta = 0.1$ and our experiments show that small values of $\eta \in (0, 0.25)$ work well. In the literature of extreme value theory, the selection of k is an important and challenging problem. k can be viewed as the effective sample size for tail extrapolation. A smaller k leads to estimators with larger variances, while larger k results in more bias. In practice, for the traditional tail index γ estimated by quantiles, a commonly used heuristic approach for choosing k is to plot the estimator of γ versus k (Hill plot) and then choose a suitable k corresponding to the first stable component of the plot, see Section 3 in de Haan and Ferreira (2006) and the references therein. However, for the recently proposed expectile-based tail index estimator, there are few studies in the literature offer constructive suggestions for the selection of k. Despite that, we still employ Hill plots for choosing k in our studies. Based on vast simulation experiments, we conclude that our proposed method for selecting k corresponding to the inflection points of the quantile-based and expectile-based Hill plots is more accurate than existing methods.

A risk measure that is commonly used as an alternative to VaR is ES. ES is favored by practitioners, who are mostly concerned with the risk of exposure to a catastrophic event that may wipe out an investment in terms of the size of the potential losses. The quantile-based risk measure of ES for $\tau \in [0, 1)$ is defined as follows (see Acerbi, 2002):

$$QES_\tau := \frac{1}{1-\tau}\int_\tau^1 q_t\,dt. \tag{3.11}$$

When Y is continuous, QES_τ is coherent and identical to the Conditional VaR (Rockafellar and Uryasev, 2000), known also as Tail Conditional Expectation (TCE), defined as $QTCE_\tau := E[Y|Y \geq q_\tau]$. Both QES_τ and $QTCE_\tau$ can then be explained as the average loss incurred in the tail event of a loss higher than the q_τ. However, QES_τ is a coherent risk measure while $QTCE_\tau$ isn't (Wirch and Hardy, 1999). Similarly to this intuitive tail conditional expectation, Taylor (2008) has introduced and used the expectile-based form $XTCE_\tau := E[Y|Y \geq \xi_\tau]$ as the basis for estimating the standard quantile-based measure $QTCE_\tau$. Because both $XTCE_\tau$ and $QTCE_\tau$ are not coherent risk measures in general, Daouia *et al.* (2018) suggested estimating the coherent expectile-based form of ES

$$XES_\tau := \frac{1}{1-\tau}\int_\tau^1 \xi_t\,dt, \tag{3.12}$$

obtained by substituting the expectile ξ_t for the quantile q_t in QES_τ. This definition is more credible than $XTCE_\tau$ as it induces a proper coherent risk measure, while maintaining the intuitive meaning of the conditional expectation. Due to the asymptotic equivalence of $XES_\tau \sim XTCE_\tau$ as $\tau \to 1$ (see Daouia *et al.*, 2018), the tail values XES_τ and $XTCE_\tau$ share the same estimators, for both intermediate levels $\tau = \tau_n$ and extreme expectile levels $\tau = \tau_n'$.

To estimate both the expectile-based and quantile-based forms of ES at a very high level that may approach one at an arbitrarily fast rate, Daouia *et al.* (2020b) suggested that the first step is to estimate these risk measures at an intermediate level ($\tau_n \to 1$ and $n(1-\tau_n) \to \infty$), and then extrapolate the resulting estimates to the far tail by making use of the traditional Hill estimator for the tail index γ after that. Here, we extend their method by replacing the traditional Hill estimator with our proposed expectile-based estimator (expectHill estimator) $\widehat{\gamma}$. The following asymptotic equation, which is the same as Proposition 1 in Daouia *et al.* (2020b), provides a guidance role in the estimation procedure.

Proposition 1. (*Daouia et al., 2020b*). *Assume that Y follows a Pareto-type distribution with tail index $0 < \gamma < 1$ and $Y_- := \min\{Y, 0\}$, where $E|Y_-| < \infty$. Then,*

$$\frac{XES_\tau}{QES_\tau} \sim \frac{\xi_\tau}{q_\tau} \sim \frac{E[Y|Y > \xi_\tau]}{E[Y|Y > q_\tau]} \quad \textit{and} \quad \frac{XES_\tau}{\xi_\tau} \sim \frac{1}{1-\gamma} \sim \frac{E[Y|Y > \xi_\tau]}{\xi_\tau}, \quad \tau \to 1.$$

Under the model assumptions that $E|Y_-| < \infty$ and Y has a heavy-tailed distribution (2.2), we want to estimate the value of the expectile-based form $XES_{\tau_n'}$, where $\tau_n' \to 1$ and $n(1-\tau_n') \to c < \infty$. From Proposition 1, we can estimate $XES_{\tau_n'}$ using the asymptotic equivalence $XES_{\tau_n'} \sim (1-\gamma)^{-1}\xi_{\tau_n'}$. By replacing γ and $\xi_{\tau_n'}$ with their estimators $\widehat{\gamma}$ and $\widehat{\xi}_{\tau_n'}$, which were described in (3.9) and (3.10), respectively, we define the estimator $XES_{\tau_n'}$ as follows:

$$\widehat{XES}_{\tau_n'} := (1-\widehat{\gamma})^{-1}\widehat{\xi}_{\tau_n'}. \tag{3.13}$$

Then, we return to the estimation of the usual form QES_{p_n} of quantile-based ES, for a pre-specified tail level $p_n \to 1$ with $n(1-p_n) \to c < \infty$. Here, we wish to derive alternative families of composite expectile-based estimators from $XES_{\tau_n'}$ introduced above, where $\tau_n' = \tau_n'(p_n)$ is chosen to make the establishment of $\xi_{\tau_n'} \equiv q_{p_n}$, so that the asymptotic equivalences $QES_{p_n} \sim E[Y|Y > q_{p_n}]$ and $XES_{\tau_n'} \sim E[Y|Y > \xi_{\tau_n'}]$ in Proposition 1 lead to $QES_{p_n} \sim XES_{\tau_n'(p_n)}$. In this way, QES_{p_n} inherits the extreme value estimator of $XES_{\tau_n'}$ itself, namely $\widehat{XES}_{\tau_n'(p_n)}$, which can be similarly computed by the same method as (3.13). Thus, a remaining task is to estimate the extreme level $\tau_n' = \tau_n'(p_n)$ such that $\xi_{\tau_n'} = q_{p_n}$. It has been found in Proposition 3 of Daouia *et al.* (2018) that such a level satisfies

$$1 - \tau_n'(p_n) \sim (1-p_n)\frac{\gamma}{1-\gamma} \quad \text{as} \quad n \to \infty,$$

under the model assumption of heavy tails with $0 < \gamma < 1$. Plugging in our expectHill estimator $\widehat{\gamma}$, we can estimate $\tau_n'(p_n)$ by

$$\widehat{\tau}_n'(p_n) := 1 - (1-p_n)\frac{\widehat{\gamma}}{1-\widehat{\gamma}}. \tag{3.14}$$

By replacing $\tau_n'(p_n)$ with $\widehat{\tau}_n'(p_n)$, these two estimators $\widehat{\xi}_{\widehat{\tau}_n'(p_n)}$ and $\widehat{XES}_{\widehat{\tau}_n'(p_n)}$ can be similarly calculated using the methods in (3.10) and (3.13), and they can be seen as the estimates of q_{p_n} and QES_{p_n}, respectively.

3.1. Asymptotic properties

Denote $\boldsymbol{Z} = (\boldsymbol{1}, \boldsymbol{X}^T)^T$, $\boldsymbol{z}_i = (1, \mathbf{x}_i)^T$, $\boldsymbol{\theta}(\tau) = (\alpha(\tau), \boldsymbol{\beta}(\tau)^T)^T$, and $\mu_{\boldsymbol{Z}} = E(\boldsymbol{Z})$. We now make further assumptions as follows:

(A1) The variable $\boldsymbol{z}_i$ has a compact support $\mathcal{Z}$, and $E(\boldsymbol{Z}\boldsymbol{Z}^T)$ is positive definite.

(A2) There exists an auxiliary line $z \to z^T\theta(r)$ with $0 < r < 1$ and a bounded vector $\theta(r)$ such that for $Y^* = Y - z^T\theta(r)$ and some Pareto-type distribution function $F_0(\cdot)$ with tail index $0 < \gamma < 1/2$,

$$\frac{1 - F_{Y^*}(t|z)}{\boldsymbol{K}(z)\{1 - F_0(t)\}} - 1 = \{1 - F_0(t)\}^{\delta}\widetilde{\boldsymbol{K}}(z)\{1 + o(1)\},$$

uniformly in $z \in \mathcal{Z}$ as $t \to \infty$, where $\boldsymbol{K}(\cdot)$ and $\widetilde{\boldsymbol{K}}(\cdot)$ are positive, continuous, and bounded functions on $\mathcal{Z}$ and $\delta > 0$ is a constant.

(A3) (a) $\frac{\partial}{\partial\tau}F_{Y^*}^{-1}(\tau|z) \sim \frac{\partial}{\partial\tau}F_0^{-1}\{\tau/\boldsymbol{K}(z)\}$ uniformly in $z \in \mathcal{Z}$ as $\tau \to 1$;
(b) $\frac{\partial}{\partial\tau}F_0^{-1}(1-\tau)$ is regularly varying at zero with index $-\gamma - 1$.

(A4) $U_0(t) = F_0^{-1}(1 - 1/t)$ satisfies the second-order condition (2.5) with $0 < \gamma < 1/2$, $\rho < 0$, and $A(t) = \gamma d t^{\rho}$ with $d \neq 0$.

Conditions A1–A4 are similar as Conditions B1–B4 in Wang *et al.* (2012). However, a minor change in the condition $0 < \gamma < 1/2$ is required to ensure that the asymmetric least squares estimators of ξ_τ are asymptotically Gaussian. Note that for U_0, the second-order condition (2.5) is equivalent to

$$U_0(t) = ct^{\gamma}\left[1 + \frac{A(t)}{\rho}\{1 + o(1)\}\right] \quad \text{as} \quad t \to \infty, \tag{3.15}$$

with some constant $c > 0$. We call c the corresponding constant. According to $A(t) = \gamma d t^{\rho}$ with $d \neq 0$ and some straightforward calculations, it follows that (3.15) is equivalent to

$$1 - F_0(x) = \left(\frac{x}{c}\right)^{-1/\gamma}\left[1 + \frac{d}{\rho}\left(\frac{x}{c}\right)^{\rho/\gamma}\{1 + o(1)\}\right] \quad \text{as} \quad x \to \infty, \tag{3.16}$$

for example, see Gomes and Pestana (2007).

Most conventional regression models are covered by these regularity conditions, for example, the location-scale shift regression model: $Y = \alpha + \mathbf{x}^T\boldsymbol{\beta} + (1 + \mathbf{x}^T\boldsymbol{\sigma})\epsilon$, where $\epsilon \sim F_0(\cdot)$ (see also the example in Wang *et al.*, 2012); for a given $\mathbf{x}$, by denoting $z = (1, \mathbf{x})^T$, $z^T\boldsymbol{\sigma} = 1 + \mathbf{x}^T\widetilde{\boldsymbol{\sigma}}$, $\boldsymbol{\theta} = (\alpha, \boldsymbol{\beta}^T)^T$, and $Y^* = Y - z^T\boldsymbol{\theta}$, we have $F_{Y^*}(y|z) = F_0\{y/(z^T\boldsymbol{\sigma})\}$. Using Taylor expansion and (3.16), we have

$$\begin{aligned}\frac{1 - F_{Y^*}(y|z)}{1 - F_0(y)} &= \frac{1 - F_0\{y/(z^T\boldsymbol{\sigma})\}}{1 - F_0(y)} \\ &= (z^T\boldsymbol{\sigma})^{(1/\gamma)}\left[1 + \frac{d}{\rho}\Big((z^T\boldsymbol{\sigma})^{-\rho/\gamma} - 1\Big)\left(\frac{y}{c}\right)^{\rho/\gamma}\Big(1 + o(1)\Big)\right],\end{aligned}$$

where c is a positive constant as defined in (3.15). Define $\boldsymbol{K}(z) = (z^T\boldsymbol{\sigma})^{(1/\gamma)}$ and $\widetilde{\boldsymbol{K}}(z) = \frac{d}{\rho}\{(z^T\boldsymbol{\sigma})^{-\rho/\gamma} - 1\}$. Then

$$\frac{1 - F_{Y^*}(y|z)}{\boldsymbol{K}(z)\{1 - F_0(y)\}} - 1 = \widetilde{\boldsymbol{K}}(z)\left(\frac{y}{c}\right)^{\rho/\gamma}(1 + o(1)).$$

Since $1 - F_0(y) \approx (y/c)^{-1/\gamma}$, Condition A2 holds with $\delta = -\rho$; thus

$$\frac{1 - F_{Y^*}(t|z)}{\boldsymbol{K}(z)\{1 - F_0(t)\}} - 1 = (1 - F_0(t))^{-\rho}\widetilde{\boldsymbol{K}}(z)(1 + o(1)). \tag{3.17}$$

In addition, $F_{Y^*}(y|z) = F_0\{y/(z^T\boldsymbol{\sigma})\}$ leads to $U_{Y^*}(t|z) = (z^T\boldsymbol{\sigma})U_0(t)$, and thus, Condition A3(a) holds.

We now provide the asymptotic properties of the proposed estimators.

Theorem 3.1. *Suppose that Model (3.7) and Conditions A1–A4 hold, $k^{-1/2}n^{\eta}\log k \to 0$, $\sqrt{k}(n/k)^{\widetilde{\rho}} \to C \geq 0$ as $n \to \infty$, $k = k_n \to \infty$, and $k/n \to 0$, where $\widetilde{\rho} = \max(-\gamma, \rho)$ and C is a constant. And the bias conditions $\sqrt{k}\widetilde{A}(n/k) \to \lambda_1 \in \mathbb{R}$ and $\sqrt{k}/q_{\tau_{n-k}} \to \lambda_2 \in \mathbb{R}$ are satisfied. Then, for any given $z = (1, \mathbf{x}^T)^T$, we have:*

$$\sqrt{k}(\widehat{\gamma}(\boldsymbol{x}) - \gamma) \xrightarrow{d} N(b_z, v_z), \tag{3.18}$$

with

$$b_z = \frac{(1-\gamma)(\gamma^{-1}-1)^{-\widetilde{\rho}}}{(1-\widetilde{\rho})(1-\gamma-\widetilde{\rho})}\lambda_1 - E(Y|z)\frac{\gamma^2(\gamma^{-1}-1)^{\gamma}}{\gamma+1}\lambda_2,$$

and

$$v_z = \frac{2\gamma^3}{1-2\gamma}z^T\boldsymbol{H}^{-1}\boldsymbol{\Sigma}\{\boldsymbol{K}(z)\}^{-2\gamma}\boldsymbol{H}^{-1}z,$$

where $\boldsymbol{H} = E[\{\boldsymbol{K}(\boldsymbol{Z})\}^{-\gamma}\boldsymbol{Z}\boldsymbol{Z}^T]$, $\boldsymbol{\Sigma} = E[\boldsymbol{Z}\boldsymbol{Z}^T]$ and$\widetilde{A}(\cdot)$ is defined below by (A.1) of the Appendix.

When the common index assumption is violated, the tail index $\widehat{\gamma} = \widehat{\gamma}(\mathbf{x})$ is expected to vary with $\mathbf{x}$. In Tail Index Regression (TIR) (see Wang and Tsai, 2009), the tail index is often assumed to be linear in $\mathbf{x}$ after some parametric link transformations are performed. Our estimator $\widehat{\gamma}(\mathbf{x})$ can be viewed as a nonparametric estimation of $\gamma(\mathbf{x})$ and thus can provide some guidance for the choice of the required link function required in TIR.

To derive the asymptotic normality of the conventional Hill estimator based on quantiles, with an asymptotic bias $\lambda_1/(1-\widetilde{\rho})$ and an asymptotic variance γ^2 (see Theorem 3 in Wang *et al.*, 2012), conditions involving the auxiliary function $\widetilde{A}(\cdot)$ in Theorem 3.1 are also required. Theorem 3.1 also features a further bias condition involving the quantile function $q_{\tau_{n-k}}$, and this is to be expected in view of Proposition 1 in Daouia *et al.* (2020b), of which a consequence is that the remainder term in the approximation $\xi_{1-(1-\tau_n)s}/\xi_{\tau_n} \sim s^{-\gamma}$ depends on both A and q_{τ_n}.

Theorem 3.2. *Suppose that the conditions of Theorem 3.1 hold,* $n(1-\tau_n')=o(k)$ *and* $\log(n(1-\tau_n'))=o(\sqrt{k})$. *Then,*

$$\frac{\sqrt{k}}{\log[k/(n(1-\tau_n'))]}\left(\frac{\widehat{\xi}_Y(\tau_n'|z)}{\xi_Y(\tau_n'|z)}-1\right)\xrightarrow{d} N(b_z, v_z), \tag{3.19}$$

and by replacing p_n *with* τ_n', *we have*

$$\frac{\sqrt{k}}{\log[k/(n(1-p_n))]}\left(\frac{\widehat{\xi}_Y(\widehat{\tau}_n'(p_n)|z)}{q_Y(p_n|z)}-1\right)\xrightarrow{d} N(b_z, v_z), \tag{3.20}$$

$$\frac{\sqrt{k}}{\log[k/(n(1-p_n))]}\left(\frac{\widehat{XES}_{\widehat{\tau}_n'(p_n)}}{QES_{p_n}}-1\right)\xrightarrow{d} N(b_z, v_z), \tag{3.21}$$

with (b_z, v_z) *as in Theorem 3.1.*

It is clear that the estimators in Theorems 3.1 and 3.2 share the same asymptotic behavior from a theoretical point of view.

4. Simulation study

We conduct simulation studies to investigate the performance of our proposed method by estimating the conditional quantile and ES at a given high level of p_n. The results show that the proposed estimates outperform the usual quantile regression estimates for high tails in terms of mean squared error (MSE). Moreover, our proposed method is superior to that of Wang *et al.* (2012) in some cases. For comparison, the data are generated from the same model as Wang *et al.* (2012), which is formulated as follows:

$$y_i = x_{i1} + x_{i2} + (1 + r x_{i1})e_i, \quad i = 1, \ldots, n, \tag{4.22}$$

where $x_{ij} \sim \text{Uniform}(-1, 1)$, $j = 1, 2$, $e_i's$ are independent and identically distributed random variables, and r is a constant that controls the degree of heteroscedasticity. Therefore, the τth conditional expectile of Y is

$$\xi_Y(\tau|\mathbf{x}_i) = \alpha(\tau) + \mathbf{x}_i^T \boldsymbol{\beta}(\tau),$$

where $\mathbf{x}_i = (x_{i1}, x_{i2})^T$, $\alpha(\tau) = \xi_e(\tau)$, $\boldsymbol{\beta}(\tau) = (1 + r\xi_e(\tau), 1)^T$, and $\xi_e(\tau)$ is the τth expectile of e_i. We consider three sample sizes, $n = 200, 500, 1000$, and two r

values, $r = 0$ and 0.9, so that the coefficients $\boldsymbol{\beta}(\tau)$ are constant in homoscedastic models with $r = 0$ and vary across τ in heteroscedastic models with $r = 0.9$.

We also choose three different models for generating $e_1, \ldots, e_n$. In models 1 and 2, the $e_i's$ are from Pareto distributions with the extreme value index parameters $\gamma = 0.2$ and 0.45, respectively. In model 3, the $e_i's$ are generated from the t distribution with four degrees of freedom so that the tail index parameter is $\gamma = 0.25$.

For each simulated dataset, we apply the proposed method to estimate q_{p_n} and QES_{p_n} at the high levels $p_n = 0.999$ and 0.9999, respectively. The number of repetitions of the Monte Carlo simulation is 300 for each scenario. We compare three estimation methods: the traditional quantile regression method (QR), the method by Wang *et al.* (2012) based on quantile regression without a common slope (NCS), and our proposed method based on expectile regression without a common slope (ENCS). For both NCS and ENCS methods, let $k = [cn^{1/3}]$ with $c = 4.5$, where n is the number of observations, and let $\eta = 0.1$ so that $m = n - [n^{0.1}]$. Our numerical investigation suggests that the tail index becomes stable around this choice of k and both of the NCS and ENCS methods clearly outperform QR for $c \in [3, 10]$.

Let us first take a quick look at the asymptotic variance comparison of NCS and ENCS methods. According to asymptotic theory,

$$\frac{\mathrm{Var(NCS)}}{\mathrm{Var(ENCS)}} \sim \frac{\gamma^2}{2\gamma^3/(1-2\gamma)} = \frac{1-2\gamma}{2\gamma},$$

holds, and this suggests that NCS method is likely to have a larger variance with $0 < \gamma < 0.25$, while ENCS method is likely to have a larger variance with $0.25 < \gamma < 0.5$. On the other hand, it is noticed that the asymptotic variance comparisons are not fully informative for the estimation of high quantiles and QES, because both bias and variance are important in areas where data are sparse. We focus on comparisons of the mean square errors (MSE) obtained from finite sample simulations in this article. Tables 1, 2, and 3 summarize the MSE of different estimators for $\widehat{\xi}_{\tilde{\tau}_n'(p_n)}/q_{p_n}$ under different scenarios at $\mathbf{x} = (0, 0)^T$ or $\mathbf{x} = (0.5, 0.5)^T$. Tables 4, 5, and 6 summarize the MSE of different estimators for $\widehat{XES}_{\tilde{\tau}_n'(p_n)}/QES_{p_n}$ under different scenarios at $\mathbf{x} = (0, 0)^T$ or $\mathbf{x} = (0.5, 0.5)^T$. The reason we consider the MSE of their quotients is that they represent the relative MSE, which are more convincing than the traditional absolute MSE.

The results in Tables 1, 2, and 3 suggest that the proposed NCS and ENCS methods provide more efficient estimations than that of QR in almost all cases. In Table 1, with relatively light tails, both NCS and ENCS perform well in terms of estimating high conditional quantiles, and NCS is slightly better than our proposed ENCS method. However, both of them are much better than QR. In Table 2, with relatively heavy-tailed distributions, NCS is superior to ENCS. One reason for this is that according to asymptotic theory, ENCS has a larger variance than NCS when we set $\gamma = 0.45 > 0.25$. Another potential

TABLE 1

MEAN SQUARED ERRORS ×100 OF DIFFERENT ESTIMATORS FOR $\widehat{\xi}_{\widetilde{\tau}_n'(p_n)}/q_{p_n}$ IN CASE 1 WITH PARETO(0.2) ERRORS, WHERE $p_n = 0.999$ AND 0.9999.

		$\mathbf{x} = (0, 0)^T$		$\mathbf{x} = (0.5, 0.5)^T$	
r	Method	0.999	0.9999	0.999	0.9999
			$n = 200$		
0	QR	7.85	18.41	10.07	24.09
	NCS	3.5	7.52	4.58	10.03
	ENCS	3.47	11.16	4.10	11.96
0.9	QR	10.69	29.98	9.68	27.05
	NCS	4.76	9.09	4.78	10.73
	ENCS	5.61	11.56	5.88	12.81
			$n = 500$		
0	QR	5.26	15.27	4.96	17.78
	NCS	2.06	4.6	2.63	6.68
	ENCS	2.09	6.84	2.62	7.89
0.9	QR	5.75	22.48	5.35	25.78
	NCS	2.52	5.51	2.67	8.87
	ENCS	2.77	7.49	2.83	10.84
			$n = 1000$		
0	QR	4.13	10.65	3.89	10.35
	NCS	1.21	2.45	1.51	4.21
	ENCS	1.23	2.75	1.52	4.24
0.9	QR	4.25	11.76	4.13	18.27
	NCS	1.44	3.11	1.67	5.45
	ENCS	1.47	4.15	1.73	6.87

Note: QR is the conventional quantile regression method, NCS is the method proposed by Wang *et al.* (2012) assuming non-common slopes based on conditional quantiles, and ENCS is our proposed method based on conditional expectiles.

cause is the mean value of a Pareto distribution is not equal to zero, so ENCS is more biased than NCS based on the asymptotic theory. Moreover, a heavier tail means more outliers and expectile regression is more sensitive to these outliers than quantile regression, so it brings a larger variance. In Table 3, with $t(4)$ distribution of tail index $\gamma = 0.25$, NCS and ENCS have the same asymptotic error, and the mean value of $t(4)$ is zero. The results show that our proposed ENCS is best, and suggest that our proposed method is more suitable for financial data analysis because it is now commonly accepted that financial asset returns are, in fact, heavy-tailed with zero mean. In general, the NCS and ENCS methods have a larger MSE when $n = 200$, but the errors start to shrink as the sample size increases.

Tables 4, 5, and 6 show the MSE of the $\widehat{XES}_{\widetilde{\tau}_n'(p_n)}/QES_{p_n}$ estimator by using the NCS and ENCS methods, respectively. For comparison, we first use the method of Wang *et al.* (2012) to estimate the extreme level quantile $\widetilde{q}_{p_n}$, and

TABLE 2

MEAN SQUARED ERRORS ×100 OF DIFFERENT ESTIMATORS FOR $\widehat{\xi}_{\widetilde{\tau}_n'(p_n)}/q_{p_n}$ IN CASE 2 WITH PARETO(0.45) ERRORS, WHERE $p_n = 0.999$ AND 0.9999.

		$\mathbf{x} = (0, 0)^T$		$\mathbf{x} = (0.5, 0.5)^T$	
r	Method	0.999	0.9999	0.999	0.9999
		$n = 200$			
0	QR	19.29	22.27	20.96	37.58
	NCS	10.74	12.60	11.63	13.68
	ENCS	15.58	17.54	19.62	23.89
0.9	QR	21.11	32.75	25.35	43.78
	NCS	11.54	15.51	16.61	18.77
	ENCS	15.39	19.49	19.15	22.65
		$n = 500$			
0	QR	12.13	15.27	14.12	19.69
	NCS	5.16	8.63	6.54	7.66
	ENCS	8.07	9.88	7.61	10.59
0.9	QR	15.87	24.38	20.15	26.79
	NCS	5.47	9.11	8.14	8.94
	ENCS	7.91	11.58	10.59	12.82
		$n = 1000$			
0	QR	6.42	13.25	8.13	13.27
	NCS	2.20	2.79	2.60	5.23
	ENCS	3.41	3.88	3.76	7.32
0.9	QR	10.2	14.83	12.52	20.12
	NCS	3.47	3.68	4.03	6.16
	ENCS	4.79	5.75	6.41	8.77

Note: QR is the conventional quantile regression method, NCS is the method proposed by Wang *et al.* (2012) assuming non-common slopes based on conditional quantiles, and ENCS is our proposed method based on conditional expectiles.

then use $\widetilde{q}_{p_n}/(1-\widetilde{\gamma})$ to estimate $\widehat{QES}_{p_n}$, where $\widetilde{\gamma}$ is the tail index estimated based on conditional quantiles. Not surprisingly, the MSE results in Tables 4, 5, and 6 have the same properties as those in Tables 1, 2, and 3. In Tables 4 and 5, for Pareto distributions of light- or high- tailed indices, NCS is better than ENCS. However, in Table 6, with the t distribution, ENCS performs better.

Our empirical studies suggest that the proposed method is not worse than that of Wang *et al.* (2012) for estimating quantiles and QES at extreme levels where the tail index is small, but it performs better than NCS when the t distribution is used. Inspired by this advantage, it is suitable for us to apply the proposed approach to financial data analysis, and this is in accordance with our initial goal. The method in this paper should be used in larger sample problems, especially when the data exhibit clear heteroscedasticity in the tail. Our estimates are better than those of QR in general, and it is consistent with what we imagine at an extremely high level.

TABLE 3

MEAN SQUARED ERRORS ×100 OF DIFFERENT ESTIMATORS FOR $\widehat{\xi}_{\widehat{\tau}_n'(p_n)}/q_{p_n}$ IN CASE 3 WITH $t(4)$ ERRORS, WHERE $p_n = 0.999$ AND 0.9999.

		$\mathbf{x} = (0, 0)^T$		$\mathbf{x} = (0.5, 0.5)^T$	
r	Method	0.999	0.9999	0.999	0.9999
			$n = 200$		
0	QR	16.91	34.02	17.52	41.13
	NCS	8.12	20.78	8.87	22.82
	ENCS	5.18	11.86	5.99	12.19
0.9	QR	17.34	44.73	19.69	44.93
	NCS	12.88	19.71	14.76	20.03
	ENCS	8.42	10.54	8.89	10.19
			$n = 500$		
0	QR	10.35	15.09	12.74	21.57
	NCS	4.76	11.48	5.02	12.43
	ENCS	2.69	5.98	3.04	6.77
0.9	QR	11.64	16.85	13.32	24.97
	NCS	4.25	9.96	4.97	11.15
	ENCS	2.38	5.72	2.96	5.49
			$n = 1000$		
0	QR	5.01	6.83	5.77	8.37
	NCS	2.34	5.72	2.41	6.19
	ENCS	1.28	2.47	1.48	3.09
0.9	QR	6.32	7.89	6.95	11.26
	NCS	2.21	5.58	2.39	6.08
	ENCS	1.14	2.33	1.18	2.76

Note: QR is the conventional quantile regression method, NCS is the method proposed by Wang *et al.* (2012) with non-common slopes based on conditional quantiles, and ENCS is our proposed method based on conditional expectiles.

5. REAL DATA ANALYSIS

To illustrate the practical usefulness of our proposed extrapolation method, we consider the daily data of two publicly traded Chinese firms China Life (CL) and the Bank of China (BOC), respectively. CL is one of the biggest and most well-known insurance corporations in China, and BOC is one of the biggest banks in China. Both of them belong to the Shanghai Stock Exchange. The data are downloaded through R package *quantmod* from Yahoo Finance and cover the period from January 13, 2012 to January 5, 2018, which contains the turbulence in Chinese stock market from June 2015 to February 2016. The sample has 1453 observations which are calculated as the difference of the log transformation of the price, that is, $y_t = \log(P_t/P_{t-1})$, where P_t is the daily price. Our computations focus on $-y_t$ so that extremely high levels of VaR

TABLE 4

MEAN SQUARED ERRORS $\times 100$ OF DIFFERENT ESTIMATORS FOR $\widehat{XES}_{\hat{\tau}_n'(p_n)}/QES_{p_n}$ IN CASE 1 WITH PARETO(0.2) ERRORS, WHERE $p_n = 0.999$ AND 0.9999.

		$\mathbf{x} = (0, 0)^T$		$\mathbf{x} = (0.5, 0.5)^T$	
r	Method	0.999	0.9999	0.999	0.9999
			$n = 200$		
0	NCS	5.10	9.82	7.29	13.62
	ENCS	6.27	15.41	7.77	16.96
0.9	NCS	7.12	11.87	7.40	14.35
	ENCS	8.86	15.56	9.47	17.98
			$n = 500$		
0	NCS	3.26	6.24	4.44	9.5
	ENCS	4.12	10.2	4.66	12.16
0.9	NCS	3.99	7.51	4.55	10.81
	ENCS	4.85	10.88	5.27	12.86
			$n = 1000$		
0	NCS	2.89	3.57	3.05	4.79
	ENCS	3.75	4.61	3.81	5.18
0.9	NCS	2.85	3.26	3.14	4.86
	ENCS	3.24	4.12	3.86	5.24

Note: NCS is the method proposed by Wang *et al.* (2012) with non-common slopes based on conditional quantiles and ENCS is our proposed method based on conditional expectiles.

and QES can be used to measure extremely high levels of loss. We start the analysis with the following model:

$$y_t = \alpha(\tau) + \boldsymbol{M}_{t-1}\boldsymbol{\beta}(\tau) + \epsilon_t,$$

where $\boldsymbol{M}_{t-1}$ is a lagged macroeconomic state vector, similar to those suggested by Fang *et al.* (2018): (i) the volatility index, VIX, which is defined as the average of the squared daily return of the SSE (Shanghai Stock Exchange) 50 index with a weekly frequency; (ii) the liquidity spread, LS, which is defined as the difference between the 3-month collateral repo rate and the 3-month treasury bill rate; (iii) the spread term, ST, which is defined as the difference between the 10-year treasury bill rate and the 3-month treasury bill rate; (iv) credit spread, CS, which is defined as the difference between the 10-year BAA rated bond and the treasury bill rate.

In practice, the first objective is to test whether the slope estimates $\widehat{\boldsymbol{\beta}}(\tau)$ vary across τ at high levels. Therefore, we apply the analysis of variance (ANOVA) test developed by Newey and Powell (1987) and consider the special case of common slopes (the null hypothesis) at $\tau_j, j = 1, \ldots, 10$, where $\tau_1, \ldots, \tau_{10}$ are equally spaced between 0.95 and 0.99. The p-value of the ANOVA test are

TABLE 5

MEAN SQUARED ERRORS ×100 OF DIFFERENT ESTIMATORS FOR $\widehat{XES}_{\hat{\tau}_n'(p_n)}/QES_{p_n}$ IN CASE 2 WITH PARETO(0.45) ERRORS, WHERE $p_n = 0.999$ AND 0.9999.

		$\mathbf{x} = (0, 0)^T$		$\mathbf{x} = (0.5, 0.5)^T$	
r	Method	0.999	0.9999	0.999	0.9999
			$n = 200$		
0	NCS	25.03	29.11	35.80	41.21
	ENCS	28.85	34.89	43.92	48.25
0.9	NCS	28.15	32.49	46.41	48.12
	ENCS	37.84	44.59	55.43	57.64
			$n = 500$		
0	NCS	13.56	16.43	14.48	19.50
	ENCS	17.72	19.12	18.96	21.43
0.9	NCS	13.89	17.48	16.99	20.87
	ENCS	14.85	20.59	19.67	24.93
			$n = 1000$		
0	NCS	4.88	5.61	5.07	7.11
	ENCS	6.79	8.63	7.84	10.17
0.9	NCS	5.87	7.02	7.21	8.87
	ENCS	8.25	10.92	9.74	12.59

Note: NCS is the method proposed by Wang *et al.* (2012) with non-common slopes based on conditional quantiles and ENCS is our proposed method based on conditional expectiles.

6.5×10^{-14} for CL and 4.3×10^{-73} for BOC, suggesting that the slope coefficients $\widehat{\boldsymbol{\beta}}(\tau)$ of both the CL and BOC datasets are significantly different at upper levels $\tau \in [0.95, 0.99]$. Another issue is to select a reasonable value of k.

A commonly used heuristic approach for choosing k in the EVT literature is to plot the estimates of γ versus k and then choose a suitable k corresponding to the first stable component of the plot. However, there are few literature that discuss that how to choose a suitable k for the proposed method based on conditional expectiles. Following previous opinions, we also plot the estimates of γ as a function of k and the results are presented in Figure 1. For illustration, in Figure 1, when the macroeconomic state variable is given at $\overline{\boldsymbol{M}} = \frac{1}{T-1}\sum_{t=1}^{T-1} \boldsymbol{M}_t$, the NCS estimates of $\widetilde{\gamma}$ for CL (left) are relatively stable when k is between 95 and 100 and the $\widehat{\gamma}$ estimated by the ENCS method also become relatively stable in this interval and are close to $\widetilde{\gamma}$. Thus, we choose $k = 98$ for the CL data. Similarly, as we can see in the right plot of the BOC data in Figure 1, the first stable interval is [65, 75] where the ENCS estimates of $\widehat{\gamma}$ are very close to the NCS estimates. Thus, the chosen $k = 70$ is reasonable for the BOC data.

As the mainstream approaches for measuring the risks of financial firms, VaR and ES are widely used and well explained in numerous applications. In this article, we aim to estimate VaR and ES at an extremely high level. After

TABLE 6

MEAN SQUARED ERRORS ×100 OF DIFFERENT ESTIMATORS FOR $\widehat{XES}_{\widetilde{\tau}_n'(p_n)}/QES_{p_n}$ IN CASE 3 WITH $t(4)$ ERRORS, WHERE $p_n = 0.999$ AND 0.9999.

r	Method	$\mathbf{x} = (0, 0)^T$		$\mathbf{x} = (0.5, 0.5)^T$	
		0.999	0.9999	0.999	0.9999
			$n = 200$		
0	NCS	15.43	40.01	18.54	45.43
	ENCS	8.03	14.88	9.36	17.82
0.9	NCS	19.5	35.87	25.63	36.04
	ENCS	14.61	15.31	17.67	15.57
			$n = 500$		
0	NCS	8.67	19.28	9.64	25.76
	ENCS	4.96	9.93	4.77	10.83
0.9	NCS	8.56	18.92	11.78	21.09
	ENCS	4.67	9.39	4.24	9.79
			$n = 1000$		
0	NCS	4.48	10.13	5.42	12.73
	ENCS	2.77	4.96	2.85	5.12
0.9	NCS	4.51	9.18	5.09	11.74
	ENCS	2.32	4.57	2.48	4.68

Note: NCS is the method proposed by Wang *et al.* (2012) with non-common slopes based on conditional quantiles and ENCS is our proposed method based on conditional expectiles.

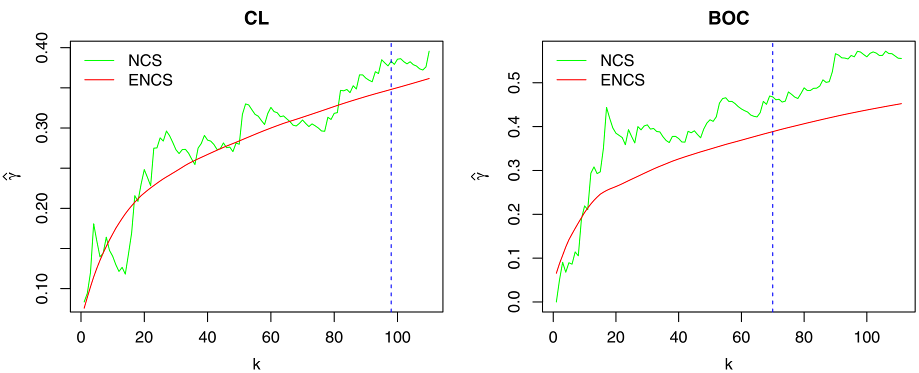

FIGURE 1: The plot of the estimates of $\widehat{\gamma}$ as a function of k, for CL (left) and BOC (right) with both the NCS (green) and ENCS (red) methods. Note: The macroeconomic state variables of the two plots are set as $\overline{\boldsymbol{M}} = \frac{1}{T-1}\sum_{t=1}^{T-1}\boldsymbol{M}_t$. Throughout the two figures, NCS estimated $\widetilde{\gamma}$ for CL are first stable when k is between 95 and 100, while those for BOC are first stable when k is between 65 and 75. In addition, ENCS estimated $\widehat{\gamma}$ of CL and BOC are relatively close to that NCS estimates. Thus, we choose $k = 98$ and $k = 70$ for CL and BOC, respectively.

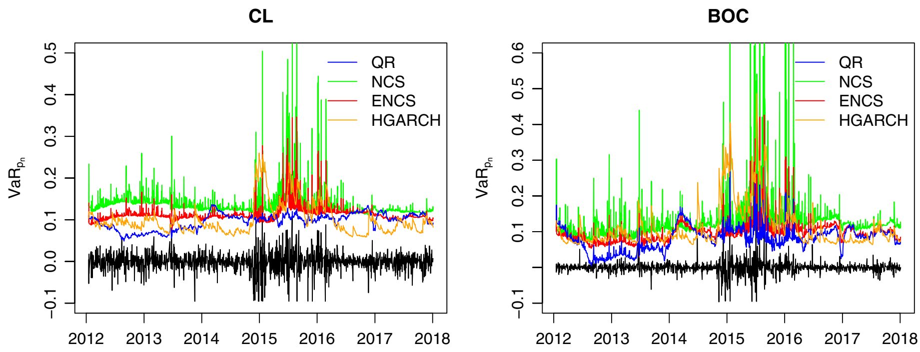

FIGURE 2: The estimated VaR of CL (left) and BOC (right) at the extremely high level $p_n = 0.999$. Note: The black line represents the original data,the blue, green, red, and orange lines represent the extremely high-level VaR estimates obtained by traditional quantile regression (QR), NCS, ENCS, and HGARCH, respectively.

determining the value of k, we can use the estimated tail index $\widehat{\gamma}$ to extrapolate our results to extremely high level based on (3.10), (3.13), and (3.14). In this analysis, we focus on the extreme level $p_n = 0.999$. Figure 2 presents the estimated results of $\widehat{\mathrm{VaR}}_{p_n}$ or equally the $\widehat{\xi}_{\widehat{\tau}_n'(p_n)}$ of CL (left) and BOC (right). For comparison purposes, we consider three different methods QR (blue line), NCS (green line), and ENCS (red line). We also consider GARCH models with heavy-tailed innovations (see Chan *et al.*, 2007). To distinguish it from the general GARCH model we denoted it by HGARCH (orange line). For simplicity, we consider the following HGARCH(1,1) model:

$$y_t = \sigma_t \epsilon_t, \quad \sigma_t^2 = c + b y_{t-1}^2 + a \sigma_{t-1}^2, \tag{5.23}$$

where ϵ_t follows a Pareto-type distribution, and $c > 0, b > 0, a > 0, b + a < 1$. Once $\widehat{\sigma}_t$ have been estimated by model (5.23), then $\widehat{\epsilon}_t$ can be estimated by $\widehat{\epsilon}_t = y_t/\widehat{\sigma}_t$. By the assumption that ϵ_t in (5.23) have heavy tails, the relative tail index is estimated by the Hill estimator

$$\widehat{\gamma} = \frac{1}{k} \sum_{i=1}^{k} \log \frac{\widehat{\epsilon}_{n,n-i+1}}{\widehat{\epsilon}_{n,n-k}},$$

where $\widehat{\epsilon}_{n,1} \le \cdots \le \widehat{\epsilon}_{n,n}$ denote the order statistics of $\widehat{\epsilon}_t$, and k is similarly selected based on the Hill plot. Given an extreme level p_n, the p_n-quantile of ϵ_t can be estimated by

$$\widehat{VaR}_{\epsilon,p_n} = \left(\frac{k}{n(1-p_n)} \right)^{\widehat{\gamma}} \widehat{\epsilon}_{n,n-k}.$$

Then, the p_n-quantile of y_t can be estimated by

$$\widehat{VaR}_{y_t,p_n} = \widehat{\sigma}_t \widehat{VaR}_{\epsilon,p_n}.$$

The black line represents the original data, and it shows that extreme values mainly occur during 2015–2016, the period of Chinese stock market turbulence. When we focus on the left plot for the CL data in Figure 2, the VaR estimated at an extremely high level of CL by QR is usually underestimated during the whole period, and this will cause the company to have inadequate risk resistance. However, NCS method (green line) is usually above the QR method (blue line), especially during crisis period, so NCS may be rejected by investors because it is too conservative and reduces revenue too much margin is set to withstand risks. The result of our proposed ENCS (red line) method is between the lines of QR and NCS for most of the period and can provide guidance for risk management and investors. The estimated values obtained by the HGARCH method are relatively small during the steady economic period, but extremely large during the crisis period. Similarly, we can estimate the extremely high VaR of the BOC by the same way as that of CL, and the results are shown in the right plot of Figure 2. It can be seen that NCS and ENCS deliver similar pattern as VaR estimates, but both methods estimate higher risks than those of QR in most cases. However, the NCS and HGARCH methods sometimes estimate more extreme values of VaR than those of ENCS method during the crisis period. To quantify the conservatism of each method, for each method we also report the realized number of violations (NV)

$$NV = \sum_{t=1}^{n} \mathbb{I}(y_t > VaR_{\tau,t}),$$

where $VaR_{\tau,t}$ is the quantile estimator and the corresponding asymmetric piecewise linear score computed by averaging

$$S(VaR_\tau, \boldsymbol{Y}) = \frac{1}{n}\sum_{t=1}^{n} S(VaR_{\tau,t}, y_t),$$

where

$$S(VaR_{\tau,t}, y_t) = \begin{cases} (1-\tau)(VaR_{\tau,t} - y_t), & \text{if} \quad y_t \le VaR_{\tau,t}, \\ \tau(y_t - VaR_{\tau,t}), & \text{if} \quad VaR_{\tau,t} < y_t. \end{cases}$$

The results are presented in Table 7. Since the sample size is 1453 and the extreme level of $p_n = 0.999$, ENCS is the most accurate method in which model the number of violations is just between 1 and 2. From the averaging asymmetric piecewise linear score method, the QR method is usually the least conservative, which could underestimate the potential risk. On the contrary, the NCS method is usually the most conservative that not accepted by investors. While the ENCS method steers a middle case between QR and NCS method, which could be more attractive in practice.

With regard to estimating ES at an extremely high level, there is no straightforward regression method for achieving this goal. Thus, we only take NCS

TABLE 7

THE NUMBER OF VIOLATIONS (NV) AND THE AVERAGING ASYMMETRIC PIECEWISE LINEAR SCORE $S(VaR_\tau, \boldsymbol{Y}) \times 1000$ FOR EACH METHODS OF CL AND BOC, RESPECTIVELY.

	CL		BOC	
Method	NV	$S(VaR_\tau, \boldsymbol{Y})$	NV	$S(VaR_\tau, \boldsymbol{Y})$
QR	4	0.097	0	0.081
NCS	0	0.139	0	0.132
ENCS	1	0.118	2	0.107
HGARCH	0	0.101	1	0.116

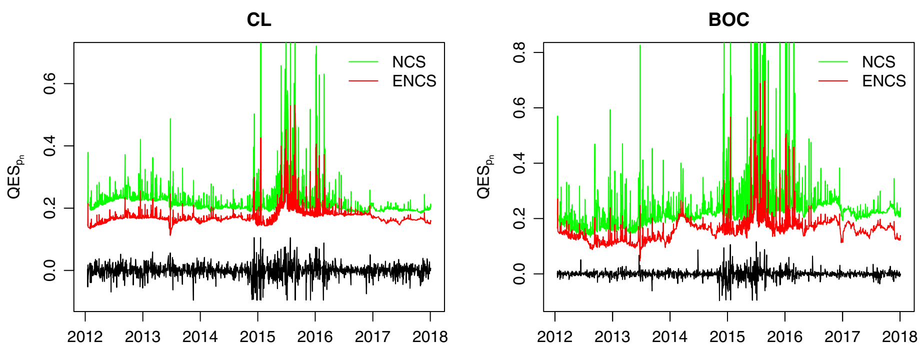

FIGURE 3: The Expected Shortfalls (QES) of CL (left) and BOC (right) are at the extremely high level $p_n = 0.999$. Note: The black line represents the original data, the green and red lines represent the extremely high-level QES estimated by NCS and ENCS methods, respectively.

and ENCS for our comparison. Figure 3 presents the estimations of $\widehat{QES}_{p_n}$ or equally $\widehat{XES}_{\widehat{\tau}'_n(p_n)}$ at $p_n = 0.999$ for CL (left) and BOC (right). The left plot in Figure 3 shows that the NCS method estimates a larger QES for CL than the ENCS method does during most of the sample period. When we examine the right plot in Figure 3, similar as that of CL estimates, the green line is usually above the red line, which means that the NCS method estimates a larger QES for BOC than the ENCS method does during most of the sample period.

Overall, our proposed ENCS method is not worse than the NCS method proposed by Wang *et al.* (2012) for estimating extremely high conditional VaR and ES. Of course, both NCS and ENCS methods perform better than QR in the high-tailed risk analysis.

6. DISCUSSION

Estimating extreme tails is difficult, if not infeasible, without any distributional assumptions on the tails. In this article, we develop a new method for high conditional expectile estimation by assuming that the error distribution is in the maximum domain of attraction of a Pareto-type distribution. This assumption simplifies the complex conditions required for the proof of asymptotic theory and makes it possible to extrapolate expectile estimates from intermediate levels of the data range to the high end. We rigorously establish the consistency and asymptotic normality of the proposed extreme value index estimators and the extrapolated high conditional expectile estimators. Numerical studies show that extrapolation leads to more accurate conditional quantile and ES at high tails than those obtained by conventional quantile regression for heavy-tailed distributions. Moreover, our proposed ENCS method performs better than the NCS method in zero-mean, heavy-tailed error distributions, such as the t distribution.

For the financial analysis conducted in Section 5, the challenging work is to select a suitable value of k. Like many other extreme value studies suggest, we plot the estimates of γ as a function of k and choose the first relatively stable component of the plot. Through a comparison of the results of the two stock datasets, we suggest an alternative method for selecting k, that is, to select the intersection of the conventional Hill plot and the expectile-based Hill plot. In most cases, the estimated value is close to those obtained by quantile-based estimators over the first stable interval. In our analysis, the ENCS method is no worse than NCS method proposed by Wang *et al.* (2012) for estimating extremely high conditional VaR and ES values. However, sometimes our method provides more eclectic opinions than those of NCS for both risk managers and investors.

The estimation of the standard error of any high quantile or expectile estimates can be challenging. Although the asymptotic variance can be calculated by Theorem 3.2, the asymptotic variance of the proposed ENCS method is difficult to estimate due to its complicated structure. Whether other methods such as the bootstrap can provide decent approximations needs further investigation. Another inadequacy of the proposed method is that there is additional bias induced when the conditional mean is not zero. There is also a trade-off between bias and variance with regard to the complexity of the tail index as a function of $\mathbf{x}$. Due to the data sparsity in the tails of a distribution, a constant tail index is often realistic and helpful in many situations. Additional research is needed to evaluate how to strike a better balance between the bias and variance when there is clear evidence against the common index assumption.

Acknowledgments

The work is supported by the National Key Research and Development Plan (No. 2016YFC0800104) and the National Science Foundation of China (No. 71771203).

REFERENCES

ACERBI, C. (2002) Spectral measures of risk: a coherent representation of subjective risk aversion. *Journal of Banking and Finance*, **26**, 1505–1518.

ARTZNER, P., DELBAEN, F., EBER, J.-M. and HEATH, D. (1999) Coherent measures of risk. *Mathematical Finance*, **9**, 203–228.

BELLINI, F., KLAR, B., MÜLLER, A. and ROSAZZA GIANINA, E. (2014) Generalized quantiles as risk measures. *Insurance: Mathematics and Economics*, **54**, 41–48.

BELLINI, F. and DI BERNARDINO, E. (2017). Risk management with expectiles. *The European Journal of Finance*, **23**, 487–506

CHAN, N.H., DENG, S. J., PENG, L. and XIA, Z. (2007) Interval estimation of value-at-risk based on GARCH models with heavy-tailed innovations. *Journal of Econometrics* **137**(2), 556–576.

CAI, Z., FANG, Y. and TIAN, D. (2018). Assessing tail risk using expectile regressions with partially varying coefficients. *Journal of Management Science and Engineering*, **3**, 183–213.

DE HAAN, L. and FERREIRA, A. (2006) *Extreme Value Theory: An Introduction*. New York: Springer.

DAOUIA, A., GIRARD, S. and STUPFLER, G. (2018) Estimation of tail risk based on extreme expectiles. *Journal of the Royal Statistical Society, Series B*, **80**, 263–292.

DAOUIA, A., GIRARD, S. and STUPFLER, G. (2020a) Tail expectile process and risk assessment. *Bernoulli*, **26**(1), 531–556.

DAOUIA, A., GIRARD, S. and STUPFLER, G. (2020b) ExpectHill estimation, extreme risk and heavy tails. *Journal of Econometrics*, preprint.

FANG, L.B., SUN, B.Y., LI, H.J. and YU, H. (2018) Systemic risk network of chinese financial institutions. *Emerging Markets Review*, **35**, 190–206.

GOMES, M. and PESTANA, D. (2007) A simple second-order reduced bias tail index estimator. *Journal of Statistical Computation and Simulation*, **77**, 487–504.

HILL, B.M. (1975) A simple general approach to inference about the tail of a distribution. *The Annals of Statistics*, **3**, 1163–1174.

KOENKER, R. and BASSETT, G. (1978) Regression quantiles. *Econometrica*, **46**, 33–50.

KUAN, C.-M., YEH, J.-H. and HSU, Y.-C. (2009) Assessing value at risk with CARE, the Conditional Autoregressive Expectile models. *Journal of Econometrics*, **2**, 261–270.

LI, D., PENG, L. and YANG, J. (2010) Bias reduction for high quantiles. *Journal of Statistical Planning and Inference*, **140**, 2433–2441.

NEWEY, W.K. and POWELL, J.L. (1987) Asymmetric least squares estimation and testing. *Econometrica*, **55**, 819–847.

ROCKAFELLAR, R.T. and URYASEV, S. (2000) Optimization of conditional value-at-risk. *Journal of Risk*, **2**, 21–42.

SOBOTKA, F., RADICE, R., MARRA, G. and KNEIB, T. (2013a) Estimating the relationship between women's education and fertility in Bostwana by using and instumental variable approach to semiparametric expectile regression. *Journal of the Royal Statistical Society*, **62**, 25–45.

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

WEISSMAN, I. (1978) Estimation of parameters and large quantiles based on the k largest observations. *Journal of the American Statistical Association*, **73**, 812–815.

WIRCH, J.L. and HARDY, M.R. (1999) A synthesis of risk measures for capital adequacy. *Insurance: Mathematics and Economics*, **25**, 337–347.

WANG, H. and TSAI, C.L. (2009) Tail index regression. *Journal of the American Statistical Association*, **104**, 1233–1240.

WANG, H.J., LI, D. and HE, X. (2012) Estimation of high conditional quantiles for heavy-tailed distributions. *Journal of the American Statistical Association*, **107**, 1453-1464.

YAO, Q. and TONG, H. (1996) Asymmetric least squares regression estimation: a nonparametric approach. *Journal of Nonparametric Statistics*, **6**, 273–292.

JIE HU
Department of Statistics and Finance, School of Management
University of Science and Technology of China, Hefei, China

YU CHEN (Corresponding author)
Department of Statistics and Finance, School of Management
University of Science and Technology of China, Hefei, China
E-Mail: cyu@ustc.edu.cn

KEQI TAN
Department of Statistics and Finance, School of Management
University of Science and Technology of China, Hefei, China

APPENDIX A. PROOFS OF THEOREMS

Lemma 1. *Condition A3 and Equation(3.17) imply that* $U_{Y^*}(t|\boldsymbol{z}) = F_{Y^*}^{-1}(1-1/t|\boldsymbol{z})$ *satisfies the second-order condition (2.5) with index* (γ, ρ, A^*), *where* $0 < \gamma < 1/2$, *and* $A^*(t) = \gamma d^* t^\rho$, $d^* = \{\boldsymbol{K}(\boldsymbol{z})\}^\rho\{d + \rho\widetilde{\boldsymbol{K}}(\boldsymbol{z})\}$. *Then,* $U_Y(t|\boldsymbol{z}) = F_Y^{-1}(1-1/t|\boldsymbol{z})$ *satisfies the second-order condition (2.5) with index* $(\gamma, \widetilde{\rho}, \widetilde{A})$, *where*

$$\widetilde{A}(t) = \gamma \widetilde{d} t^{\widetilde{\rho}}, \qquad \widetilde{\rho} = \max(-\gamma, \rho), \tag{A.1}$$

$\widetilde{d} = d^*\mathbb{I}(\rho \geq -\gamma) - c^{*-1}\boldsymbol{z}^T\boldsymbol{\theta}(r)\mathbb{I}(\rho \leq -\gamma)$, $c^* = c\{\boldsymbol{K}(\boldsymbol{z})\}^\gamma$ *and* $c > 0$ *is a constant.*

Proof of Lemma 1. This lemma has been proved in Wang *et al.* (2012). □

Lemma 2. *Suppose Model (3.6) and Conditions A1–A4 hold. Define* $\mathcal{T} = \{\tau_{n-k} < \cdots < \tau_m\}$ *with* $m = n - [n^\eta]$ *for* $0 < \eta < 1$, $\tau_j = j/(n+1)$ *for* $j = n-k, \ldots, m$, *and* $k > [n^\eta]$, $k/n \to 0$. *then we have*

$$\frac{\sqrt{n(1-\tau)}}{\xi_{0,\tau}}\{\widehat{\boldsymbol{\theta}}(\tau) - \boldsymbol{\theta}(\tau)\} \xrightarrow{d} N\left(0, \frac{2\gamma^3}{1-2\gamma}\boldsymbol{H}^{-1}\boldsymbol{\Sigma}\boldsymbol{H}^{-1}\right),$$

uniformly for $\tau \in \mathcal{T}$, *where* $\xi_{0,\tau}$ *is the* τ*th expectile of* F_0, $\boldsymbol{H} = E\left[\{\boldsymbol{K}(\boldsymbol{Z})\}^{-\gamma}\boldsymbol{Z}\boldsymbol{Z}^T\right]$ *and* $\boldsymbol{\Sigma} = E(\boldsymbol{Z}\boldsymbol{Z}^T)$ *are positive definite matrices.*

Proof of lemma 2. Define

$$M_n(\boldsymbol{u}, \tau) = \sum_{i=1}^n [\varrho_\tau(y_i - \boldsymbol{z}_i^T\boldsymbol{\theta}(\tau) - \boldsymbol{z}_i^T\boldsymbol{u}/\sqrt{n}) - \varrho_\tau(y_i - \boldsymbol{z}_i^T\boldsymbol{\theta}(\tau))].$$

Then, denote the first and second derivatives by $\varrho_\tau(y_i - \boldsymbol{z}_i^T\boldsymbol{\theta}(\tau) - t)$ at $t = 0$ as follows:

$$g_\tau(y_i^*) = \varrho_\tau'(y_i^* - t)|_{t=0} = -2[\tau y_i^*\mathbb{I}(y_i^* \geq 0) + (1-\tau)y_i^*\mathbb{I}(y_i^* < 0)],$$

and

$$h_\tau(y_i^*) = \varrho_\tau''(y_i^* - t)|_{t=0} = 2[\tau\mathbb{I}(y_i^* \geq 0) + (1-\tau)\mathbb{I}(y_i^* < 0)].$$

Therefore, we can decompose $M_n(\boldsymbol{u}, \tau)$ by Taylor expansion,

$$M_n(\boldsymbol{u}, \tau) = \sum_{i=1}^{n}\left[-g_\tau(y_i^*)\frac{\boldsymbol{z}_i^T\boldsymbol{u}}{\sqrt{n}} + \frac{h_\tau(y_i^*)}{2}\left(-\frac{\boldsymbol{z}_i^T\boldsymbol{u}}{\sqrt{n}}\right)^2 + o\left(\frac{1}{n}\right)\right],$$

where $y_i^* = y_i - \boldsymbol{z}_i^T\boldsymbol{\theta}(\tau)$. Denote $W_n(\tau) = \sum_{i=1}^n g_\tau(y_i^*)\frac{z_i}{\sqrt{n}}$, $\mu_{h_\tau} = E(h_\tau(y_i^*))$. Then we calculate $Var(W_n(\tau))$ and μ_{h_τ} based on heavy-tail distribution assumption.

By Condition A4, F_0 satisfies the second-order condition $\mathcal{C}_2(\gamma, \rho, A)$, and let $\overline{F}_0$ be the survival function. Then we have

$$\forall x > 0, \quad \lim_{t\to\infty}\frac{1}{A(1/\overline{F}_0(t))}\left[\frac{\overline{F}_0(tx)}{\overline{F}_0(t)} - x^{-1/\gamma}\right] = x^{-1/\gamma}\frac{x^{\rho/\gamma}-1}{\gamma\rho}. \tag{A.2}$$

Furthermore, by Bellini *et al*. (2014), we have

$$\frac{\overline{F}_0(\xi_{0,\tau})}{1-\tau} \to (\gamma^{-1} - 1) \quad \text{as} \quad \tau \to 1. \tag{A.3}$$

Since $\tau \uparrow 1$, $1-\tau \downarrow 0$, $\xi_\tau \to \infty$, the main source term of variance $Var(W_n(\tau))$ comes from $\tau y_i^*\mathbb{I}(y_i^* \geq 0)$. Assume that $\epsilon_i, i = 1, \ldots, n$ are iid from F_0, by Condition A2, $y_i^* = y_i - \boldsymbol{z}_i^T\boldsymbol{\theta}(\tau) = \{\boldsymbol{K}(z_i)\}^\gamma(\epsilon_i - \xi_{0,\tau})$, thus

$$\begin{aligned} Var(W_n(\tau)) &= 4\tau^2\frac{\sum_{i=1}^n\{\boldsymbol{K}(z_i)\}^{2\gamma}z_i z_i^T}{n}E[(\epsilon - \xi_{0,\tau})^2\mathbb{I}(\epsilon \geq \xi_{0,\tau})] \\ &\quad + o\left((1-\tau)\frac{\sum_{i=1}^n\{\boldsymbol{K}(z_i)\}^{2\gamma}z_i z_i^T}{n}\right) \\ &\to 4\tau^2E[\{\boldsymbol{K}(\boldsymbol{Z})\}^{2\gamma}\boldsymbol{Z}\boldsymbol{Z}^T]\sigma_{g_\tau}^2, \end{aligned} \tag{A.4}$$

where $\sigma_{g_\tau}^2 = E[(\epsilon - \xi_{0,\tau})^2\mathbb{I}(\epsilon \geq \xi_{0,\tau})]$, and we can calculate $\sigma_{g_\tau}^2$ as follows:

$$\begin{aligned} \sigma_{g_\tau}^2 &= \xi_{0,\tau}^2E\left[\left(\frac{\epsilon}{\xi_{0,\tau}} - 1\right)^2\mathbb{I}(\epsilon/\xi_{0,\tau} \geq 1)\right] \\ &= \xi_{0,\tau}^2\int_1^\infty 2(x-1)\overline{F}_0(\xi_{0,\tau}x)\,dx \\ &= \xi_{0,\tau}^2\overline{F}_0(\xi_{0,\tau})\left(\frac{2\gamma^2}{(1-2\gamma)(1-\gamma)} + \int_1^\infty 2(x-1)\left[\frac{\overline{F}_0(\xi_{0,\tau}x)}{\overline{F}_0(\xi_{0,\tau})} - x^{-1/\gamma}\right]dx\right). \end{aligned} \tag{A.5}$$

Plugging (A.2) and (A.3) into this equality, we thus get

$$\begin{aligned} \sigma_{g_\tau}^2 &= \xi_{0,\tau}^2(1-\tau) \\ &\quad\times\left(\frac{2\gamma}{1-2\gamma} + 2A\left(\frac{1}{\overline{F}_0(\xi_{0,\tau})}\right)\left[\frac{1-\gamma+o(1)}{\gamma(1-2\gamma)(1-2\gamma-\rho)} - \frac{1+o(1)}{\gamma(1-\gamma-\rho)}\right] + o(1)\right). \end{aligned} \tag{A.6}$$

Next we calculate μ_{h_τ} directly:

$$\mu_{h_\tau} = 2[\tau E(\mathbb{I}(\epsilon \geq \xi_{0,\tau})) + (1-\tau)E(\mathbb{I}(\epsilon < \xi_{0,\tau}))] = 2[\tau \overline{F}_0(\xi_{0,\tau}) + (1-\tau)F_0(\xi_{0,\tau})],$$

substituting (A.3) into this equality, we have

$$\mu_{h_\tau} = 2\tau(1-\tau)\Big[\gamma^{-1} - (1-\tau)\frac{\gamma^{-1}-2}{\tau}\Big] = 2\tau(1-\tau)[\gamma^{-1} + o(1)]. \tag{A.7}$$

Recall that

$$M_n(\boldsymbol{u},\tau) = \sum_{i=1}^{n}\left[-g_\tau(y_i^*)\frac{\boldsymbol{z}_i^T\boldsymbol{u}}{\sqrt{n}} + \frac{h_\tau(y_i^*)}{2}\left(-\frac{\boldsymbol{z}_i^T\boldsymbol{u}}{\sqrt{n}}\right)^2 + o\left(\frac{1}{n}\right)\right],$$

and

$$\sum_{i=1}^{n}\frac{h_\tau(y_i^*) - \mu_{h_\tau}}{2}\left(-\frac{\boldsymbol{z}_i^T\boldsymbol{u}}{\sqrt{n}}\right)^2 = \sum_{i=1}^{n}\frac{h_\tau(y_i^*) - \mu_{h_\tau}}{2n}(\boldsymbol{z}_i^T\boldsymbol{u})^2 \to 0 \ \ as \ \ n\to\infty,$$

thus,

$$\begin{aligned}\sum_{i=1}^{n}\frac{h_\tau(y_i^*)}{2}\left(-\frac{\boldsymbol{z}_i^T\boldsymbol{u}}{\sqrt{n}}\right)^2 &= \sum_{i=1}^{n}\frac{\mu_{h_\tau}}{2}\left(-\frac{\boldsymbol{z}_i^T\boldsymbol{u}}{\sqrt{n}}\right)^2 + \sum_{i=1}^{n}\frac{h_\tau(y_i^*) - \mu_{h_\tau}}{2}\left(-\frac{\boldsymbol{z}_i^T\boldsymbol{u}}{\sqrt{n}}\right)^2\\ &\to \frac{\mu_{h_\tau}}{2}\boldsymbol{u}^T\boldsymbol{\Sigma}\boldsymbol{u} \quad \text{as} \quad n\to\infty.\end{aligned}$$

Then, we have:

$$M_n(\boldsymbol{u},\tau) \to -\boldsymbol{u}^T W_n(\tau) + \frac{\mu_{h_\tau}}{2}\boldsymbol{u}^T\boldsymbol{\Sigma}\boldsymbol{u} \quad \text{as} \quad n\to\infty.$$

In addition, $E(W_n(\tau)) = \boldsymbol{0}$, it can be readily obtain from Lindeberg–Feller CLT that

$$W_n(\tau) \xrightarrow{d} N\Big(\boldsymbol{0}, 4\tau^2\sigma_{g_\tau}^2 E[\{\boldsymbol{K}(\boldsymbol{Z})\}^{2\gamma}\boldsymbol{Z}\boldsymbol{Z}^T]\Big),$$

then, it follows that

$$\sqrt{n}\{\widehat{\boldsymbol{\theta}}(\tau) - \boldsymbol{\theta}(\tau)\} = (\mu_{h_\tau}\boldsymbol{\Sigma})^{-1}W_n(\tau)\{1 + o_p(1)\} \xrightarrow{d} N(\boldsymbol{0}, \frac{4\tau^2\sigma_{g_\tau}^2}{\mu_{h_\tau}^2}\boldsymbol{H}^{-1}\boldsymbol{\Sigma}\boldsymbol{H}^{-1}), \tag{A.8}$$

uniformly for $\tau \in \mathcal{T}$.

By plugging in (A.6) and (A.7), we have

$$\begin{aligned}&\frac{(1-\tau)4\tau^2\sigma_{g_\tau}^2}{\xi_{0,\tau}^2\mu_{h_\tau}^2}\\ &\quad = \frac{4\tau^2\xi_{0,\tau}^2(1-\tau)^2\left(\frac{2\gamma}{1-2\gamma} + 2A\left(\frac{1}{\overline{F}_0(\xi_{0,\tau})}\right)\left[\frac{1-\gamma+o(1)}{\gamma(1-2\gamma)(1-2\gamma-\rho)} - \frac{1+o(1)}{\gamma(1-\gamma-\rho)}\right] + o(1)\right)}{4\tau^2\xi_{0,\tau}^2(1-\tau)^2[\gamma^{-1}+o(1)]^2}\end{aligned}$$

$$= \frac{2\gamma^3}{1-2\gamma} + 2\gamma^2 A\left(\frac{1}{\overline{F}_0(\xi_{0,\tau})}\right)\left[\frac{1-\gamma+o(1)}{\gamma(1-2\gamma)(1-2\gamma-\rho)} - \frac{1+o(1)}{\gamma(1-\gamma-\rho)}\right]$$
$$+\gamma^2 o(1)$$
$$\to \frac{2\gamma^3}{1-2\gamma} \quad \text{as} \quad \tau \to 1.$$

Therefore,

$$\frac{\sqrt{n(1-\tau)}}{\xi_{0,\tau}}\{\widehat{\boldsymbol{\theta}}(\tau) - \boldsymbol{\theta}(\tau)\} \xrightarrow{d} N\left(0, \frac{2\gamma^3}{1-2\gamma}\boldsymbol{H}^{-1}\boldsymbol{\Sigma}\boldsymbol{H}^{-1}\right),$$

uniformly for $\tau \in \mathcal{T}$, hence we complete the proof. □

Proof of Theorem 3.1. We first denote that $\xi_j = \boldsymbol{z}^T\boldsymbol{\theta}(\tau_j)$ for $j = n-k, \ldots, m$. Note that

$$\widehat{\gamma} = \frac{1}{k-[n^\eta]}\sum_{j=[n^\eta]}^{k} \log \frac{\xi_{n-j}\left(1+\frac{\widehat{\xi}_{n-j}-\xi_{n-j}}{\xi_{n-j}}\right)}{\xi_{n-k}\left(1+\frac{\widehat{\xi}_{n-k}-\xi_{n-k}}{\xi_{n-k}}\right)}$$
$$= \left[\frac{1}{k}\sum_{j=[n^\eta]}^{k}\log\frac{\xi_{n-j}}{\xi_{n-k}} + \frac{1}{k}\sum_{j=[n^\eta]}^{k}\frac{\widehat{\xi}_{n-j}-\xi_{n-j}}{\xi_{n-j}}\{1+o_p(1)\}\right.$$
$$\left. -\frac{1}{k}\sum_{j=[n^\eta]}^{k}\frac{\widehat{\xi}_{n-k}-\xi_{n-k}}{\xi_{n-k}}\{1+o_p(1)\}\right]\frac{k}{k-[n^\eta]}$$
$$=: (E_{1n}+E_{2n}-E_{3n})\{k/(k-[n^\eta])\}.$$

By Lemma 1, $U_Y(t)$ satisfies the second-order condition (2.5) indexed by $(\gamma, \widetilde{\rho}, \widetilde{A})$, where $0<\gamma<1/2, \widetilde{A}(t) = \gamma\widetilde{d}t^{\widetilde{\rho}}$, $\widetilde{\rho} = max(-\gamma, \rho)$, and $\widetilde{d} = d^*\mathbb{I}(\rho \geq -\gamma) - c^{*-1}\boldsymbol{z}^T\boldsymbol{\theta}(r)\mathbb{I}(\rho \leq -\gamma)$, with corresponding constant c^*. Then we consider the three subitems $E_{in}, i = 1, 2, 3$, separately.

Note that

$$\frac{\xi_j}{\xi_{n-k}} \sim \frac{q_j}{q_{n-k}} \sim \frac{U_Y\{(n+1)/(j+i)\}}{U_Y\{(n+1)/(k+i)\}},$$

where q_j is the τ_jth quantile for $j = n-k, \ldots, m$. Thus, we can write E_{1n} as

$$E_{1n} = \frac{1}{k}\sum_{j=1}^{k}\log\frac{\xi_{n-j}}{\xi_{n-k}} - \frac{1}{k}\sum_{j=1}^{[n^\eta-1]}\log\frac{U_Y\{(n+1)/(j+i)\}}{U_Y\{(n+1)/(k+i)\}}$$
$$= \frac{1}{k}\sum_{j=1}^{k}\log\frac{\xi_{n-j}}{\xi_{n-k}} + O(1)\frac{n^\eta}{k}\log\frac{U_Y\left(\frac{n+1}{k+1}\cdot(k+1)\right)}{U_Y\left(\frac{n+1}{k+1}\right)}.$$

Then, by Proposition 1 in Daouia *et al.* (2020a) and Riemann integral approximation, it follows that

$$E_{1n} = -\gamma \int_0^1 \log s\, ds + o(1/k) + \frac{\gamma(\gamma^{-1}-1)^{\gamma} E(Y|z)}{q_{n-k}} \int_o^1 (s^{\gamma}-1)\, ds + o\left(\frac{E(Y|z)}{q_{n-k}}\right)$$
$$+ \frac{(1-\gamma)(\gamma^{-1}-1)^{-\widetilde{\rho}}}{1-\gamma\widetilde{\rho}} \times \widetilde{A}(n/k) \int_0^1 \frac{s^{-\widetilde{\rho}}-1}{\widetilde{\rho}}\, ds + o(\widetilde{A}(n/k)) + O(1)\frac{n^{\eta}\log k}{k}$$
$$= \gamma + \frac{(1-\gamma)(\gamma^{-1}-1)^{-\widetilde{\rho}}}{(1-\widetilde{\rho})(1-\gamma-\widetilde{\rho})}\widetilde{A}(n/k) - \frac{E(Y|z)}{q_{n-k}} \frac{\gamma^2(\gamma^{-1}-1)^{\gamma}}{\gamma+1}$$
$$+ O(k^{-1}n^{\eta}\log k) + o(\widetilde{A}(n/k)). \tag{A.9}$$

We now consider E_{2n}. By Lemma 2, we have

$$\frac{\sqrt{n(1-\tau)}}{\xi_{0,\tau}}\{\widehat{\boldsymbol{\theta}}(\tau) - \boldsymbol{\theta}(\tau)\} = \sqrt{2\gamma^3/(1-2\gamma)}\boldsymbol{H}^{-1} W_n^*(\tau)\{1+o_p(1)\}, \tag{A.10}$$

uniformly for $\tau \in \mathcal{T}$, where $W_n^*(\tau) = (2\tau\sigma_{g_\tau})^{-1} E\{\boldsymbol{K}(\boldsymbol{Z})\}^{-\gamma} W_n(\tau)$ converges to a mean zero and covariance matrix $\boldsymbol{\Sigma}$ Gaussian process on $\mathcal{T}$.
Then

$$E_{2n} = \frac{1}{k}\sum_{j=[n^{\eta}]}^{k} \frac{\xi_{0,\tau_{n-j}}\sqrt{2\gamma^3/(1-2\gamma)}\boldsymbol{z}^T\boldsymbol{\Sigma}^{-1} W_n^*(\tau_{n-j})}{\xi_{\tau_{n-j}}\sqrt{n(1-\tau_{n-j})}}\{1+o_p(1)\},$$

it follows that

$$\frac{\xi_{0,\tau_{n-j}}}{\xi_{\tau_{n-j}}\sqrt{n(1-\tau_{n-j})}} = \left(\frac{j+1}{k+1}\right)^{-1/2} \frac{\{\boldsymbol{K}(\boldsymbol{z})\}^{-\gamma}}{\sqrt{k}}(1+o(1))$$
$$= \{\boldsymbol{K}(\boldsymbol{z})\}^{-\gamma} k^{-1/2}\int_0^1 u^{-1/2}\, du(1+o(1))$$
$$= 2k^{-1/2}\{\boldsymbol{K}(\boldsymbol{z})\}^{-\gamma}(1+o(1)).$$

Thus,

$$E_{2n} = 2k^{-1/2}\sqrt{2\gamma^3/(1-2\gamma)}\boldsymbol{z}^T\boldsymbol{H}^{-1} W_n^*(1)\{\boldsymbol{K}(\boldsymbol{z})\}^{-\gamma}\{1+o_p(1)\}, \tag{A.11}$$

where $W_n^*(1) = lim_{\tau\to 1} W_n^*(\tau)$. Similar to the approximation of E_{2n}, it is easy to see that

$$E_{3n} = k^{-1/2}\sqrt{2\gamma^3/(1-2\gamma)}\boldsymbol{z}^T\boldsymbol{H}^{-1} W_n^*(1)\{\boldsymbol{K}(\boldsymbol{z})\}^{-\gamma}\{1+o_p(1)\}. \tag{A.12}$$

By (A.9), (A.11), and (A.12), and the assumption $\sqrt{k}\widetilde{A}(n/k) \to \lambda_1 \in R$, $\sqrt{k}/q_{\tau_{n-j}} \to \lambda_2 \in R$, and $k^{-1/2}n^{\eta}\log k \to 0$, we obtain that

$$\sqrt{k}(\widehat{\gamma}-\gamma) = \frac{(1-\gamma)(\gamma^{-1}-1)^{-\widetilde{\rho}}}{(1-\widetilde{\rho})(1-\gamma-\widetilde{\rho})}\lambda_1 - E(Y|z)\frac{\gamma^2(\gamma^{-1}-1)^{\gamma}}{\gamma+1}\lambda_2$$
$$+ \sqrt{2\gamma^3/(1-2\gamma)}\boldsymbol{z}^T\boldsymbol{H}^{-1} W_n^*(1)\{\boldsymbol{K}(\boldsymbol{z})\}^{-\gamma}\{1+o_p(1)\}.$$

Thus,

$$\sqrt{k}(\widehat{\gamma}-\gamma)\xrightarrow{d}N(b_z,v_z),$$

and we complete the proof. □

Proof of Theorem 3.2. Note that

$$\log\left(\frac{\widehat{\xi}_{\tau_n'}}{\xi_{\tau_n'}}\right)=(\widehat{\gamma}-\gamma)\log\left(\frac{k}{n(1-\tau_n')}\right)+\log\left(\frac{\widehat{\xi}_{n-k}}{\xi_{n-k}}\right)-\log\left(\left[\frac{n(1-\tau_n')}{k}\right]^{\gamma}\frac{\xi_{\tau_n'}}{\xi_{n-k}}\right).$$

The convergence $\log[k/(n(1-\tau_n'))]\to\infty$ yields

$$\frac{\sqrt{k}}{\log[k/(n(1-\tau_n'))]}\log\left(\frac{\widehat{\xi}_{n-k}}{\xi_{n-k}}\right)=O_p(1/\log[k/(n(1-\tau_n'))])=o_p(1), \tag{A.13}$$

and

$$\begin{aligned}
&\frac{\sqrt{k}}{\log[k/(n(1-\tau_n'))]}\log\left(\left[\frac{n(1-\tau_n')}{k}\right]^{\gamma}\frac{\xi_{\tau_n'}}{\xi_{n-k}}\right)\\
&=\frac{\sqrt{k}}{\log[k/(n(1-\tau_n'))]}\left(\log\left(\frac{\xi_{\tau_n'}}{q_{\tau_n'}}\right)-\log\left(\frac{\xi_{n-k}}{q_{n-k}}\right)+\log\left(\left[\frac{n(1-\tau_n')}{k}\right]^{\gamma}\frac{q_{\tau_n'}}{q_{n-k}}\right)\right)\\
&=O\left(\frac{\sqrt{k}}{\log[k/(n(1-\tau_n'))]}\left[\frac{1}{q_{n-k}}+|\widetilde{A}(n/k)|+\frac{1}{q_{\tau_n'}}\right]\right)\\
&=O\left(\frac{\sqrt{k}}{\log[k/(n(1-\tau_n'))]}\left[\frac{1}{q_{n-k}}+|\widetilde{A}(n/k)|\right]\right)\\
&=o(1).
\end{aligned} \tag{A.14}$$

Here, convergence (A.13) is a consequence of Lemma 2. Convergence (A.14) follows from a combination of Proposition 1 in Daouia *et al.* (2020a), Theorem 2.3.9 in de Haan and Ferreira (2006), and the regular variation of $|\widetilde{A}|$. Combining these convergences and using the delta method leads to the first desired result.

The key idea to prove the second problem is to write

$$\widehat{\xi}_{\widehat{\tau}_n'(p_n)}=\left(\frac{1-\widehat{\tau}_n'(p_n)}{1-\tau_{n-k}}\right)^{-\widehat{\gamma}}\widehat{\xi}_{\tau_{n-k}}=\left(\frac{1-\widehat{\tau}_n'(p_n)}{1-\tau_n'(p_n)}\right)^{-\widehat{\gamma}}\times\left[\left(\frac{1-\tau_n'(p_n)}{1-\tau_{n-k}}\right)^{-\widehat{\gamma}}\widehat{\xi}_{\tau_{n-k}}\right]. \tag{A.15}$$

Moreover, as shown in the part of the proof of Theorem 6 in Daouia *et al.* (2018),

$$\frac{1-\widehat{\tau}_n'(p_n)}{1-\tau_n'(p_n)}=1+O_p\left(\frac{1}{\sqrt{n(1-\tau_{n-k})}}\right).$$

Therefore, we have

$$\left(\frac{1-\widehat{\tau}_n'(p_n)}{1-\tau_n'(p_n)}\right)^{-\widehat{\gamma}} = \exp\left(-\widehat{\gamma}\log\left[\frac{1-\widehat{\tau}_n'(p_n)}{1-\tau_n'(p_n)}\right]\right)$$
$$= \exp\left(-\left[\gamma + O_p\left(\frac{1}{\sqrt{n(1-\tau_{n-k})}}\right)\right] \times O_p\left(\frac{1}{\sqrt{n(1-\tau_{n-k})}}\right)\right)$$
$$= 1 + O_p\left(\frac{1}{\sqrt{n(1-\tau_{n-k})}}\right), \tag{A.16}$$

by a Taylor expansion. Furthermore

$$\widehat{\xi}_{\widehat{\tau}_n'(p_n)} = \left(\frac{1-\widehat{\tau}_n'(p_n)}{1-\tau_{n-k}}\right)^{-\widehat{\gamma}} \widehat{\xi}_{\tau_{n-k}},$$

by definition of the class of estimators $\widehat{\xi}_{\widehat{\tau}_n'(p_n)}$. From Theorem 3.2, we conclude that the conditions of Theorem 3.2 are satisfied if the parameter τ_n' is replaced by $\tau_n'(p_n)$. Then

$$\frac{\sqrt{k}}{\log\left[k/\{n(1-\tau_n'(p_n))\}\right]}\left(\frac{\widehat{\xi}_{\tau_n'(p_n)}}{\xi_{\tau_n'(p_n)}} - 1\right) \xrightarrow{d} N(b_z, v_z).$$

Finally

$$\log\left[\frac{k}{n(1-\tau_n'(p_n))}\right] = \log\left[\frac{1-\tau_{n-k}}{1-\tau_n'(p_n)}\right] = \log\left[\frac{1-\tau_{n-k}}{1-p_n}\right] + \log\left[\frac{1-p_n}{1-\tau_n'(p_n)}\right]$$

and the first term above tends to infinity, while the second term converges to a finite constant. Consequently

$$\log\left[\frac{k}{n(1-\tau_n'(p_n))}\right] = \log\left[\frac{1-\tau_{n-k}}{1-p_n}\right](1+o(1)) = \log\left[\frac{k}{n(1-p_n)}\right](1+o(1)).$$

Together with the equality $\xi_{\tau_n'(p_n)} = q_{p_n}$ which is true by definition of $\tau_n'(p_n)$, this entails

$$\frac{\sqrt{k}}{\log\left[k/(n(1-p_n))\right]}\left(\frac{\widehat{\xi}_{\tau_n'(p_n)}}{q_{p_n}} - 1\right) \xrightarrow{d} N(b_z, v_z), \tag{A.17}$$

Combining (A.15), (A.16), and (A.17), we complete the proof.

For the third part of the proof, we define the simple version $\widehat{\tau}_n' = \widehat{\tau}_n'(p_n), \tau_n' = \tau_n'(p_n)$, then we examining the convergence of $\widehat{XES}_{\widehat{\tau}_n'}$. Write

$$\log\left(\frac{\widehat{XES}_{\widehat{\tau}_n'}}{XES_{\tau_n'}}\right) = \log\left(\frac{\widehat{\xi}_{\widehat{\tau}_n'}}{\xi_{\tau_n'}}\right) + \log\left(\frac{(1-\widehat{\gamma})^{-1}}{(1-\gamma)^{-1}}\right) - \log\left(\frac{XES_{\tau_n'}}{(1-\gamma)^{-1}\xi_{\tau_n'}}\right).$$

By Theorem 3 and the delta method,

$$\frac{\sqrt{k}}{\log\left[k/(n(1-\tau_n'))\right]}\log\left(\frac{\widehat{\xi}_{\widehat{\tau}_n'}}{\xi_{\tau_n'}}\right) \xrightarrow{d} N(b_z, v_z). \tag{A.18}$$

Using Theorem 3.1, the delta method and the convergence $\log[k/(n(1-\tau_n'))] \to \infty$, we get

$$\frac{\sqrt{k}}{\log[k/(n(1-\tau_n'))]} \log\left(\frac{(1-\widehat{\gamma})^{-1}}{(1-\gamma)^{-1}}\right) \xrightarrow{P} 0. \tag{A.19}$$

Using finally a combination of Propositions 1(i) and 4 in Daouia *et al.* (2020a) and the regular variation of $|\widetilde{A}|$, we obtain

$$\frac{\sqrt{k}}{\log[k/(n(1-\tau_n'))]} \log\left(\frac{XES_{\tau_n'}}{(1-\gamma)^{-1}\xi_{\tau_n'}}\right) \to 0. \tag{A.20}$$

Combining convergences (A.17), (A.18), and (A.20), we obtain

$$\frac{\sqrt{k}}{\log[k/(n(1-\tau_n'))]} \log\left(\frac{\widehat{XES}_{\widehat{\tau}_n'}}{XES_{\tau_n'}}\right) \xrightarrow{d} N(b_z, v_z). \tag{A.21}$$

By definition of $\tau_n'(p_n)$, $XES_{\tau_n'(p_n)} = QES_{p_n}$, combining convergences (A.21), we have

$$\frac{\sqrt{k}}{\log[k/\{n(1-\tau_n'(p_n))\}]} \left(\frac{\widehat{XES}_{\widehat{\tau}_n'(p_n)}}{QES_{p_n}} - 1\right) \xrightarrow{d} N(b_z, v_z). \tag{A.22}$$

Recalling that $\log[k/\{n(1-\tau_n'(p_n))\}] = \log[k/(n(1-p_n))](1+o(1))$, we have completed the proof. □