Insurance: Mathematics and Economics 111 (2023) 142–162

Contents lists available at ScienceDirect

Insurance: Mathematics and Economics

journal homepage: www.elsevier.com/locate/ime

Statistical inference for extreme extremile in heavy-tailed heteroscedastic regression model

Yu Chen, Mengyuan Ma, Hongfang Sun *

Department of Statistics and Finance, School of Management, University of Science and Technology of China, Hefei 230026, Anhui, PR China

ARTICLE INFO

Article history:
Received July 2022
Received in revised form March 2023
Accepted 3 April 2023
Available online 12 April 2023

JEL classification:
C1
C13
C15
C51
G22

Keywords:
Conditional extremiles
Extreme value theory
Heavy-tailed distribution
Heteroscedastic regression
Inference

ABSTRACT

As a least squares analogue of quantiles, extremiles define a coherent risk measure determined by weighted expectations instead of tail probabilities. Estimating extremiles of heavy-tailed variables in a regression framework is a challenging task, especially for dependent cases. This paper develops some methods for the estimation of extreme conditional extremiles in the framework of heteroscedastic regression model with heavy-tail noises, specifically, direct and indirect methods based on the conditional extremile estimators for the residuals. We also construct corresponding bias-reduced estimators and investigate their asymptotic properties compared to the original versions. Our mathematical assumptions are satisfied in the mean-variance regression model and heteroscedastic single-index model, which makes it possible to apply our result in a series of important examples. We demonstrate our results through a simulation study and real sets of insurance and financial data analyses.

1. Introduction

Assessing the extreme behaviors of a random phenomenon is a major issue in many fields such as finance, insurance, and environmental science (see Embrechts et al., 1997; Kazama et al., 2009). A typical way of handling extreme events is to estimate extreme quantiles of random variables of interest, for instance, the daily log-return of a stock market index, the claim amounts for an insurance company, or the flood intensity in a given region. Meanwhile, expectiles also can be used to serve such a purpose which is reflected in the recent literature (see Daouia et al., 2018, 2020; Davison et al., 2022; Girard et al., 2021, etc.). Assuming the event of interest is modeled by a quantitative random variable Y. For a fixed τ in $(0, 1)$, the τ-th quantile or expectile of Y is defined as the solution of a minimization problem (see Koenker and Bassett, 1978; Newey and Powell, 1987):

$$\arg\min_{\theta} \mathbb{E}\left[\rho_\tau(Y-\theta) - \rho_\tau(Y)\right], \tag{1}$$

when the loss function takes the form of $\rho_\tau(y) = |\tau - \mathbb{1}_{\{y\le 0\}}||y|$, Eq. (1) corresponds to the τ-th quantile q_τ, and $\rho_\tau(y) = |\tau - \mathbb{1}_{\{y\le 0\}}|y^2$, Eq. (1) represents the τ-th expectile e_τ. The connection between the quantile and expectile has been studied when Y belongs to the domain of attraction of generalized extreme value distribution (see Bellini et al., 2014; Mao et al., 2015; Mao and Yang, 2015, etc.).

Even though these risk measures so far are valuable tools, both quantiles and expectiles have been criticized in the literature for either axiomatic or practical reasons. Quantiles only use the information on whether an observation is below or above some specific value, while expectiles lack transparent interpretation due to the absence of an explicit expression. Apart from these two, expected shortfall (ES) is also a popular risk measure that compensates for the shortcomings of quantile by taking an expectation for the losses above quantile at

* Corresponding author.
E-mail addresses: cyu@ustc.edu.cn (Y. Chen), mmy@mail.ustc.edu.cn (M. Ma), sunhf@mail.ustc.edu.cn (H. Sun).

https://doi.org/10.1016/j.insmatheco.2023.04.001
0167-6687/

some high level. Unfortunately, this approach induces ES to be a conservative risk measure, making it less appealing for individual financial institutions. More discussion of these risk measures can be found in the literature, e.g., Emmer et al. (2015), Mao et al. (2015), and Stupfler (2020). Extremiles, as a new least squares analogue of quantiles introduced by Daouia et al. (2019), deal with these drawbacks. The τ-th extremile is defined by the following minimization problem:

$$\xi_\tau = \text{argmin}_\theta\, \mathbb{E}\left[J_\tau(F(Y))\cdot\left(|Y-\theta|^2-|Y|^2\right)\right],$$

where F is a continuous cumulative distribution function of Y, and the special-generating function $J_\tau(\cdot)=K'_\tau(\cdot)$, with

$$K_\tau(t)=\begin{cases}1-(1-t)^{s(\tau)} & \text{if } 0<\tau\le 1/2\\ t^{r(\tau)} & \text{if } 1/2\le\tau<1\end{cases}$$

being a distribution function with support $[0,1]$, and $r(\tau)=s(1-\tau)=\log(1/2)/\log(\tau)$. As a new least squares analogue of quantiles, extremile were obtained by the alternative representation of quantiles. However, the theory of tail extremile is, in comparison to that of tail quantile and tail expectile, especially the former, relatively unexplored and still in full development. Following earlier work on quantile regression and expectile regression, the extension of extremile estimation to a regression context is a very natural step.

It is usual in practical applications that Y is recorded along with auxiliary information represented by a random covariate $\boldsymbol{X}$. In this setting, the regression τ-th extremile of Y given $\boldsymbol{X}=\boldsymbol{x}$ is defined as

$$\xi_\tau(Y\mid\boldsymbol{x})=\text{argmin}_\theta\,\mathbb{E}\left[J_\tau(F_Y)\cdot\left(|Y-\theta|^2-|Y|^2\right)\mid\boldsymbol{X}=\boldsymbol{x}\right].$$

Estimating extremiles under a regression framework was introduced in Daouia et al. (2022a), but it remains limited to independent and identically distributed heavy-tailed observations. Extending it to a framework of heteroscedastic regression model with heavy-tailed noise is challenging and is the main work of the present paper. We assume that the response variable and the covariate are linked by a heteroscedastic regression model of the form

$$Y=m(\boldsymbol{X})+\sigma(\boldsymbol{X})\varepsilon,$$

where m and $\sigma>0$ are two measurable functions of $\boldsymbol{X}$. This model is flexible since (i) no parametric assumptions are made on $m(\cdot),\sigma(\cdot)$ and ε; (ii) it allows for heterscedasticity via the function $\sigma(\cdot)$. There are some models that fit it, such as the location-scale model (see Spady and Stouli, 2018a,b; Daouia et al., 2022b) and the heteroscedastic single-index model (see Zhu et al., 2013). Moreover, we only consider the case when ε has a heavy-tailed distribution with a constant tail index, which can describe quite well the tail structure and sparseness of the data in most applications in financial and natural sciences. Then the response variable Y inherits its tail behavior from ε, which allows us to use extreme value theory as a useful tool for statistical inferences.

One issue in the above modeling is the assumption on level τ. Our focus is to make inference for extreme extremile so that it is appropriate to model τ as a sequence of the sample size n and $\tau=\tau_n\to 1$ as $n\to\infty$. Then we classifies τ into two categories: the intermediate level where $\tau_n\to 1$ and $n(1-\tau_n)\to\infty$ as $n\to\infty$, and the extreme level where $\tau=\tau'_n\to 1$ and $n(1-\tau'_n)\to c\in[0,\infty)$ as $n\to\infty$. In order to obtain the extremile in the extreme level which suffers from the lack of sample on tail regions, we consider the intermediate level first, since extreme level usually requires Weissman's extrapolation technique (Weissman, 1978) based on the results of intermediate level. The estimation of intermediate extremile is concentrated, on the one hand, by solving the weighted least squares problem called the direct method, and on the other hand, by using the asymptotic relationship between extreme extremiles and quantiles called the indirect method. Besides, the key of Weissman's extrapolation method is a consistent estimator of the tail index of the underlying heavy-tailed distribution. One of the most well-known tail index estimators is the Hill estimator (see Hill, 1975), and some bias reduction versions have been proposed (see Caeiro et al., 2005; Gomes et al., 2015, 2016)) to improve the original Hill estimator. In this article, the resulting extrapolated estimator hinges upon a biased-reduced Hill estimator in Caeiro et al. (2005), which is combined with the biased-reduced extrapolation process to realize bias reduction for extreme extremile estimation.

The main contributions within the above model of this article are three-fold. First, we provide the estimation of extreme conditional extremile in the heteroscedastic regression model with heavy-tailed noise by the methods of both direct and indirect construction. Second, we study a bias-reduced version of the tail index estimator and further provide an automatic bias reduction methodology for extreme extremile estimation. Finally, the theory for the estimation of extreme conditional extremile in a heteroscedastic regression model can be applied in a number of specific models which are shown in the simulation study.

The remaining parts of the paper are organized as follows. The heteroscedastic regression model for heavy-tailed distributions is presented in detail in Section 2. The construction of the resulting estimators is stated in Section 3 under both intermediate level and extreme level, and their asymptotic distributions are provided. Section 4 provides the bias reduction for the estimation which can be viewed as a two-aspect procedure: bias reduction for the tail-index estimator and bias reduction for Wessiman's extrapolation device. Section 5 investigates examples of regression models where the proposed form is fitted and shows how our methods adapt to a simulation. The applications of our methods and theory to real data are provided in Section 6. Finally, a conclusion is proposed in Section 7. All the auxiliary lemmas and proofs, as well as further results, are postponed to the Appendix.

2. Extremile in heavy-tailed heteroscedastic regression model

We consider the class of heteroscedastic regression models, where the relationship between a response variable $Y\in\mathbb{R}$ and a deterministic covariate vector $\boldsymbol{X}\in\mathbb{R}^d$ is in the form of

$$Y=m(\boldsymbol{X})+\sigma(\boldsymbol{X})\varepsilon,\quad \sigma(\boldsymbol{X})>0, \tag{2}$$

where the noise variable $\varepsilon \in \mathbb{R}$ is centered with unit variance and independent of $\boldsymbol{X}$. The random covariate $\boldsymbol{X}$ has a density function $f_{\boldsymbol{X}} \in \mathbb{R}^d$ whose support is a compact set denoted as S. The measurable functions m and σ are the mean and standard deviation functions of Y on the covariate $\boldsymbol{X}$. In fact, for any $\tau \in (0,1)$, the condition extremile of Y given $\boldsymbol{x}$ in this model takes the form of

$$\xi_\tau(Y \mid \boldsymbol{x}) = m(\boldsymbol{x}) + \sigma(\boldsymbol{x})\xi_\tau(\varepsilon), \tag{3}$$

by the location and scale equivariance of extremile and the assumption of the independence between the covariate $\boldsymbol{X}$ and ε.

Moreover, the real random variable ε with distribution F is assumed to be heavy-tailed, that is, its survival function $\overline{F}(\cdot) := 1 - F(\cdot)$ satisfies

$$\overline{F}(t) = t^{-1/\gamma} L(t) \tag{4}$$

for sufficiently large $t > 0$ and some $\gamma > 0$, where L is a slowly varying function at infinity, i.e. $\lim_{t\to\infty} L(tx)/L(t) = 1$, for all $x > 0$. Then, $\overline{F}$ is said to be regularly varying at infinity with index $-1/\gamma$ and γ is called tail index which tunes the tail heaviness of X. That is, the larger the index, the heavier the right tail. As usual in extreme value analysis, it is more convenient to work with the tail quantile function defined as $U(t) = \inf\left\{s \in \mathbb{R} : F(s) \geq 1 - 1/t\right\} = q_{1-1/t}$ for $t > 1$. Then the Eq. (4) is equivalent to the following condition:

$$\lim_{t\to\infty} \frac{U(tx)}{U(t)} = x^\gamma, \text{ for all } x > 0, \gamma > 0. \tag{5}$$

Let $\{(Y_i, \boldsymbol{X}_i)\}_{i=1}^n$ be a (strictly) stationary sequence of copies of the random pair $(Y, \boldsymbol{X})$ which from the model in Eq. (2), then for any $i \in \{1, 2, \ldots, n\}$

$$Y_i = m(\boldsymbol{X}_i) + \sigma(\boldsymbol{X}_i)\varepsilon_i, \ \sigma(\boldsymbol{X}_i) > 0,$$

and $\boldsymbol{X}_i$ is independent of ε_i. Moreover, throughout this paper, suppose that $\varepsilon_i = (Y_i - m(\boldsymbol{X}_i))/\sigma(\boldsymbol{X}_i), i \in \{1, 2, \ldots, n\}$ are independent and share the same distribution F. From Eq. (3), the process of conditional extremile estimation in the heteroscedastic regression model falls into two steps: the model structure estimation of the conditional mean and variance function and the extremile estimation of ε. As we know, the estimation of m and σ can be solved accurately and robustly either under a parametric or a semi-parametric model setting. Moreover, the rate of convergence of model structure estimation is easily achieved faster than unconditional extreme extremile estimation. For this reason, we concentrate on the estimation of the extreme unconditional extremile of ε which combined with the estimation of m and σ can lead to the construction of the extreme conditional extremile estimator of Y given $\boldsymbol{x}$. Throughout Section 3, we suppose that the measurable functions m and σ have been estimated by the consistent estimators $\bar{m}$ and $\bar{\sigma}$ so that the unobserved $\varepsilon_1, \ldots, \varepsilon_n$ are estimated by the residuals $\widehat{\varepsilon}_i = (Y_i - \bar{m}(\boldsymbol{X}_i))/\bar{\sigma}(\boldsymbol{X}_i)$ for all $i = 1, \ldots, n$, and then just focus on the asymptotic properties of the estimator of extreme extremile $\xi_\tau(\varepsilon)$ when $\tau = \tau_n \to 1$.

3. Estimation of extreme extremile in heavy-tailed regression model

In order to obtain the asymptotic properties of the estimators, we make the following assumptions on the distribution function, density function, and tail quantile function of the residual ε.

Condition $\mathcal{C}_1(\gamma)$: Assume the density function $f(\cdot) = F'(\cdot)$ exists and satisfies

$$\lim_{t\to+\infty} \frac{tf(t)}{1 - F(t)} = \frac{1}{\gamma}.$$

Condition $\mathcal{C}_1(\gamma)$ can be viewed as a strengthening condition of Eq. (4) since it implies that $\overline{F}(\cdot)$ is regularly varying with index $-1/\gamma$ when $\gamma > 0$ (see Theorem 1.11 of de Haan and Ferreira (2006) and Proposition 2.1 of Beirlant et al. (2004)). We further make a classical second-order refinement of Eq. (5), that is,

Condition $\mathcal{C}_2(\gamma, \rho, A)$: For all $x > 0$,

$$\lim_{t\to\infty} \frac{1}{A(t)} \left[\frac{U(tx)}{U(t)} - x^\gamma\right] = x^\gamma \frac{x^\rho - 1}{\rho},$$

where $\gamma > 0, \rho \leq 0$ and $A(\cdot)$ is a positive or negative function satisfying $\lim_{t\to\infty} A(t) = 0$. Here, $\left(x^\rho - 1\right)/\rho$ is to be understood as $\log x$ when $\rho = 0$.

In the context of extreme value analysis, the conditions $\mathcal{C}_1(\gamma)$ and $\mathcal{C}_2(\gamma, \rho, A)$ required for the heavy-tailed distribution are used, for example, in Daouia et al. (2019) and Girard et al. (2022a). In addition, we need to make the following extra condition mentioned by Daouia et al. (2019):

Condition $\mathcal{C}_3$: The distribution F is twice differentiable on its support which is an interval, and its corresponding density function f satisfies

$$\sup_{0<t<1} t(1-t)\frac{f'(q_t)}{[f(q_t)]^2} < \infty.$$

By Proposition 2.4.9 in de Haan and Ferreira (2006), under the condition $\mathcal{C}_3$, we can relate the empirical quantile process $t \mapsto \varepsilon_{[nt],n}$, for $0 < t < 1$, to a sequence of standard Brownian bridges.

According to the definition of unconditional extremile, it is required the first moment of H_τ is finite with the distribution function $F_{H_\tau} = K_\tau(F)$. Since $\mathbb{E}|\varepsilon| < \infty$ implies $\mathbb{E}|H_\tau| < \infty$ for any $0 \leq \tau \leq 1$, one of our minimal working assumptions throughout is that $0 < \gamma <$

1 and $\mathbb{E}|\varepsilon_-| < \infty$, where $\varepsilon_- = \max(-\varepsilon, 0)$. The conditions above are standard and general in the extreme value literature. Before we state the main results, we introduce a quantity defined as $\max_{1\le i\le n} \frac{|\widehat{\varepsilon}_i - \varepsilon_i|}{1+|\varepsilon_i|}$ which measures the difference between the available residual $\widehat{\varepsilon}_i$ and unseen error ε_i for all $1 \le i \le n$. The same quantity is used in Ahmad et al. (2020) and Girard et al. (2021) for the extreme conditional expectiles and quantiles under heavy-tailed distributions in the heteroscedastic regression model respectively.

Before making inference for extremile $\xi_\tau(\varepsilon)$ at a very extreme level, we first work with the extremile at the intermediate level. Section 3.1 concentrates on an M-estimator by solving the least squares problem directly. Section 3.2 provides another estimator based on the asymptotic relationship between extremiles and quantiles. Then, a tail extrapolation motivates the estimation for the extreme extremiles for ε in Section 3.3. We build a plug-in estimator of extreme conditional extremile for Y given $\boldsymbol{X}$ in Section 3.4.

3.1. Direct estimation of intermediate extremile

For the intermediate level, i.e. $\tau_n \to 1$ and $n(1-\tau_n) \to \infty$, we solve the weighted least squares problem on the accessible residuals $\widehat{\varepsilon}_i, i = 1, \ldots, n$, to obtain the direct estimator:

$$\widehat{\xi}^M_{\tau_n}(\varepsilon) = \arg\min_{t\in\mathbb{R}} \sum_{i=1}^n J_{\tau_n}\left(\frac{i}{n}\right)\left|\widehat{\varepsilon}_{i,n} - t\right|^2,$$

where $\widehat{\varepsilon}_{i,n}$ denotes the i-th order statistic of the residuals. Additionally, it is expressed explicitly as

$$\widehat{\xi}^M_{\tau_n}(\varepsilon) = \frac{\sum_{i=1}^n J_{\tau_n}\left(\frac{i}{n}\right)\widehat{\varepsilon}_{i,n}}{\sum_{i=1}^n J_{\tau_n}\left(\frac{i}{n}\right)}. \tag{6}$$

In fact, the estimator in Eq. (6) can be obtained by replacing $\varepsilon_{i,n}$ with $\widehat{\varepsilon}_{i,n}$ in the extremile estimator of the noised variable ε: $\widetilde{\xi}^M_{\tau_n}(\varepsilon) = \sum_{i=1}^n J_{\tau_n}\left(\frac{i}{n}\right)\varepsilon_{i,n} \Big/ \sum_{i=1}^n J_{\tau_n}\left(\frac{i}{n}\right)$. As the asymptotic properties of $\widetilde{\xi}^M_{\tau_n}(\varepsilon)$ have been tackled by Theorem 4 in Daouia et al. (2019), we are interested in demonstrating that $\widehat{\xi}^M_{\tau_n}(\varepsilon)$ and $\widetilde{\xi}^M_{\tau_n}(\varepsilon)$ have the same asymptotic behavior. We are now ready for our first theorem giving the consistency and asymptotic normality of the intermediate estimator $\widehat{\xi}^M_{\tau_n}(\varepsilon)$.

Theorem 3.1. *Suppose that $\mathbb{E}|\varepsilon_-| < \infty$, $\mathcal{C}_1(\gamma)$, $\mathcal{C}_2(\gamma, \rho, A)$ and $\mathcal{C}_3$ hold with $0 < \gamma < 1/2$. Let $\tau_n \uparrow 1$ be such that $n(1-\tau_n) \to \infty$ and $\sqrt{n(1-\tau_n)}A\left((1-\tau_n)^{-1}\right) = O(1)$. If the residuals $\widehat{\varepsilon}_i$, $1 \le i \le n$, satisfy*

$$\sqrt{n(1-\tau_n)} \max_{1\le i\le n} \frac{|\widehat{\varepsilon}_i - \varepsilon_i|}{1+|\varepsilon_i|} \xrightarrow{\mathbb{P}} 0, \tag{7}$$

then

$$\sqrt{n(1-\tau_n)}\left(\frac{\widehat{\xi}^M_{\tau_n}(\varepsilon)}{\xi_{\tau_n}(\varepsilon)} - 1\right) \xrightarrow{d} \frac{\gamma\sqrt{\log 2}}{\Gamma(1-\gamma)} \int_0^\infty t^{-\gamma-1}e^{-t}W(t)dt,$$

where $\Gamma(\cdot)$ is the Gamma function and W is a standard Brownian motion. In particular,

$$\sqrt{n(1-\tau_n)}\left(\frac{\widehat{\xi}^M_{\tau_n}(\varepsilon)}{\xi_{\tau_n}(\varepsilon)} - 1\right) \xrightarrow{d} \mathcal{N}(0, V(\gamma))$$

with

$$V(\gamma) = \left(\frac{\gamma\sqrt{\log 2}}{\Gamma(1-\gamma)}\right)^2 \int_0^\infty \int_0^\infty (tu)^{-\gamma-1}e^{-t-u}(t\wedge u)dtdu.$$

Remark 1. Theorem 3.1 requires the tail index γ satisfies $0 < \gamma < 1/2$ to guarantee that $\int_0^\infty t^{-\gamma-1}e^{-t}W(t)\,dt$ is well-defined, and the same condition appears in Theorem 3.4. The condition Eq. (7) can narrow the gap between $\widehat{\varepsilon}_{i,n}$ and $\varepsilon_{i,n}$ such that $\widehat{\xi}^M_{\tau_n}(\varepsilon)/\widetilde{\xi}^M_{\tau_n}(\varepsilon) - 1$ is asymptotically negligible. Obviously, the condition (7) is related to the estimation of the conditional mean function and variance function at $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n$. From Theorem 3.1, it can be seen that the direct estimator $\widehat{\xi}^M_{\tau_n}(\varepsilon)$ has the $\sqrt{n(1-\tau_n)}$-asymptotic normality, similar to the sample M-statistic $\widetilde{\xi}^M_{\tau_n}(\varepsilon)$ under the weak condition (7) exclusively. More generally, Theorem 3.1 is an analogue based on the available residuals of Theorem 4 in Daouia et al. (2019).

3.2. Indirect estimation of intermediate extremile

Exploiting the asymptotic relationship between extremiles and quantiles, another indirect estimator can be considered. It has been found in Proposition 3(i) of Daouia et al. (2019) that under condition (4) or equivalently condition (5) with $0 < \gamma < 1$, we have

$$\frac{\xi_\tau}{q_\tau} \sim g(\gamma) \text{ as } \tau \uparrow 1, \tag{8}$$

where $g(t)=\Gamma(1-t)\{\log 2\}^t$ and the notation "$\sim$" means asymptotic equivalence. By resorting to Eq. (8), we construct an indirect estimator based on quantile at the intermediate level,

$$\widehat{\xi}^{Q}_{\tau_n}(\varepsilon)=g(\bar{\gamma})\cdot\widehat{q}_{\tau_n}(\varepsilon)=g(\bar{\gamma})\cdot\widehat{\varepsilon}_{n-\lfloor n(1-\tau_n)\rfloor,n}.$$

Here, $\bar{\gamma}$ is a consistent estimator of γ. In the existing literature, amounts of tail index estimators have been proposed and studied. The classic Hill-type estimator and its bias-reduced version will be introduced in Section 4. Denote the Digamma function by $\mathcal{D}(t)=\frac{\Gamma'(t)}{\Gamma(t)}$, where $\Gamma'(\cdot)$ represents the first derivative of $\Gamma(\cdot)$. The following theorem yields the asymptotic properties of the indirect estimator $\widehat{\xi}^{Q}_{\tau_n}(\varepsilon)$ in the intermediate case.

Theorem 3.2. *Suppose that $\mathbb{E}|\varepsilon_-|<\infty$ and $\mathcal{C}_2(\gamma,\rho,A)$ holds with $0<\gamma<1$. Let $\tau_n\uparrow 1$ be such that*

$$n(1-\tau_n)\to\infty,\ \sqrt{n(1-\tau_n)}A\left((1-\tau_n)^{-1}\right)\to\lambda_1\in\mathbb{R},\ (1-\tau_n)\sqrt{n(1-\tau_n)}\to\lambda_2\in\mathbb{R}. \tag{9}$$

Assume also that the estimator $\bar{\gamma}$ satisfies $\sqrt{n(1-\tau_n)}(\bar{\gamma}-\gamma)\xrightarrow{d}\Psi$, where Ψ is nondegenerate, and is asymptotically independent from $\widehat{q}_{\tau_n}(\varepsilon)$. If the residuals $\widehat{\varepsilon}_i$, $i=1,\dots,n$, satisfy condition (7), then

$$\sqrt{n(1-\tau_n)}\left(\frac{\widehat{\xi}^{Q}_{\tau_n}(\varepsilon)}{\xi_{\tau_n}(\varepsilon)}-1\right)\xrightarrow{d}\eta(\gamma)\cdot\Psi+\Phi-\left[\lambda_1\frac{b_1(\gamma,\rho)}{g(\gamma)}+\lambda_2\frac{b_2(\gamma)}{g(\gamma)}\right],$$

where $\eta(\gamma)=-\mathcal{D}(1-\gamma)+\log(\log 2)$, $\Phi\sim\mathcal{N}\left(0,\gamma^2\right)$ is independent of Ψ,

$$b_1(\gamma,\rho)=\begin{cases}\frac{g(\gamma+\rho)-g(\gamma)}{\rho}, & \rho<0\\ (\log 2)^{\gamma}\int_0^{\infty}t^{-\gamma}e^{-t}(\log(\log 2)-\log(t))dt, & \rho=0\end{cases}$$

and

$$b_2(\gamma)=-\frac{1}{2}g(\gamma)+\left(1+\frac{\log 2}{2}\right)g(\gamma-1)-\frac{\log 2}{2}g(\gamma-2).$$

Remark 2. Theorem 3.2 shows the $\sqrt{n(1-\tau_n)}$-asymptotic consistency of the indirect estimator $\widehat{\xi}^{Q}_{\tau_n}(\varepsilon)$. It can be noted that it holds under weaker regularity conditions than Theorem 3.1. The condition (9) guarantees that the bias arising from the asymptotic equivalent (8) is controlled, and the bias condition $\sqrt{n(1-\tau_n)}A\left((1-\tau_n)^{-1}\right)\to\lambda_1\in\mathbb{R}$ is similar to the condition $\sqrt{n(1-\tau_n)}A\left((1-\tau_n)^{-1}\right)=O(1)$ in Theorem 3.1. The condition is a little stricter, but still standard in extreme value theory. One can refer to P.77 in de Haan and Ferreira (2006) for more details. The independence between Ψ and Φ actually inherits that between $\bar{\gamma}$ and $\widehat{q}_{\tau_n}(\varepsilon)$.

3.3. Estimation of extreme extremile based on extrapolation

We now deal with the extremile of ε at a very extreme level τ_n' such that $n\left(1-\tau_n'\right)\to c<\infty$. Motivated by the first-order regular variation condition (5) and the asymptotic relationship (8), it can be made that

$$\frac{\xi_{\tau_n'}}{\xi_{\tau_n}}\approx\frac{q_{\tau_n'}}{q_{\tau_n}}\approx\left(\frac{1-\tau_n'}{1-\tau_n}\right)^{-\gamma}\quad\text{as } n\to\infty. \tag{10}$$

This leads to an extremile estimator built on the extrapolation procedure

$$\bar{\xi}^{\star}_{\tau_n'}(\varepsilon)=\left(\frac{1-\tau_n'}{1-\tau_n}\right)^{-\bar{\gamma}}\bar{\xi}_{\tau_n}(\varepsilon),$$

which is called the Weissman-type estimator (see Weissman, 1978). Here, $\bar{\gamma}$ and $\bar{\xi}_{\tau_n}(\varepsilon)$ are consistent estimators of γ and $\xi_{\tau_n}(\varepsilon)$. We explore the asymptotic behavior of the extrapolated estimator.

Theorem 3.3. *Suppose that $\mathbb{E}|\varepsilon_-|<\infty$ and $\mathcal{C}_2(\gamma,\rho,A)$ holds with $0<\gamma<1$ and $\rho<0$. Let the sequences $\tau_n\uparrow 1$ and $\tau_n'\uparrow 1$ be such that*

$$n(1-\tau_n)\to\infty,\ n\left(1-\tau_n'\right)\to c<\infty,\ \sqrt{n(1-\tau_n)}/\log\left[(1-\tau_n)/(1-\tau_n')\right]\to\infty, \tag{11}$$

$$\sqrt{n(1-\tau_n)}\max\left(A\left((1-\tau_n)^{-1}\right),1-\tau_n\right)=O(1). \tag{12}$$

Assume also that $\sqrt{n(1-\tau_n)}(\bar{\gamma}-\gamma)\xrightarrow{d}\Psi$, where Ψ is nondegenerate. If $\bar{\xi}_{\tau_n}(\varepsilon)$ is a consistent estimator of $\xi_{\tau_n}(\varepsilon)$ satisfying $\sqrt{n(1-\tau_n)}\left(\frac{\bar{\xi}_{\tau_n}(\varepsilon)}{\xi_{\tau_n}(\varepsilon)}-1\right)=O_{\mathbb{P}}(1)$, then

$$\frac{\sqrt{n(1-\tau_n)}}{\log\left[(1-\tau_n)/\left(1-\tau_n'\right)\right]}\left(\frac{\bar{\xi}^{\star}_{\tau_n'}(\varepsilon)}{\xi_{\tau_n'}(\varepsilon)}-1\right)\xrightarrow{d}\Psi\quad\textit{as } n\to\infty.$$

Remark 3. Theorem 3.3 states that the limiting distribution of $\bar{\xi}_{\tau_n}(\varepsilon)$ is dominated by $\sqrt{n(1-\tau_n)}$-asymptotic distribution of the tail index estimator $\bar{\gamma}$, which is a common conclusion about Weissman-type estimator. For the consistent estimator $\bar{\xi}_{\tau_n}(\varepsilon)$ in Theorem 3.3, we provide $\widehat{\xi}^{M}_{\tau_n}(\varepsilon)$ and $\widehat{\xi}^{Q}_{\tau_n}(\varepsilon)$ to estimate the intermediate extremile $\xi_{\tau_n}(\varepsilon)$. In particular, $\bar{\xi}_{\tau_n}(\varepsilon)$ may also be expressed as a convex combination of them, that is $\kappa\widehat{\xi}^{M}_{\tau_n}(\varepsilon)+(1-\kappa)\widehat{\xi}^{Q}_{\tau_n}(\varepsilon)$, where κ is a constant satisfying $0\le\kappa\le 1$. The condition (12) ensures that the bias terms due to the relationship (8) and the Eq. (10) can be neglected asymptotically.

3.4. Estimation of extreme conditional extremile

In this section, we aim to estimate the conditional extremiles of Y given $\boldsymbol{x}$ in extremely high levels τ. Recalling Eq. (3), we estimate the intermediate and extreme conditional extremile by

$$\begin{aligned} &\bar{\xi}_{\tau_n}(Y \mid \boldsymbol{x}) = \bar{m}(\boldsymbol{x}) + \bar{\sigma}(\boldsymbol{x})\bar{\xi}_{\tau_n}(\varepsilon) \\ \text{and} \quad &\bar{\xi}^{\star}_{\tau_n'}(Y \mid \boldsymbol{x}) = \bar{m}(\boldsymbol{x}) + \bar{\sigma}(\boldsymbol{x})\bar{\xi}^{\star}_{\tau_n'}(\varepsilon) = \bar{m}(\boldsymbol{x}) + \bar{\sigma}(\boldsymbol{x})\left(\frac{1-\tau_n'}{1-\tau_n}\right)^{-\bar{\gamma}}\bar{\xi}_{\tau_n}(\varepsilon). \end{aligned} \tag{13}$$

Specifically, we exploit the direct method to estimate $\xi_{\tau_n}(Y \mid \boldsymbol{x})$ and $\xi_{\tau_n'}(Y \mid \boldsymbol{x})$, which is denoted by $\widehat{\xi}^M_{\tau_n}(Y \mid \boldsymbol{x})$ and $\widehat{\xi}^{M,\star}_{\tau_n'}(Y \mid \boldsymbol{x})$, where $\bar{\xi}_{\tau_n}(\varepsilon)$ in Eq. (13) is essentially replaced with $\widehat{\xi}^M_{\tau_n}(\varepsilon)$. Facilitated by the results above, we study the asymptotic properties of these two estimators.

Theorem 3.4. *Suppose that $\mathbb{E}|\varepsilon_-| < \infty$, $\mathcal{C}_1(\gamma)$, $\mathcal{C}_2(\gamma,\rho,A)$ and $\mathcal{C}_3$ hold with $0<\gamma<1/2$ and $\rho<0$. Assume also that the sequences $\tau_n \uparrow 1$ and $\tau_n' \uparrow 1$ satisfy conditions (11) and (12). Let the estimator $\bar{\gamma}$ be such that $\sqrt{n(1-\tau_n)}(\bar{\gamma}-\gamma) \xrightarrow{d} \Psi$, where Ψ is nondegenerate. If the consistent estimators $\bar{m}$ and $\bar{\sigma}$ satisfy*

$$\sqrt{n(1-\tau_n)}\max_{1\le i\le n}\left|\frac{\bar{m}(\boldsymbol{X}_i)-m(\boldsymbol{X}_i)}{\sigma(\boldsymbol{X}_i)}\right| \xrightarrow{\mathbb{P}} 0, \quad \sqrt{n(1-\tau_n)}\max_{1\le i\le n}\left|\frac{\bar{\sigma}(\boldsymbol{X}_i)}{\sigma(\boldsymbol{X}_i)}-1\right| \xrightarrow{\mathbb{P}} 0, \tag{14}$$

$$\sqrt{n(1-\tau_n)}\left(\bar{m}(\boldsymbol{x})-m(\boldsymbol{x})\right) = \mathrm{O}_{\mathbb{P}}(1), \quad \sqrt{n(1-\tau_n)}\left(\frac{\bar{\sigma}(\boldsymbol{x})}{\sigma(\boldsymbol{x})}-1\right) = \mathrm{o}_{\mathbb{P}}(1), \tag{15}$$

then as $n\to\infty$, we have

$$\sqrt{n(1-\tau_n)}\left(\frac{\widehat{\xi}^M_{\tau_n}(Y \mid \boldsymbol{x})}{\xi_{\tau_n}(Y \mid \boldsymbol{x})}-1\right) \xrightarrow{d} \mathcal{N}(0, V(\gamma))$$

and

$$\frac{\sqrt{n(1-\tau_n)}}{\log\left[(1-\tau_n)/(1-\tau_n')\right]}\left(\frac{\widehat{\xi}^{M,\star}_{\tau_n'}(Y \mid \boldsymbol{x})}{\xi_{\tau_n'}(Y \mid \boldsymbol{x})}-1\right) \xrightarrow{d} \Psi.$$

Remark 4. We observe that $\widehat{\xi}^M_{\tau_n}(Y \mid \boldsymbol{x})$ and $\widehat{\xi}^{M,\star}_{\tau_n'}(Y \mid \boldsymbol{x})$ have the same asymptotic distribution as $\widehat{\xi}^M_{\tau_n}(\varepsilon)$ and $\widehat{\xi}^{M,\star}_{\tau_n'}(\varepsilon)$, respectively. Note that condition (7) is replaced by condition (14) on the residuals $\widehat{\varepsilon}_i$. Compared with the former, condition (14) works exclusively on $\bar{m}$ and $\bar{\sigma}$ which is more available for the estimation of model structure, although condition (7) is weaker. If we only require the consistency of $\widehat{\xi}^{M,\star}_{\tau_n'}(Y \mid \boldsymbol{x})$, condition (15) can be reduced to

$$\frac{\sqrt{n(1-\tau_n)}}{\log\left(\frac{1-\tau_n}{1-\tau_n'}\right)}\left(\bar{m}(\boldsymbol{x})-m(\boldsymbol{x})\right) = \mathrm{O}_{\mathbb{P}}(1), \quad \frac{\sqrt{n(1-\tau_n)}}{\log\left(\frac{1-\tau_n}{1-\tau_n'}\right)}\left(\frac{\bar{\sigma}(\boldsymbol{x})}{\sigma(\boldsymbol{x})}-1\right) = \mathrm{o}_{\mathbb{P}}(1), \tag{16}$$

or

$$\bar{m}(\boldsymbol{x})-m(\boldsymbol{x}) = \mathrm{O}_{\mathbb{P}}(1), \quad \frac{\sqrt{n(1-\tau_n)}}{\log\left(\frac{1-\tau_n}{1-\tau_n'}\right)q_{\tau_n}(\varepsilon)} = \mathrm{O}(1), \quad \frac{\sqrt{n(1-\tau_n)}}{\log\left(\frac{1-\tau_n}{1-\tau_n'}\right)}\left(\frac{\bar{\sigma}(\boldsymbol{x})}{\sigma(\boldsymbol{x})}-1\right) = \mathrm{o}_{\mathbb{P}}(1). \tag{17}$$

The condition (17) relaxes the assumption for the conditional mean function $m(\cdot)$ by strengthening the control of the intermediate quantile $q_{\tau_n}(\varepsilon)$ in contrast with condition (16). In practice, the consistent estimator $\bar{m}(\cdot)$ converges as fast or faster than $\bar{\sigma}(\cdot)$. This means that the condition (16) is more likely to be fulfilled. Moreover, the rate of convergence of $\bar{m}$ and $\bar{\sigma}$ at the observations $\{\boldsymbol{X}_i\}_{i=1}^n$ is assumed to be faster than at the ordinary point $\boldsymbol{x}$ clearly.

Similarly, the next result concentrates on the indirect estimators

$$\begin{aligned} \widehat{\xi}^Q_{\tau_n}(Y \mid \boldsymbol{x}) &= \bar{m}(\boldsymbol{x}) + \bar{\sigma}(\boldsymbol{x})\cdot g(\bar{\gamma})\widehat{\varepsilon}^{(n)}_{n-\lfloor n(1-\tau_n)\rfloor,n}, \\ \widehat{\xi}^{Q,\star}_{\tau_n'}(Y \mid \boldsymbol{x}) &= \bar{m}(\boldsymbol{x}) + \bar{\sigma}(\boldsymbol{x})\cdot\left(\frac{1-\tau_n'}{1-\tau_n}\right)^{-\bar{\gamma}} g(\bar{\gamma})\widehat{\varepsilon}^{(n)}_{n-\lfloor n(1-\tau_n)\rfloor,n}. \end{aligned}$$

Corollary 3.1. *Suppose that $\mathbb{E}|\varepsilon_-| < \infty$, and $\mathcal{C}_2(\gamma,\rho,A)$ holds with $0<\gamma<1$ and $\rho<0$. Assume also that the sequences $\tau_n \uparrow 1$ and $\tau_n' \uparrow 1$ satisfies conditions (9) and (11). Let the estimator $\bar{\gamma}$ be such that $\sqrt{n(1-\tau_n)}(\bar{\gamma}-\gamma) \xrightarrow{d} \Psi$, where Ψ is nondegenerate, and is asymptotically independent from $\widehat{q}_{\tau_n}(\varepsilon)$. If the consistent estimators $\bar{m}$ and $\bar{\sigma}$ satisfy conditions (14) and (15), then we have*

$$\sqrt{n(1-\tau_n)}\left(\frac{\widehat{\xi}^Q_{\tau_n}(Y \mid \boldsymbol{x})}{\xi_{\tau_n}(Y \mid \boldsymbol{x})}-1\right) \xrightarrow{d} \eta(\gamma)\cdot\Psi + \Phi - \left[\lambda_1\frac{b_1(\gamma,\rho)}{g(\gamma)} + \lambda_2\frac{b_2(\gamma)}{g(\gamma)}\right]$$

$$\text{and} \quad \frac{\sqrt{n(1-\tau_n)}}{\log\left[(1-\tau_n)/(1-\tau_n')\right]}\left(\frac{\widehat{\xi}^{Q,\star}_{\tau_n'}(Y \mid \boldsymbol{x})}{\xi_{\tau_n'}(Y \mid \boldsymbol{x})}-1\right) \xrightarrow{d} \Psi, \quad \text{as } n\to\infty.$$

Under appropriate assumptions, the indirect estimator $\widehat{\xi}^{Q,\star}_{\tau_n'}(Y \mid \boldsymbol{x})$ shares consistency with the direct estimator $\widehat{\xi}^{M,\star}_{\tau_n'}(Y \mid \boldsymbol{x})$. A simulated performance assessment of these two extrapolated estimators is conducted in Section 5.

4. Bias reduction for extreme extremile estimation

Due to the relationship between the auxiliary function $A(\cdot)$ and the bias term, there are inevitable difficulties in analyzing the bias in extreme extremile estimation without extra assumptions on $A(\cdot)$. Consequently, we suppose $A(t) = \gamma b t^{\rho}$ with a nonzero constant b in the whole section. Furthermore, it is convenient to denote a sequence $k(n) = \lfloor n(1-\tau_n) \rfloor$, which is the size of the highest values in the residuals. We start by disposing of the estimation of the tail index in Section 4.1. The bias-reduced versions of the conditional extrapolated estimators $\widehat{\xi}^{M,\star}_{\tau_n'}(Y \mid \boldsymbol{x})$ and $\widehat{\xi}^{Q,\star}_{\tau_n'}(Y \mid \boldsymbol{x})$ are introduced and studied in Section 4.2.

4.1. Bias reduction for tail index estimator

The conditional estimators $\widehat{\xi}^{M,\star}_{\tau_n'}(Y \mid \boldsymbol{x})$ and $\widehat{\xi}^{Q,\star}_{\tau_n'}(Y \mid \boldsymbol{x})$ inherit the asymptotic distribution of the tail index γ. From this, we focus on reducing the bias caused by tail index estimators naturally. Here, consider the classical Hill estimator (see Hill, 1975),

$$\widehat{\gamma}_k^{\mathrm{H}} = \frac{1}{k}\sum_{i=1}^{k} \log \frac{\widehat{\varepsilon}_{n-i+1,n}}{\widehat{\varepsilon}_{n-k,n}},$$

where $\lfloor \cdot \rfloor$ is the floor function. Corollary 2.1 in Girard et al. (2021) has shown that under the assumptions $\mathcal{C}_2(\gamma,\rho,A)$ with $\sqrt{n(1-\tau_n)}A\left((1-\tau_n)^{-1}\right) \to \lambda \in \mathbb{R}$ and (7), the estimator $\widehat{\gamma}_k^{\mathrm{H}}$ constructed by the residuals is $\sqrt{n(1-\tau_n)}$-asymptotically normal with expectation $\lambda/(1-\rho)$ and variance γ^2 and is asymptotically independent from Φ. Inspired by bias-reduced strategy of Caeiro et al. (2005), it yields the bias-reduced Hill estimator

$$\widehat{\gamma}_k^{\mathrm{H,RB}} = \widehat{\gamma}_k^{\mathrm{H}}\left(1 - \frac{\bar{b}}{1-\bar{\rho}}\left(\frac{n}{k}\right)^{\bar{\rho}}\right),$$

where $\bar{b}$ and $\bar{\rho}$ are consistent estimators of b and ρ respectively. It is obtained that $\sqrt{k}\left(\widehat{\gamma}_k^{\mathrm{H,RB}} - \gamma\right)$ is asymptotically Gaussian with variance equal to γ^2 and zero mean from Corollary B.1 in Appendix B. The tail index estimators mentioned above both require the optimal choice of the sequence $k = k(n)$. Here, we select the intermediate level k by minimizing the asymptotic mean-squared error (AMSE) of the Hill estimator (see P.77 of de Haan and Ferreira, 2006):

$$\widehat{k} = \left\lfloor \left(\frac{(1-\bar{\rho})^2}{-2\bar{\rho}\bar{b}^2}\right)^{1/(1-2\bar{\rho})} n^{-2\bar{\rho}/(1-2\bar{\rho})} \right\rfloor.$$

4.2. Bias reduction for extrapolation

In fact, even though we adopt the bias-reduced estimator $\widehat{\gamma}_k^{\mathrm{H,RB}}$, the conditional extreme estimators $\widehat{\xi}^{M,\star}_{\tau_n'}(Y \mid \boldsymbol{x})$ and $\widehat{\xi}^{Q,\star}_{\tau_n'}(Y \mid \boldsymbol{x})$ are still biased due to heavy-tailed extrapolation as shown in Eq. (10). Analyze the extrapolation process for the direct method in detail,

$$\left(\frac{1-\tau_n'}{1-\tau_n}\right)^{\gamma}\frac{\xi_{\tau_n'}(\varepsilon)}{\xi_{\tau_n}(\varepsilon)} = \underbrace{\frac{g(\gamma)q_{\tau_n}(\varepsilon)}{\xi_{\tau_n}(\varepsilon)}}_{1+\mathcal{R}_{n,1}} \times \underbrace{\frac{\xi_{\tau_n'}(\varepsilon)}{g(\gamma)q_{\tau_n'}(\varepsilon)}}_{1+\mathcal{R}_{n,2}} \times \underbrace{\left(\frac{1-\tau_n'}{1-\tau_n}\right)^{\gamma}\frac{q_{\tau_n'}(\varepsilon)}{q_{\tau_n}(\varepsilon)}}_{1+\mathcal{R}_{n,3}}. \tag{18}$$

It's motivated by the bias-reduced strategy for extreme expectiles introduced by Girard et al. (2022b). Both $\mathcal{R}_{n,1}$ and $\mathcal{R}_{n,2}$ are connected with the precise expansion of Eq. (8) at the intermediate level and extreme level, which is expressed as

$$\mathcal{R}_{n,1} = \left(1 + \frac{C_1(\gamma,\rho)A\left((1-\tau_n)^{-1}\right)}{g(\gamma)} + \frac{C_2(\gamma)(1-\tau_n)}{g(\gamma)} + \mathrm{o}(1)\right)^{-1} - 1,$$
$$\mathcal{R}_{n,2} = \frac{C_1(\gamma,\rho)A\left((1-\tau_n')^{-1}\right)}{g(\gamma)} + \frac{C_2(\gamma)(1-\tau_n')}{g(\gamma)} + \mathrm{o}(1).$$

According to Theorem 2.3.9 in de Haan and Ferreira (2006), the bias term $\mathcal{R}_{n,3}$ is defined as

$$\mathcal{R}_{n,3} = \frac{1}{\rho}\left[\left(\frac{1-\tau_n'}{1-\tau_n}\right)^{-\rho} - 1\right] A\left(\frac{1}{1-\tau_n}\right)(1+o(1)).$$

Under the assumption $A(t) = \gamma b t^{\rho}$, by plugging in the consistent estimators $\bar{\gamma}$, $\bar{b}$ and $\bar{\rho}$ to estimate $\mathcal{R}_{n,1}$, $\mathcal{R}_{n,2}$ and $\mathcal{R}_{n,3}$, this results in

$$\bar{\mathcal{R}}_{n,1} = \left(1 + \frac{C_1(\bar{\gamma}, \bar{\rho}) \cdot \bar{\gamma} \bar{b} (1-\tau_n)^{-\bar{\rho}}}{g(\bar{\gamma})} + \frac{C_2(\bar{\gamma})(1-\tau_n)}{g(\bar{\gamma})}\right)^{-1} - 1,$$

$$\bar{\mathcal{R}}_{n,2} = \frac{C_1(\bar{\gamma}, \bar{\rho}) \cdot \bar{\gamma} \bar{b} \left(1-\tau_n'\right)^{-\bar{\rho}}}{g(\bar{\gamma})} + \frac{C_2(\bar{\gamma})(1-\tau_n')}{g(\bar{\gamma})},$$

$$\bar{\mathcal{R}}_{n,3} = \frac{\bar{\gamma}\bar{b}}{\bar{\rho}} \left[\left(1-\tau_n'\right)^{-\bar{\rho}} - (1-\tau_n)^{-\bar{\rho}}\right].$$

Thus, from Eq. (18), we propose a bias-reduced version of $\widehat{\xi}_{\tau_n'}^{M,\star}(\varepsilon)$ as follows:

$$\widehat{\xi}_{\tau_n'}^{M,\star,\mathrm{RB}}(\varepsilon) = \widehat{\xi}_{\tau_n'}^{M,\star}(\varepsilon) \left(1 + \bar{\mathcal{R}}_{n,1}\right) \left(1 + \bar{\mathcal{R}}_{n,2}\right) \left(1 + \bar{\mathcal{R}}_{n,3}\right).$$

Analogously, we apply another multiplicative correction for the indirect estimator based on quantile, that is

$$\widehat{\xi}_{\tau_n'}^{Q,\star,\mathrm{RB}}(\varepsilon) = \widehat{\xi}_{\tau_n'}^{Q,\star}(\varepsilon) \left(1 + \bar{\mathcal{R}}_{n,2}\right) \left(1 + \bar{\mathcal{R}}_{n,3}\right).$$

The next result provides the asymptotic properties of these bias-reduced estimators $\widehat{\xi}_{\tau_n'}^{M,\star,\mathrm{RB}}(\varepsilon)$ and $\widehat{\xi}_{\tau_n'}^{Q,\star,\mathrm{RB}}(\varepsilon)$.

Theorem 4.1. *Suppose that $\mathbb{E}|\varepsilon_-| < \infty$ and $\mathcal{C}_2(\gamma, \rho, A)$ holds with $A(t) = b\gamma t^{\rho}$, $0 < \gamma < 1$ and $\rho < 0$. The sequences $\tau_n \uparrow 1$ and $\tau_n' \uparrow 1$ satisfy condition (11) and the residuals $\widehat{\varepsilon}_i$ satisfy condition (7). Further assume that $\sqrt{n(1-\tau_n)}(\bar{\gamma} - \gamma) \xrightarrow{d} \Psi$, where Ψ is nondegenerate. Let consistent estimators $\bar{\rho}$ of ρ and $\bar{b}$ of b be such that $(\bar{\rho} - \rho)\log(n) = o_{\mathbb{P}}(1)$.*

(i) *If $0 < \gamma < 1/2$, $\mathcal{C}_3$ and condition (12) hold, then*

$$\frac{\sqrt{n(1-\tau_n)}}{\log\left[(1-\tau_n)/\left(1-\tau_n'\right)\right]} \left(\frac{\widehat{\xi}_{\tau_n'}^{M,\star,\mathrm{RB}}(\varepsilon)}{\xi_{\tau_n'}(\varepsilon)} - 1\right) \xrightarrow{d} \Psi \quad \text{as} \quad n \to \infty.$$

(ii) *If $\sqrt{n(1-\tau_n)} A\left((1-\tau_n)^{-1}\right) \to \lambda_1 \in \mathbb{R}$, then*

$$\frac{\sqrt{n(1-\tau_n)}}{\log\left[(1-\tau_n)/\left(1-\tau_n'\right)\right]} \left(\frac{\widehat{\xi}_{\tau_n'}^{Q,\star,\mathrm{RB}}(\varepsilon)}{\xi_{\tau_n'}(\varepsilon)} - 1\right) \xrightarrow{d} \Psi \quad \text{as} \quad n \to \infty.$$

Remark 5. Theorem 4.1 states that the two bias-reduced estimators $\widehat{\xi}_{\tau_n'}^{M,\star,\mathrm{RB}}(\varepsilon)$ and $\widehat{\xi}_{\tau_n'}^{Q,\star,\mathrm{RB}}(\varepsilon)$ have a limiting distribution controlled by the tail index estimator $\bar{\gamma}$, similar to their naive counterparts. The weak condition $(\bar{\rho} - \rho)\log(n) = o_{\mathbb{P}}(1)$ in Theorem 4.1 is used to control $\bar{\mathcal{R}}_{n,1}$ and $\bar{\mathcal{R}}_{n,2}$ and $\bar{\mathcal{R}}_{n,3}$.

We now construct bias-reduced extreme conditional estimators, that is

$$\widehat{\xi}_{\tau_n'}^{M,\star,\mathrm{RB}}(Y \mid \boldsymbol{x}) = \bar{m}(\boldsymbol{x}) + \bar{\sigma}(\boldsymbol{x}) \cdot \left(\frac{1-\tau_n'}{1-\tau_n}\right)^{-\bar{\gamma}} \widehat{\xi}_{1-k/n}^{M}(\varepsilon) \left(1 + \bar{\mathcal{R}}_{n,1}\right) \left(1 + \bar{\mathcal{R}}_{n,2}\right) \left(1 + \bar{\mathcal{R}}_{n,3}\right)$$

$$\text{and} \quad \widehat{\xi}_{\tau_n'}^{Q,\star,\mathrm{RB}}(Y \mid \boldsymbol{x}) = \bar{m}(\boldsymbol{x}) + \bar{\sigma}(\boldsymbol{x}) \cdot \left(\frac{1-\tau_n'}{1-\tau_n}\right)^{-\bar{\gamma}} g(\widehat{\gamma}_k^{\mathrm{H,RB}}) \widehat{\varepsilon}_{n-k,n} \left(1 + \bar{\mathcal{R}}_{n,2}\right) \left(1 + \bar{\mathcal{R}}_{n,3}\right)$$

with the intermediate sequence $k = \widehat{k}$. It follows by Theorem 4.1 straightly that the bias-reduced estimators $\widehat{\xi}_{\tau_n'}^{M,\star,\mathrm{RB}}(Y \mid \boldsymbol{x})$ and $\widehat{\xi}_{\tau_n'}^{Q,\star,\mathrm{RB}}(Y \mid \boldsymbol{x})$ have the same asymptotic behaviors as their original versions.

5. Simulation study

Some examples of heteroscedastic regression models are considered and the performances of our conditional estimators on each situation are explored in this section.

Mean-variance regression model: We first study the conditional mean and variance functions m and σ^2 in model (2) with parametric forms such that

$$Y = \alpha + \boldsymbol{\beta}^{\top} \boldsymbol{X} + s\left(\boldsymbol{\theta}^{\top} \boldsymbol{X} + 1\right) \varepsilon.$$

Here, the positive scale function $s(\cdot)$ is given e.g. $s(t) = t$ or $s(t) = \exp(t)$ and the vector $\boldsymbol{\theta}$ satisfies $\mathbb{P}\left(s\left(\boldsymbol{\theta}^{\top} \boldsymbol{X} + 1\right) > 0\right) = 1$. The task is to estimate the parameters α, $\boldsymbol{\beta}$ and $\boldsymbol{\theta}$. Here, we apply a simultaneous mean-variance regression loss function which is a globally convex problem introduced by Spady and Stouli (2018b):

$$(\widehat{\alpha}, \widehat{\boldsymbol{\beta}}, \widehat{\boldsymbol{\theta}}) := \underset{\alpha \in \mathbb{R},\ \boldsymbol{\beta}, \boldsymbol{\theta} \in \mathbb{R}^d}{\arg\min} \frac{1}{n} \sum_{i=1}^{n} \frac{1}{2} \left\{ \left(\frac{y_i - \alpha - \boldsymbol{\beta}^\top \boldsymbol{X}_i}{s\left(\boldsymbol{\theta}^\top \boldsymbol{X}_i + 1\right)} \right)^2 + 1 \right\} s\left(\boldsymbol{\theta}^\top \boldsymbol{X}_i + 1\right), \quad \text{s.t.} \quad s\left(\boldsymbol{\theta}^\top \boldsymbol{X}_i + 1\right) > 0, \quad i = 1, \ldots, n. \tag{19}$$

Under suitable assumptions, especially for the conditional variance function $\sigma(\boldsymbol{x})^2 = s(\boldsymbol{\theta}^\top \boldsymbol{x} + 1)^2$, the estimators $\widehat{\alpha}$ of α, $\widehat{\boldsymbol{\beta}}$ of $\boldsymbol{\beta}$ and $\widehat{\boldsymbol{\theta}}$ of $\boldsymbol{\theta}$ are $\sqrt{n}$-asymptotically normal (see Theorem 6 of Spady and Stouli, 2018b). We use an algorithm called constrained optimization by linear approximation (COBYLA), for derivative-free optimization implemented with *nloptr* package. Plugging in these consistent estimators, one can write the following bias-reduced versions of the M-estimator and Q-estimator:

$$\widehat{\xi}^{M,\star}_{\tau_n'}(Y \mid \boldsymbol{x}) = \widehat{\alpha} + \widehat{\boldsymbol{\beta}}^\top \boldsymbol{x} + s\left(\widehat{\boldsymbol{\theta}}^\top \boldsymbol{x} + 1\right) \cdot \widehat{\xi}^{M,\star,\mathrm{RB}}_{\tau_n'}(\varepsilon)$$
$$\text{and} \quad \widehat{\xi}^{Q,\star}_{\tau_n'}(Y \mid \boldsymbol{x}) = \widehat{\alpha} + \widehat{\boldsymbol{\beta}}^\top \boldsymbol{x} + s\left(\widehat{\boldsymbol{\theta}}^\top \boldsymbol{x} + 1\right) \cdot \widehat{\xi}^{Q,\star,\mathrm{RB}}_{\tau_n'}(\varepsilon).$$

Heteroscedastic single-index model: Concentrating on a more flexible model, we assume that both the conditional mean function and conditional variance function fit various single-index structures,

$$Y = m\left(\boldsymbol{\beta}^\top \boldsymbol{X}\right) + \sigma\left(\boldsymbol{\theta}^\top \boldsymbol{X}\right)\varepsilon,$$

where the support S has a nonempty interior S°. Both the conditional mean function $m(\cdot)$ and variance function $\sigma^2(\cdot)$ admit the single-index structure where the direction of projections in $m(\cdot)$ and $\sigma^2(\cdot)$ are the vectors β and θ, respectively. The vectors $\boldsymbol{\beta}$ and $\boldsymbol{\theta}$ are mainly estimated using the algorithm in section 2.2 of Zhu et al. (2013). A slight improvement is the use of *orthoDr* package in Zhu et al. (2018) which adapts the algorithm proposed by Wen and Yin (2013) for optimization and is comparable with existing manifold optimization approaches instead of using Newton's method. Then, we estimate m with

$$\widehat{m}\left(\widehat{\boldsymbol{\beta}}^{\mathrm{T}} \mathbf{x}_i\right) = \frac{\sum_{j=1}^{n} Z_{h_1}\left(\widehat{\boldsymbol{\beta}}^{\mathrm{T}} \mathbf{x}_j - \widehat{\boldsymbol{\beta}}^{\mathrm{T}} \mathbf{x}_i\right) Y_j \mathbb{1}\left\{\left|Y_j\right| \le t_n\right\}}{\sum_{j=1}^{n} Z_{h_1}\left(\widehat{\boldsymbol{\beta}}^{\mathrm{T}} \mathbf{x}_j - \widehat{\boldsymbol{\beta}}^{\mathrm{T}} \mathbf{x}_i\right)},$$

where $Z_h(\cdot) = Z(./h)/h$ is the kernel function and $t_n \to \infty$ is a positive truncating sequence. Next we calculate the residual $\widehat{\zeta}_i = Y_i - \widehat{m}\left(\widehat{\boldsymbol{\beta}}^{\mathrm{T}} \mathbf{x}_i\right)$ to obtain the estimator

$$\widehat{\sigma}^2\left(\widehat{\boldsymbol{\theta}}^{\mathrm{T}} \mathbf{x}_i\right) = \frac{\sum_{j=1}^{n} Z_{h_2}\left(\widehat{\boldsymbol{\theta}}^{\mathrm{T}} \mathbf{x}_j - \widehat{\boldsymbol{\theta}}^{\mathrm{T}} \mathbf{x}_i\right) \widehat{\zeta}_j^2 \mathbb{1}\left\{\left|\widehat{\zeta}_j\right| \le t_n\right\}}{\sum_{j=1}^{n} Z_{h_2}\left(\widehat{\boldsymbol{\theta}}^{\mathrm{T}} \mathbf{x}_j - \widehat{\boldsymbol{\theta}}^{\mathrm{T}} \mathbf{x}_i\right)}.$$

It shows that both of these estimators converge uniformly at the rate $n^{2/5}/\sqrt{\log n}$ under the condition $nh^5 \to a \in (0, \infty)$ (see supplementary material of Girard et al., 2021; Zhu et al., 2018). By Theorem 3.4 and Corollary 3.1, it yields that $k = n^c$ where $c \in (0, \frac{4}{5})$, and n and S are replaced by $N = \sum_{i=1}^{n} \mathbb{1}\{\boldsymbol{X}_i \in S_0\}$ and S_0, respectively, to ensure the asymptotic property of the extreme estimators (see Lemma C.4-C.6 in the supplementary material of Girard et al. (2021))

$$\widehat{\xi}^{M,\star}_{\tau_N'}(Y \mid \boldsymbol{x}) = \widehat{m}\left(\widehat{\boldsymbol{\beta}}^\top \boldsymbol{x}\right) + \widehat{\sigma}\left(\widehat{\boldsymbol{\theta}}^\top \boldsymbol{x}\right) \left(\frac{1 - \tau_N'}{1 - \tau_N}\right)^{-\bar{\gamma}} \cdot \widehat{\xi}_{\tau_N}(\varepsilon)$$
$$\text{and} \quad \widehat{\xi}^{Q,\star}_{\tau_N'}(Y \mid \boldsymbol{x}) = \widehat{m}\left(\widehat{\boldsymbol{\beta}}^\top \boldsymbol{x}\right) + \widehat{\sigma}\left(\widehat{\boldsymbol{\theta}}^\top \boldsymbol{x}\right) \left(\frac{1 - \tau_N'}{1 - \tau_N}\right)^{-\bar{\gamma}} \cdot g(\bar{\gamma}) \widehat{\varepsilon}^{(N)}_{N - \lfloor N(1-\tau_N) \rfloor, N}.$$

In this section, we propose to study the finite-sample performance of the estimators and their bias-reduced versions. For this purpose, we consider the linear and single-index models:

(M1) $Y = \boldsymbol{\beta}^\top \boldsymbol{X} + 2 + \left(\boldsymbol{\theta}^\top \boldsymbol{X} + 1\right)\varepsilon$,
(M2) $Y = \exp\left(\boldsymbol{\beta}^\top \boldsymbol{X} - 2\right) + 1 + \left(\exp\left(\boldsymbol{\theta}^\top \boldsymbol{X} - 1\right) + 1/2\right)\varepsilon$.

In the models, the covariate $\boldsymbol{X} \in \mathbb{R}^4$ consists of four independent components, the first two following a uniform distribution U(0, 1), and the last two following a Beta(2, 2) distribution. Set the coefficient vectors $\boldsymbol{\beta}^\top = (1, 1, 2, 2)$, $\boldsymbol{\theta}^\top = (1, 1, 1, 1)$. The noise ε is independent of $\boldsymbol{X}$ and follows a generalized Pareto distribution, that is $\varepsilon = \varepsilon_0 \cdot \sqrt{(1/2 - \gamma)(1 - \gamma)}$ where ε_0 has density $f_0(x) = (1 + \gamma|x|)^{-1/\gamma - 1}/2$.

We simulate $N = 1,000$ data sets of $n = 1,000$ observations $(\boldsymbol{X}_i, Y_i)$, $1 \le i \le n$ and consider the cases $\gamma \in \{0.1, 0.15, 0.2, 0.25, 0.3, 0.35\}$ with the second-order parameter $\rho = -\gamma$. We estimate the conditional extremile of level $\tau_n' = 0.995$ fixed $\boldsymbol{x}^\top = (2/5, 2/5, 1/2, 1/2)$ using four methods:

(1) Use the bias-reduced direct estimator $\widehat{\xi}^{M,\star,\mathrm{RB}}_{\tau_n'}(Y \mid \boldsymbol{x})$ with the bias-reduced Hill estimator $\widehat{\gamma}_k^{\mathrm{H,RB}}$ and $k = \widehat{k}$.
(2) Assume $\boldsymbol{\beta}$ and $\boldsymbol{\theta}$ are known and use the bias-reduced direct estimator $\widehat{\xi}^{M,\star,\mathrm{RB}}_{\tau_n'}(Y \mid \boldsymbol{x})$ with the bias-reduced Hill estimator $\widehat{\gamma}_k^{\mathrm{H,RB}}$ and $k = \widehat{k}$.

Table 1
RMSE and RMAD of extreme conditional extremile estimators in (M1).

Evaluation	Methodology	$\gamma = 0.1$	$\gamma = 0.15$	$\gamma = 0.2$	$\gamma = 0.25$	$\gamma = 0.3$	$\gamma = 0.35$
RMSE	(1)	**0.0182**	**0.0223**	0.0276	0.0349	0.0463	0.0986
	(2)	0.0195	0.0231	**0.0271**	**0.0321**	**0.0376**	**0.0439**
	(3)	0.0251	0.0284	0.0320	0.0361	0.0408	0.0460
	(4)	0.1070	0.1228	0.1402	0.1590	0.1791	0.1995
	(1i)	**0.0334**	**0.0392**	0.0468	0.0572	0.0739	0.1041
	(2i)	0.0351	0.0400	**0.0450**	**0.0517**	**0.0584**	**0.0658**
	(3i)	0.0456	0.0495	0.0538	0.0585	0.0637	0.0694
	(4i)	0.2729	0.3236	0.3863	0.4650	0.5656	0.6967
RMAD	(1)	**0.0867**	**0.0959**	**0.1028**	**0.1049**	0.1147	0.1219
	(2)	0.0935	0.0983	0.1047	0.1088	**0.1138**	**0.1177**
	(3)	0.1091	0.1119	0.1132	0.1174	0.1186	0.1179
	(4)	0.2759	0.2907	0.3046	0.3150	0.3219	0.3241
	(1i)	**0.1324**	**0.1361**	0.1451	0.1462	0.1559	0.1625
	(2i)	0.1410	0.1396	**0.1439**	**0.1452**	**0.1457**	**0.1455**
	(3i)	0.1659	0.1646	0.1622	0.1600	0.1560	0.1521
	(4i)	0.4573	0.4857	0.5179	0.5521	0.5912	0.6229

M1: $Y = \boldsymbol{\beta}^\top \boldsymbol{X} + 2 + \left(\boldsymbol{\theta}^\top \boldsymbol{X} + 1\right)\varepsilon$

Table 2
RMSE and RMAD of extreme conditional extremile estimators in (M2).

Evaluation	Methodology	$\gamma = 0.1$	$\gamma = 0.15$	$\gamma = 0.2$	$\gamma = 0.25$	$\gamma = 0.3$	$\gamma = 0.35$
RMSE	(1)	0.0199	0.0228	0.0257	0.0292	0.0342	0.0411
	(2)	**0.0164**	**0.0191**	**0.0225**	**0.0266**	**0.0330**	**0.0406**
	(3)	0.0205	0.0226	0.0256	0.0293	0.0347	0.0417
	(4)	0.0876	0.0973	0.1077	0.1218	0.1344	0.1509
	(1i)	0.0353	0.0386	0.0418	0.0457	0.0509	0.0752
	(2i)	**0.0271**	**0.0302**	**0.0334**	**0.0380**	**0.0451**	**0.0529**
	(3i)	0.0361	0.0378	0.0403	0.0435	0.0485	0.0554
	(4i)	0.2317	0.2652	0.3053	0.3646	0.4274	0.5263
RMAD	(1)	0.0911	0.0973	0.1010	0.1065	**0.1124**	**0.1220**
	(2)	**0.0832**	**0.0941**	**0.0975**	**0.1065**	0.1170	0.1250
	(3)	0.0951	0.0981	0.1044	0.1065	0.1126	0.1239
	(4)	0.2384	0.2425	0.2445	0.2433	0.2380	0.2217
	(1i)	0.1313	0.1269	0.1327	0.1330	0.1329	0.1355
	(2i)	**0.1059**	**0.1121**	**0.1147**	**0.1200**	**0.1263**	**0.1316**
	(3i)	0.1312	0.1273	0.1247	0.1275	0.1258	0.1292
	(4i)	0.4131	0.4289	0.4410	0.4544	0.4659	0.4692

M2: $Y = \exp\left(\boldsymbol{\beta}^\top \boldsymbol{X} - 2\right) + 1 + \left(\exp\left(\boldsymbol{\theta}^\top \boldsymbol{X} - 1\right) + 1/2\right)\varepsilon$

(3) Assume $\boldsymbol{\beta}$ and $\boldsymbol{\theta}$ are known and use the bias-reduced direct estimator $\widehat{\xi}^{M,\star,\mathrm{RB}}_{\tau_n'}(Y \mid \boldsymbol{x})$ with the bias-reduced Hill estimator $\widehat{\gamma}_k^{\mathrm{H,RB}}$ where the intermediate sequence is fixed $k = 100$.
(4) Assume $\boldsymbol{\beta}$ and $\boldsymbol{\theta}$ are known and use the direct estimator $\widehat{\xi}^{M,\star}_{\tau_n'}(Y \mid \boldsymbol{x})$ with the Hill estimator $\widehat{\gamma}_k^{\mathrm{H}}$ and $k = 100$.

Similar to methods (1)–(4), there are four methodologies for the indirect estimation numbered (1i)–(4i). To evaluate the performances of these strategies, we compute the relative mean squared error (RMSE) and the relative median absolute deviation (RMAD):

$$\mathrm{RMSE} = \frac{1}{N}\sum_{m=1}^{N}\left(\frac{\bar{\xi}^{\star}_{\tau_n',m}(Y \mid \boldsymbol{x})}{\xi_{\tau_n'}(Y \mid \boldsymbol{x})} - 1\right)^2 \text{ and } \mathrm{RMAD} = \mathrm{median}_{1\le m\le N}\left|\frac{\bar{\xi}^{\star}_{\tau_n',m}(Y \mid \boldsymbol{x})}{\xi_{\tau_n'}(Y \mid \boldsymbol{x})} - 1\right|,$$

where $\bar{\xi}^{\star}_{\tau_n',m}(Y \mid \boldsymbol{x})$ denotes the consistent extrapolated estimator calculated on the m-th replication. The results are presented in Tables 1 and 2.

In both (M1) and (M2), it appears that the relative errors are mainly influenced by the tail heaviness. In particular, as the tail index γ becomes larger, the RMSE and RMAD both increase in all cases. A comparison of (1)–(4) and (1i)–(4i) shows that the direct method is better than the indirect method. On the whole, method (2) performs the best, followed by method (1), especially in the single-index model (M2). Comparing (1) with (2), we find that the model structure estimation impacts relatively the accuracy of the extreme conditional estimation. From (2) and (3), it can be seen that the optimal choice of the intermediate sequence $k(n)$ is beneficial for the lighter-tailed model. We can conclude that the bias-reduction correction is an obvious advantage, as the comparison of (3) and (4) indicates. These conclusions are also applicable to the indirect method.

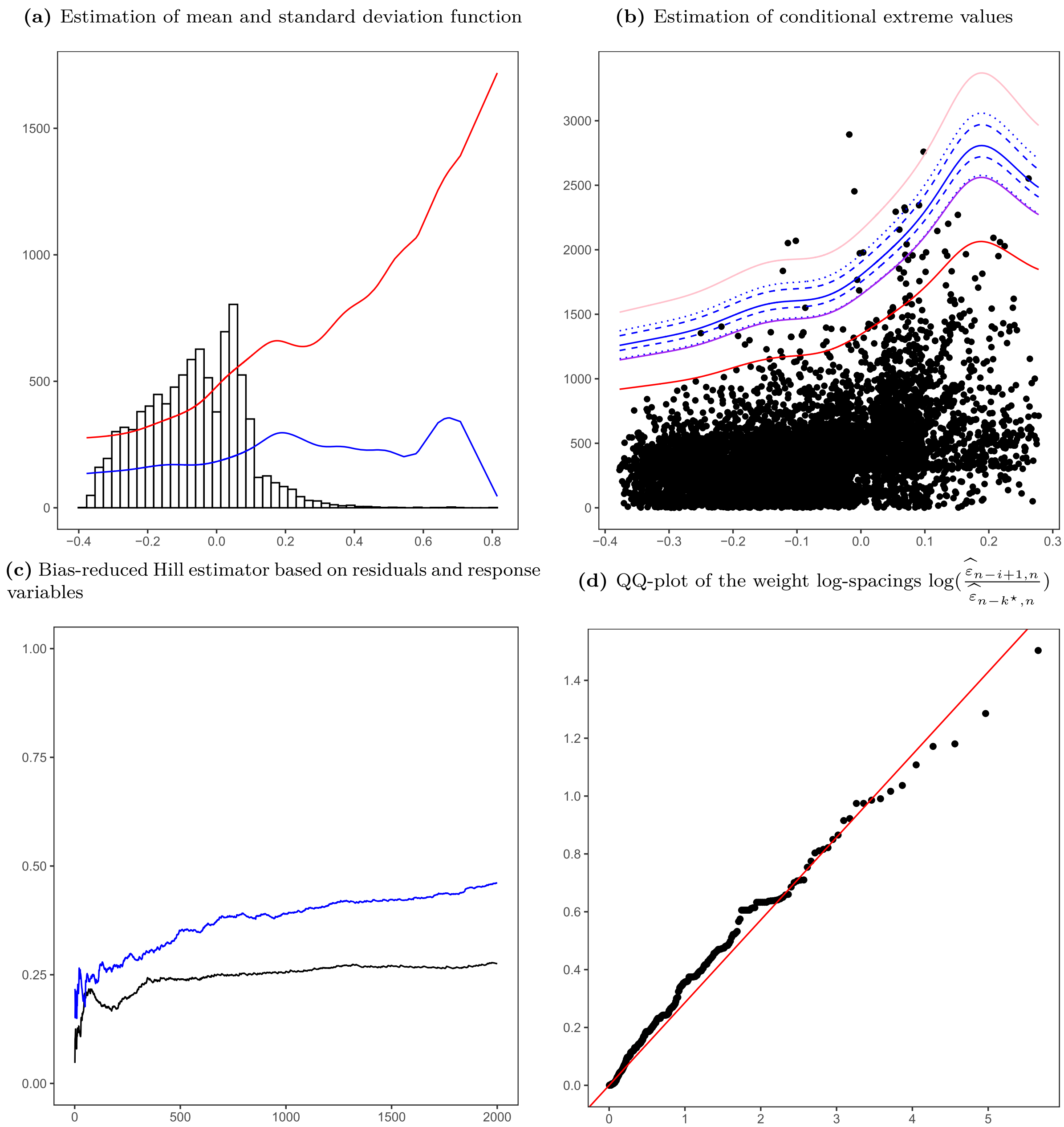

Fig. 1. Vehicle Insurance Data: (a): estimation of the mean function m (red line) and standard deviation function σ (blue line) with a histogram of the $\widehat{\boldsymbol{\beta}}^\top \boldsymbol{X}_i$. (b): estimator of the conditional extremile in the direct method (blue solid line) with bootstrap confidence intervals (blue dashed line) and Gaussian confidence intervals (blue dotted line), ES (pink line), quantile (purple line) and expectile (red line) at level $\tau_n' = 0.9977$ on the $\widehat{\boldsymbol{\beta}}^\top \boldsymbol{x} - y$ (scatter) plot. (c): bias-reduced Hill estimator based on residuals $\widehat{\varepsilon}_i$ (blue line) and response variables Y_i (black line). (d): Exponential QQ-plot of the weight log-spacings $\log\left(\widehat{\varepsilon}_{n-i+1,n}/\widehat{\varepsilon}_{n-\widehat{k},n}\right)$ for all $1 \le i \le \widehat{k} = 286$ with red line whose slope is $\widehat{\gamma}_{\widehat{k}}^{\mathrm{H,RB}} = 0.2858$. (For interpretation of the colors in the figure(s), the reader is referred to the web version of this article.)

6. Real data study

6.1. Vehicle insurance customer data

The vehicle insurance customer data set includes $n = 9,134$ customers according to their claim amounts Y, lifetime value X_1 (in usd), income X_2 (in usd), months since last claim X_3 and months since policy inception X_4. We propose to estimate the extreme extremile of the claim amount for each customer by applying the heteroscedastic single-index model to fit it. Before that, we shall test whether the conditional tail index is constant with the statistic T_4 on response variables $Y_1, \ldots, Y_n$ which is introduced in Einmahl et al. (2016). Under the null hypothesis of a constant conditional tail index, we have $k_n T_4 \xrightarrow{d} \chi^2_{m-1}$ (see Einmahl et al., 2016). Here, we choose the number of blocks $m = 4$ and calculate the test statistic $T_4 = 0.041$ and p-value is approximately 0.998. Hence, we cannot reject the hypothesis that the extreme value index γ is invariant.

For simplicity of calculation, we assume that single-index vectors $\boldsymbol{\beta}$ and $\boldsymbol{\theta}$ are equivalent. The kernel function $Z_h(\cdot)$ is the Gaussian kernel. We choose the bandwidth $h = h_1 = h_2 = n^{-1/5}/\sqrt{12} \approx 0.047$ and $t_n = \infty$ and obtain $\widehat{\boldsymbol{\beta}} \simeq (0.9146, -0.4043, -0.0040, -0.0005)^\top$

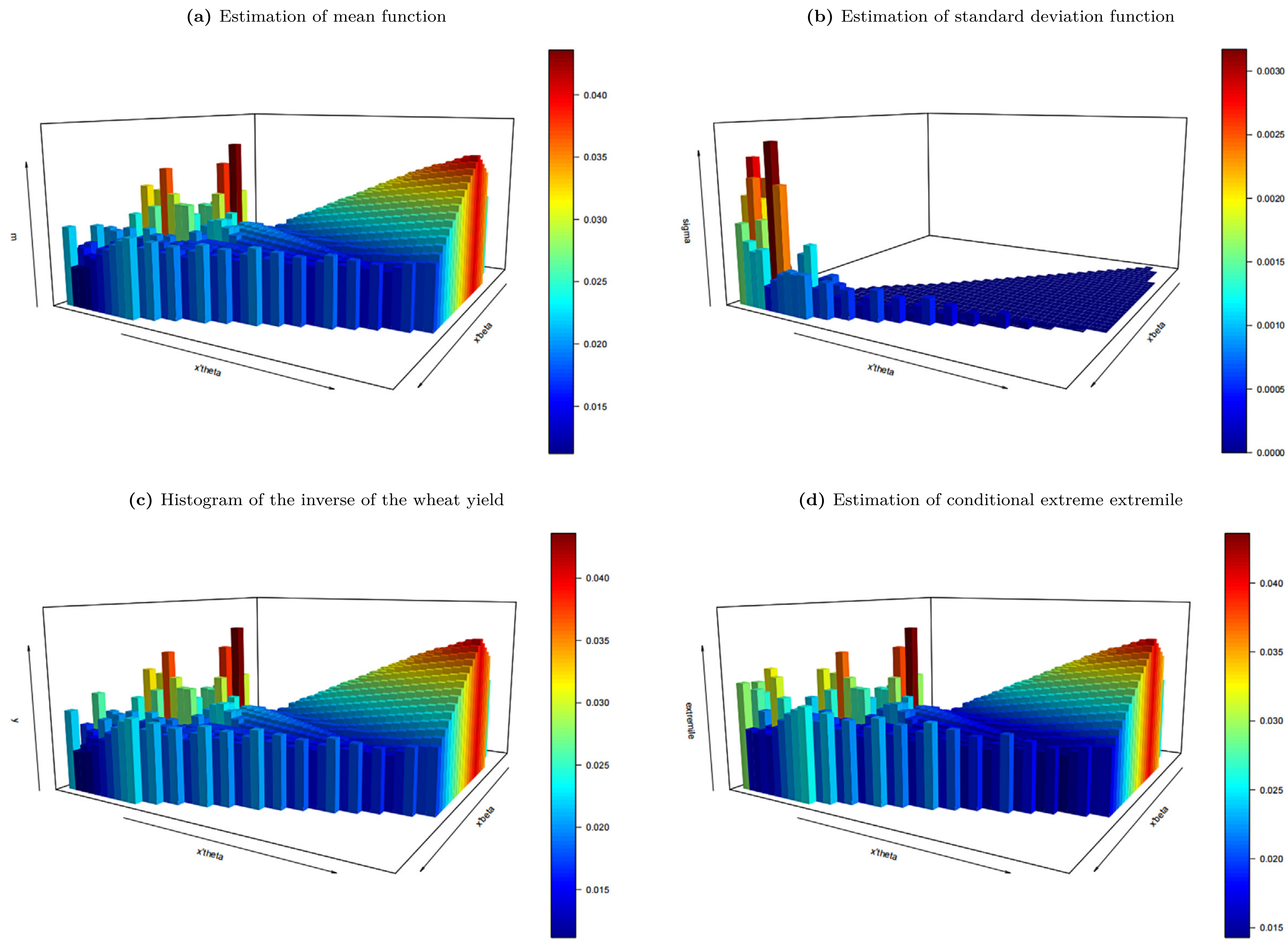

Fig. 2. Farm Data:(a): estimation of mean function $m(\cdot)$. (b): estimation of standard deviation function $\sigma(\cdot)$. (c): histogram of the inverse of the wheat yield $Y_1, \cdot, Y_n$. (d): conditional extremile direct estimators $\widehat{\xi}^{M,\star,\mathrm{RB}}_{\tau_n'}(Y \mid \boldsymbol{x_1}), \cdot, \widehat{\xi}^{M,\star,\mathrm{RB}}_{\tau_n'}(Y \mid \boldsymbol{x_n})$. On all the three-dimensional diagrams, the x'theta axis is $\boldsymbol{x}^\top\widehat{\boldsymbol{\theta}}$ and the x'beta axis is $\boldsymbol{x}^\top\widehat{\boldsymbol{\beta}}$. As the color changes from blue to red, the corresponding value increases.

(a) Bias-reduced Hill estimator based on residuals and response variables

(b) QQ-plot of the weight log-spacings $\log(\frac{\widehat{\varepsilon}_{n-i+1,n}}{\widehat{\varepsilon}_{n-k^\star,n}})$

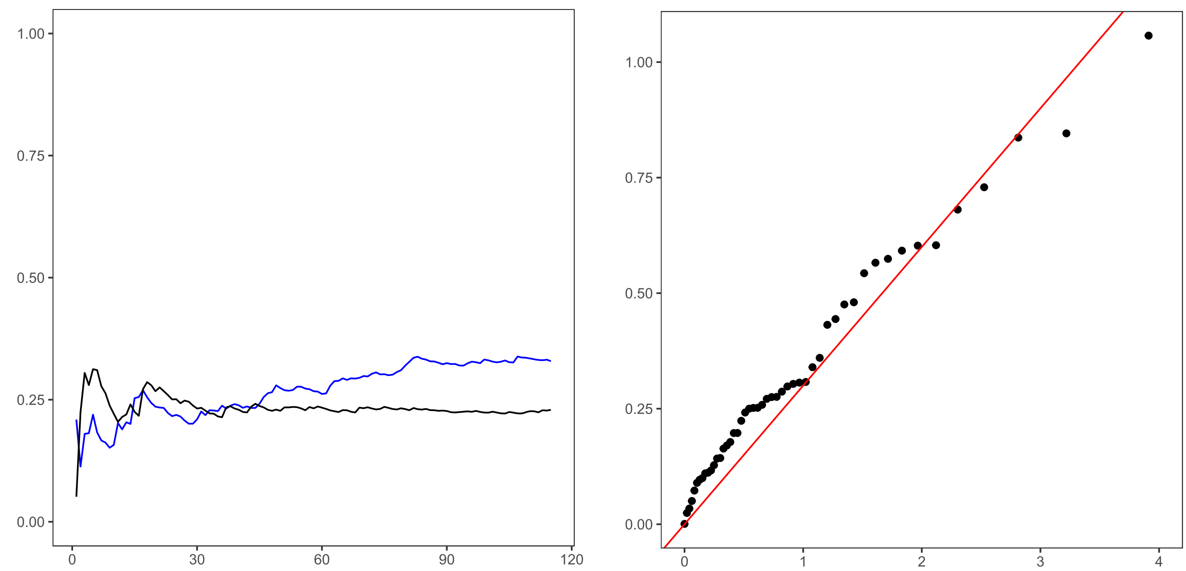

Fig. 3. Farm Data: (a): bias-reduced Hill estimator based on residuals $\widehat{\varepsilon}_i$ (blue line) and response variables Y_i (black line). (b): Exponential QQ-plot of the weight log-spacings $\log\left(\widehat{\varepsilon}_{n-i+1,n}/\widehat{\varepsilon}_{n-\widehat{k},n}\right)$ for all $1 \le i \le \widehat{k} = 70$ with red line whose slope is $\widehat{\gamma}_k^{\mathrm{H,RB}} = 0.2985$.

implying that the first two variables are crucial contributors for predicting the claim amount Y. The estimates of mean and standard deviation function $\widehat{m}$ and $\widehat{\sigma}$ are illustrated in Fig. 1(a). Obtaining the residuals $\widehat{\varepsilon}_i, \ldots, \widehat{\varepsilon}_n$, the assumption of a heavy-tailed distribution is required to be confirmed. Fig. 1(c) gives the bias-reduced Hill estimators based on residuals and response variables. It appears that the former is more stable than the latter. The bias-reduced Hill estimator $\widehat{\gamma}_k^{\mathrm{H,RB}}$ on $\widehat{\varepsilon}_i, i = 1, \ldots, n$ is approximately equal to 0.2858 with the intermediate sequence $k = \widehat{k} = 286$ by the method of Section 4.1. Fig. 1(d) shows that the points on the QQ-plot lie roughly near the line with the slope $\widehat{\gamma}_k^{\mathrm{H,RB}}$, which verifies the plausibility of the heavy-tailed assumption. Finally, we estimate the conditional extreme extremile of each customer and compare it with three other common risk measures: quantile, expectile and ES, at the same level $\tau_n' = 1 - 1/(nh) \approx 0.9977$. The results are reported in Fig. 1(b). These four curves are similar in shape. More specifically, the extreme estimator $\widehat{\xi}_{\tau_n'}^{M,\star,\mathrm{RB}}(Y \mid \boldsymbol{x})$ is more pessimistic than quantile and expectile, but is more optimistic than ES following the asymptotic proportional relationship between extreme extremile and the three risk measures (see Propositions 3 and 6 in Daouia et al., 2019). Taking into account uncertainty, we offer two pointwise 95% confidence intervals: asymptotic Gaussian confidence interval and bootstrap confidence interval for the conditional extreme extremile $\xi_{\tau_n'}(Y \mid \boldsymbol{x})$. As a comparison, we find the bootstrap confidence interval is less conservative than the Gaussian confidence interval. In addition, the lower asymptotic Gaussian confidence band is close to the conditional extreme quantile. The generation of confidence intervals is explained in detail in Appendix C.

6.2. Farm income data

The farm income data set consists of $n = 949$ observations of 6 variables namely: the inverse of the wheat yield Y (in quintal/hectare), selling price X_1 (in euro/quintal), works and services purchased for crops X_2 (in euro), the real cost of seeds and seedlings X_3 (in euro), purchases of works and services for crops X_4 (in euro) and temperature X_5 ($^\circ C$) (see Bousebata et al., 2023). For the response variable Y, the test statistic is $T_4 = 0.6264$ and the p-value is 0.8904 which provides evidence that we cannot reject the null hypothesis of constant extreme value index. Fig. 2(c) displays the inverse of the wheat yield $Y_1, \ldots, Y_n$. We consider the farm income data for the second application on the single-index model. We use the Gaussian kernel and set the bandwidth $h = n^{-1/5} \approx 0.2538$ to calculate the estimators of the index parameters $\widehat{\boldsymbol{\beta}} \simeq (0.7082, 0.4180, -0.4130, 0.1633, -0.3556)^\top$ and $\widehat{\boldsymbol{\theta}} \simeq (-0.1990, 0.9634, 0.1364, 0.0248, -0.1142)^\top$. The estimation results for $\widehat{m}(\cdot)$ and $\widehat{\sigma}(\cdot)$ are plotted in Fig. 2(a) and Fig. 2(b), respectively. Fig. 2(b) indicates the heteroscedasticity via the standard deviation function. The bias-reduced version of the Hill estimator is graphed in Fig. 3(a) both on residuals and response data. The bias-reduced Hill estimator $\widehat{\gamma}_k^{\mathrm{H,RB}}$ is 0.2985 with the intermediate level $\widehat{k} = 70$ computed on the residuals. As can be seen from Fig. 3(b), the exponential quantile–quantile plot is approximately linear which confirms that the residuals are heavy-tailed. The extreme conditional extremile estimator $\widehat{\xi}_{\tau_n'}^{M,\star,\mathrm{RB}}(Y \mid \boldsymbol{x})$ is evaluated at the level $\tau_n' = 1 - 1/(nh) \approx 0.9958$, which is shown in Fig. 2(d). As a comparison of Figs. 2(c) and 2(d), the output variable and the extreme extremile estimator from the plane $(\boldsymbol{x}^\top \widehat{\boldsymbol{\theta}}, \boldsymbol{x}^\top \widehat{\boldsymbol{\beta}})$ have similar behavior in shape. In addition, we estimate the extreme conditional quantile and observe that there are four samples where Y_i exceeds $\widehat{q}^{\star}_{\tau_n'}(Y \mid \boldsymbol{X}_i)$. It's exactly the same as the theoretical value $n(1 - \tau_n') \simeq 4$.

7. Conclusion

This paper develops a general theory for the estimation of extreme conditional extremiles in the framework of a heteroscedastic regression model with heavy-tail noises. Estimating extremiles for heavy-tailed variables under a regression framework is a challenging task, and the existing study remains limited to independent and identically distributed heavy-tailed observations. Our approach can estimate the extremile in the framework of the heteroscedastic regression model by constructing extreme conditional extremile estimators based on available residuals. According to the methods used, the estimators fall into direct and indirect estimators, and their theoretical properties are explored. Moreover, taking the bias of the extreme estimators into consideration, we provided the bias-reduced version of the estimators by a two-aspect procedure: bias reduction for the tail-index estimator and for Wessiman's extrapolation device. Finally, finite-sample simulation is conducted and reflects that tail heaviness affects the relative errors and bias correction can greatly improve the accuracy. From the analysis of real data sets, we find the extreme conditional extremile estimator is more pessimistic than quantile and expectile but is more optimistic than ES in practice.

Declaration of competing interest

The authors declare that there is no competing interest.

Data availability

Data will be made available on request.

Acknowledgements

The authors thank the editor and the anonymous referees for their helpful comments, which greatly improved the quality of this paper. The work is supported by the National Social Science Fund of China (No. 22BTJ027). Author names are listed in alphabetical order.

Appendix A. Auxiliary lemmas

Lemma A.1 states that the convergence rate of the quantity $\max_{1\le i\le n}\frac{|\widehat{\varepsilon}_i-\varepsilon_i|}{1+|\varepsilon_i|}$ can be controlled by appropriate consistent estimators $\bar{m}$ and $\bar{\sigma}$ at the sample points, which is used in the proof of Theorem 3.3. It is an adaptation of Corollary 4.1 in Ahmad et al. (2020).

Lemma A.1. *Assume there exists $\omega_n \to 0$, such that the consistent estimators $\bar{m}$ and $\bar{\sigma}$ satisfy*

$$\max_{1\le i\le n}\left|\frac{\bar{m}(\boldsymbol{X}_i)-m(\boldsymbol{X}_i)}{\sigma(\boldsymbol{X}_i)}\right| = \mathrm{O}_{\mathbb{P}}(\omega_n), \quad \max_{1\le i\le n}\left|\frac{\bar{\sigma}(\boldsymbol{X}_i)}{\sigma(\boldsymbol{X}_i)}-1\right| = \mathrm{O}_{\mathbb{P}}(\omega_n).$$

Then

$$\max_{1\le i\le n}\frac{|\widehat{\varepsilon}_i-\varepsilon_i|}{1+|\varepsilon_i|} = \mathrm{O}_{\mathbb{P}}(\omega_n).$$

Proof. First, analyze the distance between $\widehat{\varepsilon}_i$ and ε_i,

$$\begin{aligned}|\widehat{\varepsilon}_i-\varepsilon_i| &= \left|\frac{\bar{m}(\boldsymbol{X}_i)-m(\boldsymbol{X}_i)}{\sigma(\boldsymbol{X}_i)}+\left(\frac{\bar{\sigma}(\boldsymbol{X}_i)}{\sigma(\boldsymbol{X}_i)}-1\right)\varepsilon_i\right|\cdot\left|\frac{\sigma(\boldsymbol{X}_i)}{\bar{\sigma}(\boldsymbol{X}_i)}\right| \\ &\le (1+|\varepsilon_i|)\max\left\{\left|\frac{\bar{m}(\boldsymbol{X}_i)-m(\boldsymbol{X}_i)}{\sigma(\boldsymbol{X}_i)}\right|,\left|\frac{\bar{\sigma}(\boldsymbol{X}_i)}{\sigma(\boldsymbol{X}_i)}-1\right|\right\}\cdot\left|\frac{\sigma(\boldsymbol{X}_i)}{\bar{\sigma}(\boldsymbol{X}_i)}\right|,\end{aligned}$$

which leads to

$$\begin{aligned}\max_{1\le i\le n}\frac{|\widehat{\varepsilon}_i-\varepsilon_i|}{1+|\varepsilon_i|} &\le \max_{1\le i\le n}\left|\frac{\sigma(\boldsymbol{X}_i)}{\bar{\sigma}(\boldsymbol{X}_i)}\right|\cdot\max\left\{\max_{1\le i\le n}\left|\frac{\bar{m}(\boldsymbol{X}_i)-m(\boldsymbol{X}_i)}{\sigma(\boldsymbol{X}_i)}\right|,\max_{1\le i\le n}\left|\frac{\bar{\sigma}(\boldsymbol{X}_i)}{\sigma(\boldsymbol{X}_i)}-1\right|\right\} \\ &= \mathrm{O}_{\mathbb{P}}(1)\cdot\mathrm{O}_{\mathbb{P}}(\omega_n) = \mathrm{O}_{\mathbb{P}}(\omega_n).\end{aligned}$$

In detail, for all $\delta > 0$, we can write

$$P\left(\max_{1\le i\le n}\left|\frac{\sigma(\boldsymbol{X}_i)}{\bar{\sigma}(\boldsymbol{X}_i)}\right| > 2\right) \le P\left(\max_{1\le i\le n}\left|\frac{\bar{\sigma}(\boldsymbol{X}_i)}{\sigma(\boldsymbol{X}_i)}-1\right| > \frac{1}{2}\right) < \delta$$

to obtain $\max_{1\le i\le n}\left|\frac{\sigma(\boldsymbol{X}_i)}{\bar{\sigma}(\boldsymbol{X}_i)}\right| = \mathrm{O}_{\mathbb{P}}(1)$. □

Lemma A.2 states that the gap between the tail empirical quantile process of the residuals and that of the errors is uniformly bounded, which is used in the proof of Corollary B.1. In effect, it is analogous to Lemma A.3 supplementary material of Girard et al. (2021) so the proof is omitted here.

Lemma A.2. *Suppose that the sequence $\tau_n \uparrow 1$ with $n(1-\tau_n) \to \infty$. If the quantity $\delta_n := \max_{1\le i\le n}\frac{|\widehat{\varepsilon}_i-\varepsilon_i|}{1+|\varepsilon_i|}$, satisfies $\delta_n \xrightarrow{\mathbb{P}} 0$, then*

$$\max_{1\le i\le k}\left|\log\frac{\widehat{\varepsilon}_{n-i,n}}{\varepsilon_{n-i,n}}\right| = \mathrm{O}_{\mathbb{P}}(\delta_n).$$

Appendix B. Proofs of main results

Proof of Theorem 3.1. Denote the unavailable M-statistic $\widetilde{\xi}_{\tau_n}^M(\varepsilon) = \frac{\sum_{i=1}^n J_{\tau_n}\left(\frac{i}{n}\right)\varepsilon_{i,n}}{\sum_{i=1}^n J_{\tau_n}\left(\frac{i}{n}\right)}$, then

$$\begin{aligned}
\left(\widehat{\xi}_{\tau_n}^M(\varepsilon) - \widetilde{\xi}_{\tau_n}^M(\varepsilon)\right) &= \frac{\sum_{i=1}^n J_{\tau_n}\left(\frac{i}{n}\right)\left(\widehat{\varepsilon}_{i,n} - \varepsilon_{i,n}\right)}{\sum_{i=1}^n J_{\tau_n}\left(\frac{i}{n}\right)} \\
&\leq \frac{\sum_{i=1}^n J_{\tau_n}\left(\frac{i}{n}\right)\left(1 + |\varepsilon_{i,n}|\right)\delta_n}{\sum_{i=1}^n J_{\tau_n}\left(\frac{i}{n}\right)} \\
&= \delta_n\left(1 + \frac{\sum_{i=1}^n J_{\tau_n}\left(\frac{i}{n}\right)|\varepsilon_{i,n}|}{\sum_{i=1}^n J_{\tau_n}\left(\frac{i}{n}\right)}\right) \\
&= \delta_n\left(1 + \frac{\sum_{i=1}^n J_{\tau_n}\left(\frac{i}{n}\right)\varepsilon_{i,n} - 2\sum_{i=1}^n J_{\tau_n}\left(\frac{i}{n}\right)\varepsilon_{i,n}1\{\varepsilon_{i,n} \leq 0\}}{\sum_{i=1}^n J_{\tau_n}\left(\frac{i}{n}\right)}\right) \\
&= \delta_n\left(1 + \widetilde{\xi}_{\tau_n}^M(\varepsilon) - \frac{2\sum_{i=1}^n J_{\tau_n}\left(\frac{i}{n}\right)\varepsilon_{i,n}1\{\varepsilon_{i,n} \leq 0\}}{\sum_{i=1}^n J_{\tau_n}\left(\frac{i}{n}\right)}\right) \\
&= \delta_n\widetilde{\xi}_{\tau_n}^M(\varepsilon) + O_{\mathbb{P}}(\delta_n).
\end{aligned}$$

The inequality is due to $\widehat{\varepsilon}_{i,n} - \varepsilon_{i,n} \leq \delta_n(1 + |\varepsilon_{i,n}|)$ when $\delta_n < 1/2$. In addition, the last equation relies on the assumption $E|\varepsilon^-| < \infty$. Therefore,

$$\begin{aligned}
\frac{\widehat{\xi}_{\tau_n}^M(\varepsilon)}{\xi_{\tau_n}(\varepsilon)} - 1 &= \frac{\widehat{\xi}_{\tau_n}^M(\varepsilon)}{\widetilde{\xi}_{\tau_n}^M(\varepsilon)}\left(\frac{\widetilde{\xi}_{\tau_n}^M(\varepsilon)}{\xi_{\tau_n}(\varepsilon)} - 1\right) + \frac{\widehat{\xi}_{\tau_n}^M(\varepsilon)}{\widetilde{\xi}_{\tau_n}^M(\varepsilon)} - 1 \\
&= (1 + O_{\mathbb{P}}(\delta_n))\left(\frac{\widetilde{\xi}_{\tau_n}^M(\varepsilon)}{\xi_{\tau_n}(\varepsilon)} - 1\right) + O_{\mathbb{P}}(\delta_n) \\
&= \frac{\widetilde{\xi}_{\tau_n}^M(\varepsilon)}{\xi_{\tau_n}(\varepsilon)} - 1 + O_{\mathbb{P}}(\delta_n).
\end{aligned}$$

Finally, we use the asymptotic properties of $\widetilde{\xi}_{\tau_n}^M(\varepsilon)$ (see Theorem 4 in Daouia et al., 2019) to get

$$\sqrt{n(1-\tau_n)}\left(\frac{\widehat{\xi}_{\tau_n}^M(\varepsilon)}{\xi_{\tau_n}(\varepsilon)} - 1\right) \xrightarrow{d} \mathcal{N}(0, V(\gamma)),$$

with

$$V(\gamma) = \left(\frac{\gamma\sqrt{\log 2}}{\Gamma(1-\gamma)}\right)^2 \int_0^\infty\int_0^\infty (tu)^{-\gamma-1}e^{-t-u}(t \wedge u)\, dtdu. \quad \square$$

Proof of Theorem 3.2. As shown by Proposition 4 in Daouia et al. (2019), if $\tau \uparrow 1$,

$$\frac{\xi_\tau(\varepsilon)}{q_\tau(\varepsilon)} - g(\gamma) = b_1(\gamma,\rho)A\left((1-\tau)^{-1}\right) + b_2(\gamma)(1-\tau) + o\left(A\left((1-\tau)^{-1}\right)\right) + o(1-\tau).$$

Moreover, denote $r(\gamma)$ such that $\frac{\xi_\tau(\varepsilon)}{q_\tau(\varepsilon)} = g(\gamma)(1 + r(\gamma))$. Then we write

$$\begin{aligned}
\left(\frac{\widehat{\xi}_{\tau_n}^Q(\varepsilon)}{\xi_{\tau_n}(\varepsilon)} - 1\right) &= \left(\frac{g(\bar{\gamma})}{g(\gamma)} - 1\right) + \left(\frac{\widehat{q}_{\tau_n}(\varepsilon)}{q_{\tau_n}(\varepsilon)} - 1\right)\frac{g(\bar{\gamma})}{g(\gamma)} - \left(\frac{\xi_{\tau_n}(\varepsilon)}{g(\gamma)q_{\tau_n}(\varepsilon)} - 1\right)\left(\frac{g(\bar{\gamma})\widehat{q}_{\tau_n}(\varepsilon)}{g(\gamma)q_{\tau_n}(\varepsilon)} \cdot \frac{g(\gamma)q_{\tau_n}(\varepsilon)}{\xi_{\tau_n}(\varepsilon)}\right) \\
&= \left(\frac{g(\bar{\gamma})}{g(\gamma)} - 1\right) + \left(\frac{\widehat{q}_{\tau_n}(\varepsilon)}{q_{\tau_n}(\varepsilon)} - 1\right)(1 + o_{\mathbb{P}}(1)) - \left(\frac{\xi_{\tau_n}(\varepsilon)}{g(\gamma)q_{\tau_n}(\varepsilon)} - 1\right)(1 + o_{\mathbb{P}}(1)) \\
&= \left(\frac{g(\bar{\gamma})}{g(\gamma)} - 1\right) + \left(\frac{\widehat{q}_{\tau_n}(\varepsilon)}{q_{\tau_n}(\varepsilon)} - 1\right)(1 + o_{\mathbb{P}}(1)) - r(\tau_n)(1 + o_{\mathbb{P}}(1)).
\end{aligned}$$

By the delta-method, we get $\sqrt{n(1-\tau_n)}\left(\frac{g(\bar{\gamma})}{g(\gamma)} - 1\right) \xrightarrow{d} \eta(\gamma)\cdot\Psi$. Note that $\sqrt{n(1-\tau_n)}\left(\frac{\widehat{q}_{\tau_n}(\varepsilon)}{q_{\tau_n}(\varepsilon)} - 1\right) \xrightarrow{d} \Phi$ and $\sqrt{n(1-\tau_n)}r(\tau_n) \to \lambda_1\frac{b_1(\gamma,\rho)}{g(\gamma)} + \lambda_2\frac{b_2(\gamma)}{g(\gamma)}$. This completes the proof. $\square$

Proof of Theorem 3.3. To simplify the exposition, we define $d_n = \frac{1-\tau_n}{1-\tau_n'}$. Note that this is the key point:

$$\log\left(\frac{\bar{\xi}^{\star}_{\tau_n'}(\varepsilon)}{\xi_{\tau_n'}(\varepsilon)}\right) = (\bar{\gamma}-\gamma)\log(d_n) + \log\left(\frac{\bar{\xi}_{\tau_n}(\varepsilon)}{\xi_{\tau_n}(\varepsilon)}\right) - \log\left((d_n)^{-\gamma}\frac{\xi_{\tau_n'(\varepsilon)}}{\xi_{\tau_n}(\varepsilon)}\right).$$

For the first term,

$$\frac{\sqrt{n(1-\tau_n)}}{\log(d_n)}(\bar{\gamma}-\gamma)\log(d_n) = \sqrt{n(1-\tau_n)}(\bar{\gamma}-\gamma) \xrightarrow{d} \Psi \quad \text{as} \quad n\to\infty.$$

For the second term,

$$\frac{\sqrt{n(1-\tau_n)}}{\log(d_n)}\log\frac{\bar{\xi}_{\tau_n}(\varepsilon)}{\xi_{\tau_n}(\varepsilon)} = \mathrm{O}_{\mathbb{P}}(\frac{1}{\log(d_n)}) = \mathrm{o}_{\mathbb{P}}(1).$$

For the third term,

$$\begin{aligned}
&\frac{\sqrt{n(1-\tau_n)}}{\log(d_n)}\log\left((d_n)^{-\gamma}\frac{\xi_{\tau_n'(\varepsilon)}}{\xi_{\tau_n}(\varepsilon)}\right)\\
=&\frac{\sqrt{n(1-\tau_n)}}{\log(d_n)}\left(\log\frac{\xi_{\tau_n'(\varepsilon)}}{g(\gamma)q_{\tau_n'(\varepsilon)}} - \log\frac{\xi_{\tau_n(\varepsilon)}}{g(\gamma)q_{\tau_n(\varepsilon)}} + \log\left(\frac{q_{\tau_n'(\varepsilon)}}{(d_n)^{\gamma}q_{\tau_n(\varepsilon)}}\right)\right)\\
=&\frac{\sqrt{n(1-\tau_n)}}{\log(d_n)}\left(\mathrm{O}(1-\tau_n) + \mathrm{O}\left(\left|A\left(\frac{1}{1-\tau_n}\right)\right|\right) + \mathrm{O}\left(1-\tau_n'\right) + \mathrm{O}\left(\left|A\left(\frac{1}{1-\tau_n'}\right)\right|\right)\right.\\
&\left. + \mathrm{O}\left(\left|A\left(\frac{1}{1-\tau_n}\right)\right|\right)\right)\\
=&\frac{\sqrt{n(1-\tau_n)}}{\log(d_n)}\left(\mathrm{O}(1-\tau_n) + \mathrm{O}\left(\left|A\left(\frac{1}{1-\tau_n}\right)\right|\right)\right)\\
=&\mathrm{O}\left(\frac{1}{\log(d_n)}\right) = \mathrm{o}(1).
\end{aligned}$$

In detail, the second equation is due to Theorem B.2.18 in de Haan and Ferreira (2006) and Proposition 4 in Daouia et al. (2019). Finally, combine these three elements to conclude the result. □

Proof of Theorem 3.4. According to Lemma A.1, the condition (7) is implied by condition (14). Applying Theorems 3.1 and 3.3, we have

$$\sqrt{n(1-\tau_n)}\left(\frac{\widehat{\xi}^{M}_{\tau_n}(\varepsilon)}{\xi_{\tau_n}(\varepsilon)} - 1\right) \xrightarrow{d} \mathcal{N}(0, V(\gamma))$$

$$\text{and} \qquad \frac{\sqrt{n(1-\tau_n)}}{\log(d_n)}\left(\frac{\widehat{\xi}^{M,\star}_{\tau_n'}(\varepsilon)}{\xi_{\tau_n'}(\varepsilon)} - 1\right) \xrightarrow{d} \Psi \quad \text{as} \quad n\to\infty.$$

One can write

$$\begin{aligned}
\frac{\bar{\xi}_{\tau}(Y\mid\boldsymbol{x})}{\xi_{\tau}(Y\mid\boldsymbol{x})} - 1 =& \underbrace{\frac{\sigma(\boldsymbol{x})\xi_{\tau}(\varepsilon)}{m(\boldsymbol{x})+\sigma(\boldsymbol{x})\xi_{\tau}(\varepsilon)}\cdot\left(\frac{\bar{\xi}_{\tau}(\varepsilon)}{\xi_{\tau}(\varepsilon)} - 1\right)}_{B_{\tau,1}(\boldsymbol{x})} + \underbrace{\frac{\sigma(\boldsymbol{x})\xi_{\tau}(\varepsilon)}{m(\boldsymbol{x})+\sigma(\boldsymbol{x})\xi_{\tau}(\varepsilon)}\cdot\frac{\bar{\xi}_{\tau}(\varepsilon)}{\xi_{\tau}(\varepsilon)}\cdot\left(\frac{\bar{\sigma}(\boldsymbol{x})}{\sigma(\boldsymbol{x})} - 1\right)}_{B_{\tau,2}(\boldsymbol{x})}\\
&+ \underbrace{\frac{1}{m(\boldsymbol{x})+\sigma(x)\xi_{\tau}(\varepsilon)}\cdot(\bar{m}(\boldsymbol{x}) - m(\boldsymbol{x}))}_{B_{\tau,3}(\boldsymbol{x})}.
\end{aligned} \tag{20}$$

Let's consider the intermediate level $\tau = \tau_n$ for the consistent estimator $\bar{\xi}_{\tau_n}(Y\mid\boldsymbol{x}) = \widehat{\xi}^{M}_{\tau_n}(Y\mid x)$. For the first term, we have

$$\sqrt{n(1-\tau_n)}B_{\tau_n,1}(\boldsymbol{x}) = \frac{\sigma(\boldsymbol{x})\xi_{\tau_n}(\varepsilon)}{m(\boldsymbol{x})+\sigma(\boldsymbol{x})\xi_{\tau_n}(\varepsilon)}\cdot\sqrt{n(1-\tau_n)}\left(\frac{\widehat{\xi}^{M}_{\tau_n}(\varepsilon)}{\xi_{\tau_n}(\varepsilon)} - 1\right) \xrightarrow{d} \mathcal{N}(0, V(\gamma)), \tag{21}$$

applying Theorem 3.1 and the convergence $\xi_{\tau_n}(\varepsilon)\to\infty$ as $\tau_n\to 1$. For the latter two terms, we use the condition (15) to get

$$\sqrt{n(1-\tau_n)}B_{\tau_n,2}(\boldsymbol{x}) = \mathrm{o}_{\mathbb{P}}\left(\frac{\sigma(\boldsymbol{x})\xi_{\tau_n}(\varepsilon)}{m(\boldsymbol{x})+\sigma(\boldsymbol{x})\xi_{\tau_n}(\varepsilon)}\cdot\frac{\widehat{\xi}^{M}_{\tau_n}(\varepsilon)}{\xi_{\tau_n}(\varepsilon)}\right) = \mathrm{o}_{\mathbb{P}}(1) \tag{22}$$

$$\text{and} \quad \sqrt{n(1-\tau_n)}B_{\tau_n,3}(\boldsymbol{x}) = \mathrm{O}_{\mathbb{P}}\left(\frac{1}{m(\boldsymbol{x})+\sigma(\boldsymbol{x})\xi_{\tau_n}(\varepsilon)}\right) = \mathrm{o}_{\mathbb{P}}(1). \tag{23}$$

Therefore, we combine equations (21), (22) and (23) to obtain

$$\sqrt{n(1-\tau_n)}\left(\frac{\widehat{\xi}^M_{\tau_n}(Y\mid x)}{\xi_{\tau_n}(Y\mid x)}-1\right)\xrightarrow{d}\mathcal{N}(0,V(\gamma)).$$

Then, we focus on the extreme level $\tau=\tau_n'$. For the conditional extreme estimator $\widehat{\xi}^{M,\star}_{\tau_n'}(Y\mid \boldsymbol{x})$, it can be easily found that

$$\frac{\sqrt{n(1-\tau_n)}}{\log(d_n)}B_{\tau_n',1}(\boldsymbol{x})=\frac{\sigma(\boldsymbol{x})\xi_{\tau_n'}(\varepsilon)}{m(\boldsymbol{x})+\sigma(\boldsymbol{x})\xi_{\tau_n'}(\varepsilon)}\cdot\frac{\sqrt{n(1-\tau_n)}}{\log(d_n)}\left(\frac{\widehat{\xi}^{M,\star}_{\tau_n'}(\varepsilon)}{\xi_{\tau_n'}(\varepsilon)}-1\right)\xrightarrow{d}\Psi \tag{24}$$

by Theorem 3.3. Both $B_{\tau_n',2}(\boldsymbol{x})$ and $B_{\tau_n',3}(\boldsymbol{x})$ are controlled, that is

$$\begin{aligned}\frac{\sqrt{n(1-\tau_n)}}{\log(d_n)}B_{\tau_n',2}(\boldsymbol{x})&=o_{\mathbb{P}}\left(\frac{1}{\log(d_n)}\cdot\frac{\sigma(\boldsymbol{x})\xi_{\tau_n'}(\varepsilon)}{m(\boldsymbol{x})+\sigma(\boldsymbol{x})\xi_{\tau_n'}(\varepsilon)}\cdot\frac{\widehat{\xi}^{M,\star}_{\tau_n'}(\varepsilon)}{\xi_{\tau_n'}(\varepsilon)}\right)\\&=o_{\mathbb{P}}\left(\frac{1}{\log(d_n)}\right),\end{aligned} \tag{25}$$

$$\begin{aligned}\text{and}\quad\frac{\sqrt{n(1-\tau_n)}}{\log\left[(1-\tau_n)/(1-\tau_n')\right]}B_{\tau_n',3}(\boldsymbol{x})&=O_{\mathbb{P}}\left(\frac{1}{\log(d_n)}\cdot\frac{1}{m(\boldsymbol{x})+\sigma(\boldsymbol{x})\xi_{\tau_n'}(\varepsilon)}\right)\\&=o_{\mathbb{P}}\left(\frac{1}{\log(d_n)}\right).\end{aligned} \tag{26}$$

Combining equations (24), (25) and (26) yields

$$\frac{\sqrt{n(1-\tau_n)}}{\log(d_n)}\left(\frac{\widehat{\xi}^{M,\star}_{\tau_n'}(Y\mid \boldsymbol{x})}{\xi_{\tau_n'}(Y\mid \boldsymbol{x})}-1\right)\xrightarrow{d}\Psi,$$

which completes the proof. □

Proof of Corollary 3.1. The proof is almost entirely similar to that of Theorem 3.3 for estimation of the direct conditional estimator. Following the same idea as Eq. (20) and using Theorem 3.2, for the indirect intermediate estimator $\widehat{\xi}^Q_{\tau_n}(Y\mid \boldsymbol{x})$, we have

$$\begin{aligned}&\sqrt{n(1-\tau_n)}B_{\tau_n,1}(\boldsymbol{x})\xrightarrow{d}\eta(\gamma)\cdot\Psi+\Phi-\left[\lambda_1\frac{b_1(\gamma,\rho)}{g(\gamma)}+\lambda_2\frac{b_2(\gamma)}{g(\gamma)}\right],\\&\sqrt{n(1-\tau_n)}B_{\tau_n,2}(\boldsymbol{x})=o_{\mathbb{P}}(1),\ \text{and}\ \sqrt{n(1-\tau_n)}B_{\tau_n,3}(\boldsymbol{x})=o_{\mathbb{P}}(1),\end{aligned}$$

to get

$$\sqrt{n(1-\tau_n)}\left(\frac{\widehat{\xi}^Q_{\tau_n}(Y\mid \boldsymbol{x})}{\xi_{\tau_n}(Y\mid \boldsymbol{x})}-1\right)\xrightarrow{d}\eta(\gamma)\cdot\Psi+\Phi-\left[\lambda_1\frac{b_1(\gamma,\rho)}{g(\gamma)}+\lambda_2\frac{b_2(\gamma)}{g(\gamma)}\right].$$

Focusing on the extreme estimator based on quantile, one can write

$$\begin{aligned}\left(\frac{\widehat{\xi}^{Q,\star}_{\tau_n'}(\varepsilon)}{\xi_{\tau_n'}(\varepsilon)}-1\right)=&g(\bar{\gamma})\frac{q_{\tau_n'}(\varepsilon)}{\xi_{\tau_n'}(\varepsilon)}\left(\frac{\widehat{q}^{\star}_{\tau_n'}(\varepsilon)}{q_{\tau_n'}(\varepsilon)}-1\right)+\frac{q_{\tau_n'}(\varepsilon)}{\xi_{\tau_n'}(\varepsilon)}(g(\bar{\gamma})-g(\gamma))\\&+\frac{q_{\tau_n'}(\varepsilon)}{\xi_{\tau_n'}(\varepsilon)}\left(g(\gamma)-\frac{\xi_{\tau_n'}(\varepsilon)}{q_{\tau_n'}(\varepsilon)}\right),\end{aligned} \tag{27}$$

where $\widehat{q}^{\star}_{\tau_n'}(\varepsilon)=\left(\frac{1-\tau_n'}{1-\tau_n}\right)^{-\bar{\gamma}}\widehat{q}_{\tau_n}(\varepsilon)=d_n^{\bar{\gamma}}\widehat{\varepsilon}_{n-k,n}$. For the first term in Eq. (27), we have

$$\begin{aligned}\frac{\sqrt{n(1-\tau_n)}}{\log d_n}\left(\frac{\widehat{q}^{\star}_{\tau_n'}(\varepsilon)}{q_{\tau_n'}(\varepsilon)}-1\right)=&\frac{d_n^{\gamma}q_{\tau_n}(\varepsilon)}{q_{\tau_n'}(\varepsilon)}\Bigg[\sqrt{n(1-\tau_n)}\left(\frac{\widehat{q}_{\tau_n}(\varepsilon)}{q_{\tau_n}(\varepsilon)}-1\right)\frac{d_n^{\bar{\gamma}-\gamma}}{\log d_n}+\frac{\sqrt{n(1-\tau_n)}}{\log d_n}\left(d_n^{\bar{\gamma}-\gamma}-1\right)\\&-\frac{\sqrt{n(1-\tau_n)}A\left(\frac{1}{1-\tau_n}\right)}{\log d_n}\cdot\frac{q_{\tau_n'}(\varepsilon)/\left(d_n^{\gamma}q_{\tau_n}(\varepsilon)\right)-1}{A\left(\frac{1}{1-\tau_n}\right)}\Bigg].\end{aligned}$$

Note that $\sqrt{n(1-\tau_n)}\left(\frac{\widehat{q}_{\tau_n}(\varepsilon)}{q_{\tau_n}(\varepsilon)}-1\right)=O_{\mathbb{P}}(1)$ and $\frac{\sqrt{n(1-\tau_n)}}{\log d_n}\left(d_n^{\bar{\gamma}-\gamma}-1\right)\xrightarrow{d}\Psi$. It yields

$$\frac{\sqrt{n(1-\tau_n)}}{\log d_n}\left(\frac{\widehat{q}^{\star}_{\tau_n'}(\varepsilon)}{q_{\tau_n'}(\varepsilon)}-1\right)\xrightarrow{d}\Psi,$$

which is similar to the proof of Theorem 4.3.8 in de Haan and Ferreira (2006). For the second term, using delta-method, we have

$$\sqrt{n(1-\tau_n)}(g(\bar{\gamma})-g(\gamma)) = O_{\mathbb{P}}(1).$$

For the third term, we apply the assumptions for the sequences τ_n such that $\sqrt{n(1-\tau_n)}A\left((1-\tau_n)^{-1}\right) \to \lambda_1 \in \mathbb{R}$ and $(1-\tau_n)\sqrt{n(1-\tau_n)} \to \lambda_2 \in \mathbb{R}$, to obtain

$$\begin{aligned}\sqrt{n(1-\tau_n)}\left(g(\gamma) - \frac{\xi_{\tau_n'}(\varepsilon)}{q_{\tau_n'}(\varepsilon)}\right) &= \sqrt{n(1-\tau_n)}\left[O\left(A\left(\frac{1}{1-\tau_n'}\right)\right) + O\left(1-\tau_n'\right)\right] \\ &= \sqrt{n(1-\tau_n)}\left[O\left(A\left(\frac{1}{1-\tau_n}\right)\right) + O\left(1-\tau_n\right)\right] = O(1),\end{aligned}$$

controlling three terms results in

$$\frac{\sqrt{n(1-\tau_n)}}{\log d_n}\left(\frac{\widehat{\xi}^{Q,\star}_{\tau_n'}(\varepsilon)}{\xi_{\tau_n'}(\varepsilon)} - 1\right) \xrightarrow{d} \Psi.$$

Returning to the conditional indirect estimator $\widehat{\xi}^{Q,\star}_{\tau_n'}(\varepsilon)$, similar to $\widehat{\xi}^{M,\star}_{\tau_n'}(\varepsilon)$, we have

$$\frac{\sqrt{n(1-\tau_n)}}{\log(d_n)}B_{\tau_n',1}(\boldsymbol{x}) \xrightarrow{d} \Psi \quad \text{and} \quad \frac{\sqrt{n(1-\tau_n)}}{\log(d_n)}B_{\tau_n',2}(\boldsymbol{x}) = o_{\mathbb{P}}\left(\frac{1}{\log(d_n)}\right). \tag{28}$$

It follows by equations (26) and (28) that

$$\frac{\sqrt{n(1-\tau_n)}}{\log(d_n)}\left(\frac{\widehat{\xi}^{Q,\star}_{\tau_n'}(Y \mid \boldsymbol{x})}{\xi_{\tau_n'}(Y \mid \boldsymbol{x})} - 1\right) \xrightarrow{d} \Psi. \quad \square$$

Corollary B.1. *Suppose that condition $\mathcal{C}_2(\gamma, \rho, A)$ holds with $\sqrt{n(1-\tau_n)}A\left((1-\tau_n)^{-1}\right) \to \lambda \in \mathbb{R}$, $A(t) = b\gamma t^{\rho}$ and $\rho < 0$. Also assume that the residuals $\widehat{\varepsilon}_i$, $1 \le i \le n$, satisfy (7). Let consistent estimators $\bar{\rho}$ of ρ and $\bar{b}$ of b be such that $(\bar{\rho}-\rho)\log(n) = o_{\mathbb{P}}(1)$. Then*

$$\sqrt{k}\left(\widehat{\gamma}_k^{H,RB} - \gamma\right) \xrightarrow{d} \mathcal{N}\left(0, \gamma^2\right).$$

Proof of Corollary B.1. Denote $\bar{A}(t) = \bar{\gamma}\bar{b}t^{\bar{\rho}}$ and the Hill estimator based on unseen error $\widetilde{\gamma}_k^H = \frac{1}{k}\sum_{i=1}^k \log\frac{\varepsilon_{n-i+1,n}}{\varepsilon_{n-k,n}}$. One can write that

$$\sqrt{k}\left(\widehat{\gamma}_k^{H,RB} - \gamma\right) = \underbrace{\sqrt{k}\left(\widehat{\gamma}_k^H - \widetilde{\gamma}_k^H\right)}_{I_{n,1}} + \underbrace{\sqrt{k}\left(\widetilde{\gamma}_k^H - \gamma\right) - \frac{\lambda}{1-\rho}}_{I_{n,2}} - \underbrace{\left(\frac{\sqrt{k}\bar{A}(\frac{n}{k})}{1-\bar{\rho}} - \frac{\lambda}{1-\rho}\right)}_{I_{n,3}}$$

For the first term $I_{n,1}$, using Lemma A.2,

$$I_{n,1} = \sqrt{k}\left(\widehat{\gamma}_k^H - \widetilde{\gamma}_k^H\right) = \sqrt{k}\left(\frac{1}{k}\sum_{i=1}^k \log\frac{\widehat{\varepsilon}_{n-i+1,n}}{\varepsilon_{n-i+1,n}} - \log\frac{\widehat{\varepsilon}_{n-k,n}}{\varepsilon_{n-k,n}}\right) = O_{\mathbb{P}}(\sqrt{k}\delta_n) = o_{\mathbb{P}}(1).$$

Note that $I_{n,2} \xrightarrow{d} \mathcal{N}\left(0, \gamma^2\right)$. The consistency of $\bar{\rho}$ implies $(\rho - \bar{\rho})\log(1-\tau_n) = o_{\mathbb{P}}(1)$. Using Slutsky's theorem, we have $\bar{A}(\frac{n}{k})/A(\frac{n}{k}) - 1 = o_{\mathbb{P}}(1)$. Then apply $\sqrt{n(1-\tau_n)}A\left((1-\tau_n)^{-1}\right) \to \lambda \in \mathbb{R}$ to get

$$I_{n,3} = \frac{\bar{A}(\frac{n}{k})}{A(\frac{n}{k})} \cdot \frac{1-\bar{\rho}}{1-\rho} \cdot \frac{\sqrt{k}A(\frac{n}{k})}{1-\rho} - \frac{\lambda}{1-\rho} = (1+o_{\mathbb{P}}(1))\left(\frac{\lambda}{1-\rho} + o(1)\right) - \frac{\lambda}{1-\rho} = o_{\mathbb{P}}(1).$$

As a consequence,

$$\sqrt{k}\left(\widehat{\gamma}_k^{H,RB} - \gamma\right) \xrightarrow{d} \mathcal{N}\left(0, \gamma^2\right). \quad \square$$

Proof of Theorem 4.1. The key is to write that

$$\begin{aligned}\log\left(\frac{\widehat{\xi}^{M,\star,RB}_{\tau_n'}(\varepsilon)}{\xi_{\tau_n'}(\varepsilon)}\right) =& (\bar{\gamma}-\gamma)\log\left(\frac{1-\tau_n}{1-\tau_n'}\right) + \log\left(\frac{\widehat{\xi}^{M}_{\tau_n}(\varepsilon)}{\xi_{\tau_n}(\varepsilon)}\right) + \log\left(\frac{1+\bar{\mathcal{R}}_{n,1}}{1+\mathcal{R}_{n,1}}\right) \\ &+ \log\left(\frac{1+\bar{\mathcal{R}}_{n,2}}{1+\mathcal{R}_{n,2}}\right) + \log\left(\frac{1+\bar{\mathcal{R}}_{n,3}}{1+\mathcal{R}_{n,3}}\right)\end{aligned} \tag{29}$$

$$\text{and} \quad \log\left(\frac{\widehat{\xi}^{Q,\star,\mathrm{RB}}_{\tau_n'}(\varepsilon)}{\xi_{\tau_n'}(\varepsilon)}\right) = (\bar{\gamma}-\gamma)\log\left(\frac{1-\tau_n}{1-\tau_n'}\right) + \log\left(\frac{g(\bar{\gamma})}{g(\gamma)}\right) + \log\left(\frac{\widehat{q}_{\tau_n}(\varepsilon)}{q_{\tau_n}(\varepsilon)}\right) + \log\left(\frac{1+\bar{\mathcal{R}}_{n,2}}{1+\mathcal{R}_{n,2}}\right) + \log\left(\frac{1+\bar{\mathcal{R}}_{n,3}}{1+\mathcal{R}_{n,3}}\right). \tag{30}$$

By the assumptions on $\bar{\gamma}, \bar{b}$ and $\bar{\rho}$, we have

$$\log\left(\frac{1+\bar{\mathcal{R}}_{n,1}}{1+\mathcal{R}_{n,1}}\right) + \log\left(\frac{1+\bar{\mathcal{R}}_{n,2}}{1+\mathcal{R}_{n,2}}\right) + \log\left(\frac{1+\bar{\mathcal{R}}_{n,3}}{1+\mathcal{R}_{n,3}}\right) = o_{\mathbb{P}}\left(\frac{\log d_n}{\sqrt{n(1-\tau_n)}}\right). \tag{31}$$

The details are as follows. First, for $\bar{\mathcal{R}}_{n,1}$ and $\mathcal{R}_{n,1}$,

$$\begin{aligned}\frac{1}{1+\bar{\mathcal{R}}_{n,1}} - \frac{1}{1+\mathcal{R}_{n,1}} &= \frac{C_1(\bar{\gamma},\bar{\rho})\cdot\bar{\gamma}\bar{b}(1-\tau_n)^{-\bar{\rho}}}{g(\bar{\gamma})} + \frac{C_2(\bar{\gamma})(1-\tau_n)}{g(\bar{\gamma})} - \frac{C_1(\gamma,\rho)\cdot\gamma b(1-\tau_n)^{-\rho}}{g(\gamma)} - \frac{C_2(\gamma)(1-\tau_n)}{g(\gamma)} + o(1)\\ &= \left(\frac{C_1(\bar{\gamma},\bar{\rho})\cdot\bar{\gamma}\bar{b}}{g(\bar{\gamma})} - \frac{C_1(\gamma,\rho)\cdot\gamma b}{g(\gamma)}\right)(1-\tau_n)^{-\bar{\rho}} + \left((1-\tau_n)^{-\bar{\rho}} - (1-\tau_n)^{-\rho}\right)\frac{C_1(\gamma,\rho)\cdot\gamma b}{g(\gamma)}\\ &\quad + \left(\frac{C_2(\bar{\gamma})}{g(\bar{\gamma})} - \frac{C_2(\gamma)}{g(\gamma)}\right)(1-\tau_n) + o(1).\end{aligned}$$

Applying Slutsky's theorem and the continuous mapping theorem, we can get $\frac{C_1(\bar{\gamma},\bar{\rho})\cdot\bar{\gamma}\bar{b}}{g(\bar{\gamma})} - \frac{C_1(\gamma,\rho)\cdot\gamma b}{g(\gamma)} = o_{\mathbb{P}}(1)$. It yields $\sqrt{n(1-\tau_n)}\left(\frac{C_2(\bar{\gamma})}{g(\bar{\gamma})} - \frac{C_2(\gamma)}{g(\gamma)}\right) = O_{\mathbb{P}}(1)$ by the delta method. According to $(\bar{\rho}-\rho)\log(n) = o_{\mathbb{P}}(1)$, we find

$$(\rho-\bar{\rho})\log(1-\tau_n) = o_{\mathbb{P}}(1) \Rightarrow (1-\tau_n)^{-\bar{\rho}} = O_{\mathbb{P}}\left((1-\tau_n)^{-\rho}\right).$$

Using the assumption on the function $A(\cdot)$,

$$\frac{1}{1+\bar{\mathcal{R}}_{n,1}} - \frac{1}{1+\mathcal{R}_{n,1}} = O_{\mathbb{P}}\left((1-\tau_n)^{-\rho}\right) + O_{\mathbb{P}}\left(\sqrt{\frac{1-\tau_n}{n}}\right) = O_{\mathbb{P}}\left(\frac{1}{\sqrt{n(1-\tau_n)}}\right).$$

Then

$$\log\left(\frac{1+\bar{\mathcal{R}}_{n,1}}{1+\mathcal{R}_{n,1}}\right) = \log\left(\frac{1}{1+\mathcal{R}_{n,1}}\right) - \log\left(\frac{1}{1+\bar{\mathcal{R}}_{n,1}}\right) = O_{\mathbb{P}}\left(\frac{1}{\sqrt{n(1-\tau_n)}}\right).$$

The control of the second term uses the analogous technique

$$\log\left(\frac{1+\bar{\mathcal{R}}_{n,2}}{1+\mathcal{R}_{n,2}}\right) = o_{\mathbb{P}}\left(\frac{1}{\sqrt{n(1-\tau_n)}}\right).$$

Next, for $\bar{\mathcal{R}}_{n,3}$ and $\mathcal{R}_{n,3}$,

$$\begin{aligned}\bar{\mathcal{R}}_{n,3} - \mathcal{R}_{n,3} &= \frac{\bar{\gamma}\bar{b}}{\bar{\rho}}\left[(1-\tau_n')^{-\bar{\rho}} - (1-\tau_n)^{-\bar{\rho}}\right] - \frac{\gamma b}{\rho}\left[(1-\tau_n')^{-\rho} - (1-\tau_n)^{-\rho}\right] + o(1)\\ &= \underbrace{\frac{\gamma b}{\rho}(1-\tau_n)^{-\rho} - \frac{\bar{\gamma}\bar{b}}{\bar{\rho}}(1-\tau_n)^{-\bar{\rho}}}_{I_{n,4}} + \underbrace{\frac{\bar{\gamma}\bar{b}}{\bar{\rho}}(1-\tau_n')^{-\bar{\rho}} - \frac{\gamma b}{\rho}(1-\tau_n')^{-\rho}}_{I_{n,5}}.\end{aligned}$$

Focus on $I_{n,4}$,

$$I_{n,4} = \left[1 - \left(\frac{\bar{\gamma}\bar{b}}{\bar{\rho}}/\frac{\gamma b}{\rho}\right)(1-\tau_n)^{\rho-\bar{\rho}}\right]\cdot\frac{\gamma b}{\rho}(1-\tau_n)^{-\rho} = O_{\mathbb{P}}(1)\cdot(1-\tau_n)^{-\rho} = O_{\mathbb{P}}\left(\frac{1}{\sqrt{n(1-\tau_n)}}\right),$$

which is also available for $I_{n,5}$. It results in

$$\log\left(\frac{1+\bar{\mathcal{R}}_{n,3}}{1+\mathcal{R}_{n,3}}\right) = O_{\mathbb{P}}\left(\frac{1}{\sqrt{n(1-\tau_n)}}\right).$$

Combining the three terms, Eq. (31) is obtained. The result (i) follows by using Theorem 3.1 and Eq. (29). To show (ii), we should recall that $\sqrt{n(1-\tau_n)}\left(\frac{\widehat{q}_{\tau_n}(\varepsilon)}{q_{\tau_n}(\varepsilon)} - 1\right) \xrightarrow{d} \mathcal{N}\left(0,\gamma^2\right)$ (see Corollary 2.1 in Girard et al., 2021). It is then clear from Eq. (30) that

$$\frac{\sqrt{n(1-\tau_n)}}{\log(d_n)}\left(\frac{\widehat{\xi}^{Q,\star,\mathrm{RB}}_{\tau_n'}(\varepsilon)}{\xi_{\tau_n'}(\varepsilon)} - 1\right) \xrightarrow{d} \Psi. \quad \square$$

(a) Asymptotic Gaussian confidence intervals. (b) Semiparametric Pareto tail bootstrap confidence intervals

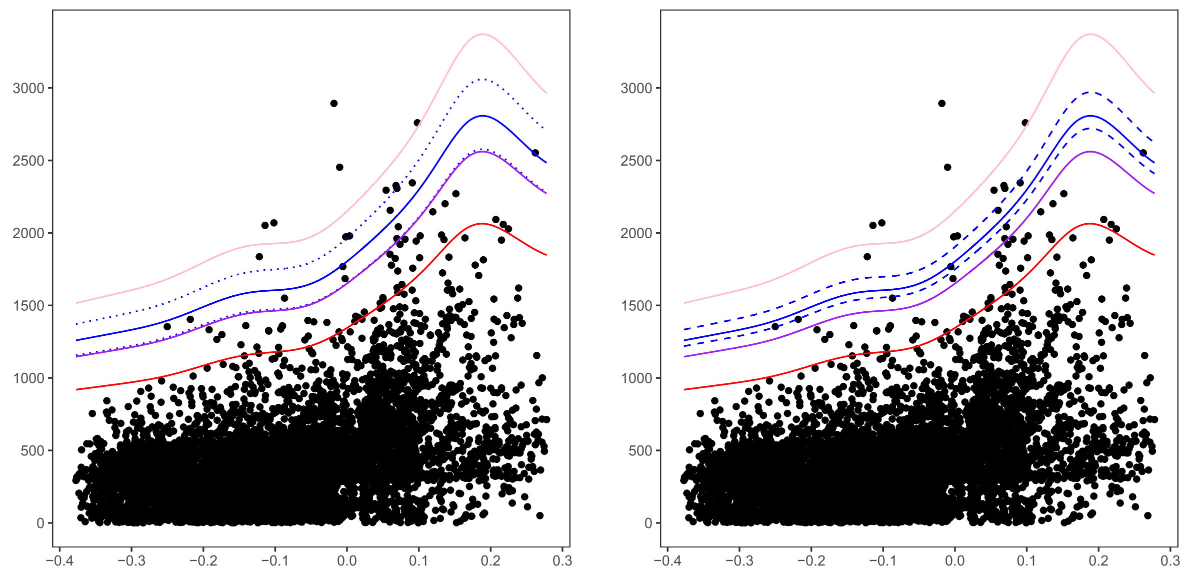

Fig. 4. Vehicle Insurance Data: Pointwise 95% confidence intervals. (a): Asymptotic Gaussian confidence intervals (blue dotted line). (b): Semiparametric Pareto tail bootstrap confidence intervals (blue dashed line). Both panels display the estimates of the conditional extremile in the direct method (blue solid line), ES (pink line), quantile (purple line) and expectile (red line) at level $\tau_n' = 0.9977$ on the $\widehat{\boldsymbol{\beta}}^\top \boldsymbol{x} - y$ (scatter) plot.

Appendix C. Supplement to the empirical procedure

Asymptotic confidence interval: If $\bar{\gamma}$ is the estimator $\widehat{\gamma}^{\mathrm{H,RB}}_{\lfloor n(1-\tau_n)\rfloor}$, Theorem 3.4 and Corollary 3.1 show,

$$\frac{\sqrt{n\,(1-\tau_n)}}{\log\left[(1-\tau_n)\,/\,\left(1-\tau_n'\right)\right]}\left(\frac{\bar{\xi}^{\star}_{\tau_n'}(Y\mid \boldsymbol{x})}{\xi_{\tau_n'}(Y\mid \boldsymbol{x})}-1\right)\xrightarrow{d}\Psi,$$

where Ψ is indeed $\sqrt{n(1-\tau_n)}$-asymptotically Gaussian with mean 0 and variance γ^2 and $\bar{\xi}^{\star}_{\tau_n'}(Y\mid \boldsymbol{x})$ is the consistent estimator of $\xi_{\tau_n'}(Y\mid \boldsymbol{x})$, e.g., $\widehat{\xi}^{M,\star}_{\tau_n'}(Y\mid \boldsymbol{x})$, $\widehat{\xi}^{Q,\star}_{\tau_n'}(Y\mid \boldsymbol{x})$ and their bias-reduced versions. We can naturally construct an asymptotic pointwise 95% confidence interval (Gaussian) for $\xi_{\tau_n'}(Y\mid \boldsymbol{x})$, that is

$$\widehat{I}^{G}_{\tau_n'}(\boldsymbol{x})=\left[\bar{\xi}^{\star}_{\tau_n'}(Y\mid \boldsymbol{x})\exp\left(\pm 1.96\frac{\log\left[(1-\tau_n)\,/\,\left(1-\tau_n'\right)\right]}{\sqrt{n\,(1-\tau_n)}}\bar{\gamma}\right)\right].$$

Simultaneously, we use the semiparametric Pareto tail bootstrap method introduced by Girard et al. (2021) to obtain the 95% bootstrap confidence interval $\widehat{I}^{B}_{\tau_n'}(\boldsymbol{x})$. The results are shown in Fig. 4.

References

Ahmad, A.A., Diop, A., Girard, S., Usseglio-Carleve, A., 2020. Estimation of extreme quantiles from heavy-tailed distributions in a location-dispersion regression model. Electronic Journal of Statistics 14, 4421–4456.
Beirlant, J., Goegebeur, Y., Segers, J., Teugels, J., 2004. Statistics of Extremes: Theory and Applications. Wiley, New York.
Bellini, F., Klar, B., Müller, A., Gianin, E.R., 2014. Generalized quantiles as risk measures. Insurance. Mathematics & Economics 54, 41–48.
Bousebata, Meryem, Enjolras, Geoffroy, Girard, Stéphane, 2023. Extreme partial least-squares. Journal of Multivariate Analysis 194, 105101.
Caeiro, F., Gomes, M.I., Pestana, D., 2005. Direct reduction of bias of the classical Hill estimator. REVSTAT Statistical Journal 3, 113–136.
Daouia, A., Gijbels, I., Stupfler, G., 2019. Extremiles: a new perspective on asymmetric least squares. Journal of the American Statistical Association 114, 1366–1381.
Daouia, A., Gijbels, I., Stupfler, G., 2022a. Extremile regression. Journal of the American Statistical Association 117, 1579–1586.
Daouia, A., Girard, S., 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., Stupfler, G., 2020. Tail expectile process and risk assessment. Bernoulli 26, 531–556.
Daouia, A., Stupfler, G., Usseglio-Carleve, A., 2022b. Inference for extremal regression with dependent heavy-tailed data. Available at https://hal.archives-ouvertes.fr/hal-03612202.
Davison, A.C., Padoan, S.A., Stupfler, G., 2022. Tail risk inference via expectiles in heavy-tailed time series. Journal of Business & Economic Statistics. Online. Available at https://doi.org/10.1080/07350015.2022.2078332.
de Haan, L., Ferreira, A., 2006. Extreme Value Theory: an Introduction. Springer.
Einmahl, J.H., de Haan, L., Zhou, C., 2016. Statistics of heteroscedastic extremes. Journal of the Royal Statistical Society, Series B, Methodological 78, 31–51.
Embrechts, P., Klüppelberg, C., Mikosch, T., 1997. Modelling Extremal Events for Insurance and Finance. Springer, Berlin-Heidelberg.
Emmer, S., Kratz, M., Tasche, D., 2015. What is the best risk measure in practice? A comparison of standard measures. The Journal of Risk 18 (2), 31–60.
Girard, S., Stupfler, G., Usseglio-Carleve, A., 2021. Extreme conditional expectile estimation in heavy-tailed heteroscedastic regression models. The Annals of Statistics 49, 3358–3382.
Girard, S., Stupfler, G., Usseglio-Carleve, A., 2022a. Nonparametric extreme conditional expectile estimation. Scandinavian Journal of Statistics 49, 78–115.
Girard, S., Stupfler, G., Usseglio-Carleve, A., 2022b. On automatic bias reduction for extreme expectile estimation. Statistics and Computing 32, 64.
Gomes, M.I., Brilhante, M.F., Caeiro, F., Pestana, D., 2015. A new partially reduced-bias mean-of-order p class of extreme value index estimators. Computational Statistics & Data Analysis 82, 223–237.

Gomes, M.I., Brilhante, M.F., Pestana, D., 2016. New reduced-bias estimators of a positive extreme value index. Communications in Statistics. Simulation and Computation 45, 833–862.
Hill, B.M., 1975. A simple general approach to inference about the tail of a distribution. The Annals of Statistics, 1163–1174.
Kazama, S., Sato, A., Kawagoe, S., 2009. Evaluating the cost of flood damage based on changes in extreme rainfall in Japan. Sustainability Science 4, 61–69.
Koenker, R., Bassett, G.S., 1978. Regression quantiles. Econometrica 46, 33–50.
Mao, T., Ng, K.W., Hu, T., 2015. Asymptotic expansions of generalized quantiles and expectiles for extreme risks. Probability in the Engineering and Informational Sciences 29, 309–327.
Mao, T., Yang, F., 2015. Risk concentration based on expectiles for extreme risks under FGM copula. Insurance. Mathematics & Economics 64, 429–439.
Newey, W.K., Powell, J.L., 1987. Asymmetric least squares estimation and testing. Econometrica 55, 819–847.
Spady, R.H., Stouli, S., 2018a. Dual regression. Biometrika 105, 1–18.
Spady, R., Stouli, S., 2018b. Simultaneous mean-variance regression. arXiv:1804.01631.
Stupfler, G., 2020. Contributions to statistical aspects of extreme value theory and risk assessment. PhD Thesis. Available at https://tel.archives-ouvertes.fr/tel-03142415.
Weissman, I., 1978. Estimation of parameters and large quantiles based on the k largest observations. Journal of the American Statistical Association 73, 812–815.
Wen, Z., Yin, W., 2013. A feasible method for optimization with orthogonality constraints. Mathematical Programming 142, 397–434.
Zhu, L., Dong, Y., Li, R., 2013. Semiparametric estimation of conditional heteroscedasticity via single-index modeling. Statistica Sinica 23, 1215.
Zhu, R., Zhang, J., Zhao, R., Xu, P., Zhou, W., Zhang, X., 2018. orthoDr: semiparametric dimension reduction via orthogonality constrained optimization. Available at https://arxiv.org/abs/1811.11733.