Statistica Sinica Preprint No: SS-2025-0187

Title	Inference for A Two-Step Joint Model of Extreme Quantile and Expected Shortfall Regression
Manuscript ID	SS-2025-0187
URL	http://www.stat.sinica.edu.tw/statistica/
DOI	10.5705/ss.202025.0187
Complete List of Authors	Qingzhao Zhong, Jingyu Ji, Liujun Chen, Yanxi Hou and Deyuan Li
Corresponding Authors	Jingyu Ji
E-mails	jingyuji@cueb.edu.cn

Statistica Sinica: Preprint
doi:10.5705/ss.202025.0187

INFERENCE FOR A TWO-STEP JOINT MODEL OF EXTREME QUANTILE AND EXPECTED SHORTFALL REGRESSION

Qingzhao Zhong[1], Jingyu Ji[2], Liujun Chen[3], Yanxi Hou[1], and Deyuan Li[4]

[1]*School of Data Science, Fudan University*

[2]*School of Statistics and Data Science, Capital University of Economics and Business*

[3]*School of Management, University of Science and Technology of China*

[4]*School of Management, Fudan University*

Abstract: As a coherent risk measure, Expected Shortfall (ES) has garnered increasing attention due to its merits in quantitative risk management, particularly its ability to capture tail risks. Consequently, the Expected Shortfall regression model has recently been proposed in conjunction with quantile regression to investigate the conditional effect of predictors on a response variable of interest. However, existing approaches have encountered challenges in effectively estimating the conditional expected shortfall regression at extreme levels, primarily due to the scarcity of observations in the tails. To address this issue, this paper first fits a joint regression model of conditional quantile and conditional ES at

Corresponding authors: Jingyu Ji, E-mail: jingyuji@cueb.edu.cn. Liujun Chen, E-mail: ljchen22@ustc.edu.cn.

an intermediate level using a two-step procedure. Subsequently, three extrapolative approaches are proposed to study the extreme conditional ES estimation. We also develop the asymptotic properties of all proposed estimators within a conditional heteroscedastic extreme framework. Furthermore, simulations are conducted to examine the finite sample performance of our methods. Finally, a real-world example underscores the practical advantages of extreme conditional ES regression.

Key words and phrases: quantile regression; expected shortfall regression; heteroscedastic extremes; tail risk.

1. Introduction

Value-at-Risk (VaR) and Expected Shortfall (ES) are two popular measures of quantitative risk management that have gained widespread adoption. VaR is favored by practitioners for its simplicity and interpretability; it represents a specific quantile of a loss distribution, making it accessible for practical use. Despite its robustness in statistical terms, VaR suffers from a significant limitation: it fails to account for tail risks beyond a certain threshold of a loss distribution. This shortcoming is particularly pronounced with heavy-tailed distributions, where VaR tends to underestimate the tail risk. To address this issue, ES was introduced. It provides a more accurate representation of potential tail risks under extreme conditions. A profound contribution by Artzner et al. (1999) introduced the concept of "coherent risk measures", which emphasizes the clear advantage of ES as a coherent measure for assessing tail risk compared to VaR.

Given that practitioners and regulators often have access to extensive datasets that can capture a more comprehensive set of tail risk characteristics, it is both theoretically intriguing and practically significant to explore inference methods for VaR and ES within the context of regression models. VaR's straightforward representation as a quantile of a loss distribution facilitates the direct application of quantile regression (QR) models, where the conditional quantiles of the loss distribution are modeled as a function of risk factors under specific scenarios. Numerous researchers have conducted fruitful studies in this area, such as Gutenbrunner and Jurecková (1992), He (1997), Koenker (2005), Zhou and Shao (2013), He et al. (2020). Upon these studies, Chernozhukov (2005) first considered a QR model at an extreme level but failed to achieve asymptotic normality. Wang et al. (2012) and Wang and Li (2013) derived the normal limit distribution using an extrapolation approach based on extreme value theory. Xu et al. (2022) studied the extreme expectile regression model by extrapolating an intermediate QR model. Girard et al. (2022) further considered the nonparametric extreme conditional expectile estimator in the context of conditional heavy-

Statistica Sinica: Preprint
doi:10.5705/ss.202025.0187

tailed distributions. Hou et al. (2024) employed a two-step procedure to estimate extreme conditional quantiles based on an extreme QR model with panel data. These studies collectively transition the classical QR model to the extreme QR model, facilitating applications in tail risk measurement.

In contrast, the ES estimation necessitates fundamentally different approaches due to its inherent non-elicitability, as formally established by Gneiting (2011). This property implies that ES cannot be directly estimated through conventional loss minimization frameworks. This characteristic poses a challenge in developing statistical inference methods for conditional ES within a regression framework. Recent studies have addressed this issue by proposing innovative approaches, such as *multi-objective elicitability*, which involves jointly modeling the quantile and expected shortfall through a minimization problem. Fissler and Ziegel (2016) showed that it is jointly elicitable with the quantile using a class of strictly consistent joint loss functions, enabling joint regression modeling for quantile and ES. Then, Dimitriadis and Bayer (2019) and Patton et al. (2019) proposed M-estimators (and Z-estimators) defined as the global minimum of these joint loss functions. However, the resulting optimization problem is computationally challenging due to the lack of differentiability and convexity in the loss function, despite established statistical properties. To mitigate the computational burden, Barendse (2020) introduced a two-step modeling procedure, bypassing the non-convexity problems, to tackle the non-eliciability of ES. In the first step, a (linear) QR model is fitted, and in the second step, a (linear) ES regression model is fitted by employing an Neyman-orthogonal score with substituting the unknown parameters of the fitted QR model. Consequently, the ES minimization problem in the second step incorporates the statistical uncertainty from the QR model in the first step. This two-step approach is more straightforward to implement in practice compared to the first, making its statistical properties particularly intriguing for further investigation.

Another challenge lies in the prediction of high-risk conditional ES within a regression framework. As ES provides a more precise assessment of tail risk under extreme conditions, there is significant interest in developing an extreme conditional ES model within a regression setting. To the best of our knowledge, no existing literature has yet addressed this specific issue. As suggested in He et al. (2023), current joint loss optimization and two-step methods perform poorly at extreme levels, as they are designed for fixed quantile levels. The main contribution of this paper lies in the integration of joint quantile and ES regression with extreme risk modeling – an area that remains largely unexplored and presents statistically

Statistica Sinica: Preprint
doi:10.5705/ss.202025.0187

intriguing properties worthy of further investigation. Moreover, we adopt the *heteroscedastic extreme* framework introduced by Einmahl et al. (2016) for modeling extreme risk, which provides an appropriate characterization of conditional distributions in regression settings. To be specific, let Y be a univariate response variable, $\mathbf{X}$ be a p-dimensional design vector, and denote $F_Y(\cdot)$ and $F_Y(\cdot|\mathbf{x})$ as the unconditional distribution of Y and the conditional distribution of Y given $\mathbf{X} = \mathbf{x}$. The first-order condition of heteroscedastic extreme is that, there exists a distribution $F_0(\cdot)$ such that

$$\lim_{t\to\infty} \frac{1 - F_Y(t|\mathbf{x})}{1 - F_0(t)} = g(\mathbf{x}), \tag{1.1}$$

uniformly for all $\mathbf{x}$ in a compact set, where g is a continuous and positive function (*scedastic function*). To analyse the asymptotic properties, Einmahl et al. (2016) presented a second-order regular variation condition (Assumption (1.b)) for the heteroscedastic extremes quantifying the rate of convergence in (1.1). Under this second-order condition, Xu et al. (2022) has shown that the conditional distribution $F_Y(\cdot|\mathbf{x})$ and unconditional distribution $F_Y(\cdot)$ fall in the same maximum domain of attraction with a uniform extreme value index γ. This allows for Hill estimator-based extrapolation in extreme conditional ES regression.

Upon the foundations of heteroscedastic extremes, we implement the extreme conditional ES regression as follows. First, we conduct the two-step

Statistica Sinica: Preprint
doi:10.5705/ss.202025.0187

procedure in (2.10) and (2.11) to fit conditional quantile and ES regression models at an intermediate level. Since the intermediate level varies with the sample size, replacing a fixed level with an intermediate one makes the theoretical analysis significantly more challenging and complex. This constitutes the main technical challenge. Stronger joint asymptotic normalities for the resulting estimators are presented in Proposition 2 and Theorem 1, which highlight the relevance between the quantile and ES regression components in the two-step procedure. Second, we develop several different approaches to extrapolate the extreme conditional ES estimators by examining the relationship between extreme ES and intermediate ES/quantile. Additionally, motivated by PELVE of Li and Wang (2023), we also develop an extrapolation method via level selection. Prior to our work, Xu et al. (2022) employed a similar approach to extrapolate the extreme expectile estimator using quantile regression via level selection. However, it fails to derive the asymptotic property. In contrast, we address this problem by selecting two intermediate orders and establish the corresponding asymptotic properties, which serve as another theoretical improvement of our work.

We organize this paper as follows. In Section 2, we first present the basic description for conditional quantile and conditional ES models and then studied the joint regression model at an intermediate level. The pro-

Statistica Sinica: Preprint
doi:10.5705/ss.202025.0187

posed methods for estimating the extreme conditional ES within a linear regression framework are discussed comprehensively in Section 3. Section 4 provides simulation evidence of the good finite-sample properties of our methods to predict the conditional ES at extreme levels. An empirical application in Section 5 further illustrates the effectiveness of our proposals.

2. A joint model for intermediate quantile and ES regression

Recall that the sample $\{(y_i, \mathbf{x}_i)\}_{i=1}^n$ are drawn independently from the distribution of a random vector $(Y, \mathbf{X})$, where $Y \in \mathbb{R}$ is the response variable and $\mathbf{X} \in \mathbb{R}^p$ is the predictor. In this paper, we consider $\mathbf{X}$ includes the unit as the first coordinate. We denote $F_Y(\cdot)$ and $F_Y(\cdot|\mathbf{x})$ to represent the unconditional distribution of Y and the conditional distribution of Y given $\mathbf{X} = \mathbf{x}$, respectively. Besides, the generalized inverse functions of $F_Y(\cdot)$ and $F_Y(\cdot|\mathbf{x})$ are denoted as $F_Y^{-1}(\cdot)$ and $F_Y^{-1}(\cdot|\mathbf{x})$, respectively. We denote $(a)_+ := \max(a, 0)$ and $(a)_- := \min(a, 0)$.

The conditional quantile of $F_Y(\cdot|\mathbf{x})$ given $\mathbf{X} = \mathbf{x}$ at a (fixed) level $\tau \in (0, 1)$ is defined as

$$Q_Y(\tau|\mathbf{x}) := F_Y^{-1}(\tau|\mathbf{x}) = \inf\{y \mid F_Y(y|\mathbf{x}) \geq \tau\},$$

or, equivalently, via an optimization such that

$$Q_Y(\tau|\mathbf{x}) = \arg\min_{q\in\mathbb{R}} E(\rho_\tau(Y-q)\,|\,\mathbf{X}=\mathbf{x}), \tag{2.1}$$

where $\rho_\tau(u) = (\tau - \mathrm{I}(u \le 0))u = |\tau - \mathrm{I}(u \le 0)|\,|u|$.

The conditional (right-tail) expected shortfall of $F_Y(\cdot|\mathbf{x})$ given $\mathbf{X}=\mathbf{x}$ at a level $\tau \in (0,1)$ is defined as

$$\mathrm{ES}_Y(\tau|\mathbf{x}) := \frac{1}{1-\tau}\int_\tau^1 F_Y^{-1}(t|\mathbf{x})\,dt = E(Y|Y \ge Q_Y(\tau|\mathbf{x}), \mathbf{X}=\mathbf{x}). \tag{2.2}$$

One can see that $\mathrm{ES}_Y(\tau|\mathbf{x})$ refers to the expectation of Y given $\mathbf{X}=\mathbf{x}$ conditional on the event $\{Y \ge Q_Y(\tau|\mathbf{x})\}$ of its distribution, indicating $\mathrm{ES}_Y(\tau|\mathbf{x})$ can describe the tail behavior of Y. Following Barendse (2020), He et al. (2023), ES can be characterized jointly with the conditional quantile by

$$\mathrm{ES}_Y(\tau|\mathbf{x}) = \arg\min_{e\in\mathbb{R}} E\left(\phi_\tau(Y-e, Y-Q_Y(\tau|\mathbf{x}))\,|\,\mathbf{X}=\mathbf{x}\right), \tag{2.3}$$

where $\phi_\tau(u,v) := \big((u-v) + \frac{v}{1-\tau}I(v \ge 0)\big)^2$. Moreover, one can derive the relationship between $Q_Y(\tau|\mathbf{x})$ and $\mathrm{ES}_Y(\tau|\mathbf{x})$ by,

$$\mathrm{ES}_Y(\tau|\mathbf{x}) = Q_Y(\tau|\mathbf{x}) + \frac{1}{1-\tau}E\left((Y-Q_Y(\tau|\mathbf{x}))_+|\mathbf{X}=\mathbf{x}\right). \tag{2.4}$$

The detailed derivations of (2.3), (2.4), as well as (2.9) below, are all contained in Supplement S3.

2.1 Tail behavior of conditional Expected Shortfall

To estimate conditional ES $\mathrm{ES}_Y(\tau_n'|\mathbf{x})$ at an extreme level τ_n', we are motivated by (2.3) to propose extrapolation methods utilizing extreme value theory. More specifically, we investigate the regular variation conditions of the extreme conditional ES, taking into account the tail behavior of the conditional distribution $F_Y(\cdot|\mathbf{x})$. We impose the following assumptions on the right tail of $F_Y(\cdot|\mathbf{x})$.

Assumption 1.

(1.a) The distribution of $\mathbf{X}$ *has a compact support* $\mathcal{X}$ *and* $E(\mathbf{X}\mathbf{X}^\top)$ *is positive definite.*

(1.b) There exist a positive and eventually decreasing function A *with* $\lim_{t\to\infty} A(t) = 0$*, and a positive continuous function* $g(\mathbf{x})$ *on* $\mathbf{X}$ *with* $E(g(\mathbf{X})) = 1$*, such that as* $t \to \infty$*,*

$$\sup_{\mathbf{x}\in\mathcal{X}} \left| \frac{1 - F_Y(t\,|\,\mathbf{x})}{1 - F_0(t)} - g(\mathbf{x}) \right| = O\left(A\left(\frac{1}{1 - F_0\,(t)} \right) \right). \tag{2.5}$$

(1.c) There exist some $\gamma > 0, \rho < 0$*, and a positive and eventually decreasing function* A_1 *with* $\lim_{t\to\infty} A_1(t) = 0$ *such that: for all* $x > 0$*,*

$$\lim_{t\to\infty} \frac{1}{A_1(1/(1 - F_0(t)))} \left(\frac{1 - F_0(tx)}{1 - F_0(t)} - x^{-\frac{1}{\gamma}} \right) = x^{-1/\gamma} \frac{x^{\rho/\gamma} - 1}{\gamma\rho}. \tag{2.6}$$

(1.d) *As* $t \to \infty$, $A(t) = o(A_1(t))$. *It also satisfies* $\lim_{n\to\infty} \sqrt{k} A_1(n/k) = 0$ *with* $k := k_n$ *such that* $k \to \infty, k/n \to 0$ *as* $n \to \infty$.

Assumption 1 characterizes the heteroscedastic extremes for the conditional distribution $F_Y(\cdot|\mathbf{x})$. Specifically, Assumption (1.a) regarding predictors is a typical condition for the asymptotic theory of quantile regression. Assumption (1.b) introduces a second-order condition for heteroscedastic extremes (see Einmahl et al. (2016)), indicating that the conditional distribution $F_Y(\cdot|\mathbf{x})$ has an equivalent tail to some $F_0(\cdot)$, but scaled by a function $g(\mathbf{x})$, where both $F_0(\cdot)$ and $g(\cdot)$ are defined in (1.1). Assumption (1.c) is a conventional second-order regular varying condition for $F_0(\cdot)$, while Assumption (1.d) specifies the convergence rate of two related auxiliary functions $A(t)$ and $A_1(t)$, with $A(t)$ converging slightly faster than $A_1(t)$. Note that $F_0(\cdot)$ can be replaced by the unconditional distribution $F_Y(\cdot)$. It is because, as shown in Lemmas S1 and S2, the conditional and unconditional distributions satisfy the same second-order regular variation as $F_0(\cdot)$ with a uniform extreme value index γ under Assumption 1. It suggests that the Hill estimator (3.1) for γ can be constructed by directly using the samples of response variable and applied it in extrapolations, regardless of the values of the predictors. Therefore, Assumptions (1.b) - (1.c) are milder, more general and more tractable, as extreme value index is usually conditional on

$\mathbf{x}$ if we assume regular variation on $F_Y(\cdot|\mathbf{x})$, which makes the extrapolation challenging to implement.

The following proposition shows the limiting behavior between $\mathrm{ES}_Y(\tau|\mathbf{x})$ and $Q_Y(\tau|\mathbf{x})$ and provides an important relationship to obtain the extrapolative approaches below.

Proposition 1. *Under Assumption 1 with $0 < \gamma < 1$, we have that as $\tau \to 1$,*

$$\sup_{\mathbf{x}\in\mathcal{X}} \left| \frac{\mathrm{ES}_Y(\tau|\mathbf{x})}{Q_Y(\tau|\mathbf{x})} - \frac{1}{1-\gamma} \right| = O\left(A_1\left(\frac{1}{1-\tau}\right)\right). \tag{2.7}$$

As indicated by Proposition 1, the ratio of $\mathrm{ES}_Y(\tau|\mathbf{x})$ and $Q_Y(\tau|\mathbf{x})$ converges to a constant $1/(1-\gamma)$ uniformly with a convergence rate dominated by $A_1(1/(1-\tau))$ as $\tau \uparrow 1$.

Remark 1. The first conclusion stems from Proposition 1 in Cai et al. (2015), which shows the limit of ratio between marginal expected shortfall $\mathbb{E}(Y|X > Q_X(\tau))$ and $Q_Y(\tau)$ as $\tau \uparrow 1$:

$$\lim_{\tau\uparrow 1} \frac{E(Y|X > Q_X(\tau))}{Q_Y(\tau)} = \int_0^\infty R(x^{-1/\gamma}, 1)\, dx.$$

It is straightforward to check that the integral $\int_0^\infty R(x^{-1/\gamma}, 1)\, dx = \frac{1}{1-\gamma}$ with $R(x, y) = x \wedge y$ when $X = Y$ for univariate case, agreeing with (2.7). The second one below stems from Proposition 3 in Li and Wang (2023),

$$\lim_{\varepsilon\downarrow 0} \frac{\mathrm{ES}_Y(1-t\varepsilon)}{Q_Y(1-\varepsilon)} = \frac{1}{1-\gamma} t^{-\gamma},$$

Statistica Sinica: Preprint
doi:10.5705/ss.202025.0187

for some $t > 0$, which can be treated as an unconditional version of (2.7) when we take a fixed value for $t = 1$.

2.2 Joint inference for intermediate quantile and ES regression

In this subsection, we study a joint model for intermediate conditional quantile regression and conditional ES regression, explicitly incorporating the tail behavior of the conditional ES. Throughout the article, we focus on the right-tailed conditional ES defined in (2.2) with a divergent risk level $\tau \uparrow 1$ to be either intermediate or extreme. More specifically, an intermediate level τ_n is a sequence of n satisfying $n(1-\tau_n) \to \infty$ and $\tau_n \to 1$ as $n \to \infty$, while an extreme level τ_n' is a sequence of n satisfying $n(1-\tau_n') \to c \in [0, \infty)$ and $\tau_n' \to 1$ as $n \to \infty$. In this paper, we focus on (conditional) linear regression models to simultaneously estimate both the intermediate conditional quantile and ES by

$$Q_Y(\tau_n|\mathbf{x}) = \boldsymbol{\beta}_0(\tau_n)^\top \mathbf{x} \quad \text{and} \quad \mathrm{ES}_Y(\tau_n|\mathbf{x}) = \boldsymbol{\theta}_0(\tau_n)^\top \mathbf{x}, \tag{2.8}$$

where the true values $\boldsymbol{\beta}_0(\tau_n)$ and $\boldsymbol{\theta}_0(\tau_n)$ are solutions to (2.1) and (2.3),

$$\boldsymbol{\beta}_0(\tau_n) := \arg\min_{\boldsymbol{\beta}} E\left(\rho_{\tau_n}(Y - \boldsymbol{\beta}^\top \mathbf{X})\right),$$

$$\boldsymbol{\theta}_0(\tau_n) := \arg\min_{\boldsymbol{\theta}} E\left(\phi_{\tau_n}\left(Y - \boldsymbol{\theta}^\top \mathbf{X}, Y - \boldsymbol{\beta}_0(\tau_n)^\top \mathbf{X}\right)\right).$$

The models given by equation (2.8) are referred to as intermediate quantile and ES regression models because the risk level τ_n is not a fixed level but an intermediate one. This distinction sets them apart from the models proposed in Dimitriadis and Bayer (2019), Patton et al. (2019), He et al. (2023). Consequently, the $\boldsymbol{\beta}_0(\tau_n)$ and $\boldsymbol{\theta}_0(\tau_n)$ are both sequences of parameter vectors. Based on the relationship (2.4), we derive that

$$\boldsymbol{\theta}_0(\tau_n) = \boldsymbol{\beta}_0(\tau_n) + \frac{1}{1-\tau_n}(E(\mathbf{X}\mathbf{X}^\top))^{-1} \cdot E\left((Y - \boldsymbol{\beta}_0(\tau_n)^\top \mathbf{X})_+ \cdot \mathbf{X}\right), \quad (2.9)$$

which plays an important role in the two-step estimation procedure.

The optimization problem for $\boldsymbol{\theta}_0(\tau_n)$ is complicated by the presence of the unknown true value $\boldsymbol{\beta}_0(\tau_n)$, which precludes direct minimization of $\boldsymbol{\theta}_0(\tau_n)$ based on the available data. To address this challenge and conduct inference for both the linear quantile regression and linear ES regression models at the intermediate level τ_n, we employ a two-step approach. In the first step, we fit an intermediate quantile regression model to obtain the estimator $\hat{\boldsymbol{\beta}}_n(\tau_n)$. In the second step, we fit an intermediate ES regression model by substituting the unknown parameter $\boldsymbol{\beta}_0(\tau_n)$ with the estimator $\hat{\boldsymbol{\beta}}_n(\tau_n)$. The procedures are given as follows:

Stage 1: Solve the following optimization to derive $\hat{\boldsymbol{\beta}}_n(\tau_n)$ by

$$\hat{\boldsymbol{\beta}}_n(\tau_n) := \arg\min_{\boldsymbol{\beta}} \sum_{i=1}^{n} \rho_{\tau_n}\left(y_i - \boldsymbol{\beta}^\top \mathbf{x}_i\right); \quad (2.10)$$

Stage 2: Given $\hat{\boldsymbol{\beta}}_n(\tau_n)$ in the first stage, solve the following optimization to derive $\hat{\boldsymbol{\theta}}_n(\tau_n)$ by

$$\hat{\boldsymbol{\theta}}_n(\tau_n) := \arg\min_{\boldsymbol{\theta}} \sum_{i=1}^{n} \phi_{\tau_n}\left(y_i - \boldsymbol{\theta}^\top \mathbf{x}_i, y_i - \hat{\boldsymbol{\beta}}_n(\tau_n)^\top \mathbf{x}_i\right), \tag{2.11}$$

which yields a closed expression,

$$\hat{\boldsymbol{\theta}}_n(\tau_n) = \hat{\boldsymbol{\beta}}_n(\tau_n) + \frac{1}{1-\tau_n}\left(\frac{1}{n}\sum_{i=1}^{n}\mathbf{x}_i\mathbf{x}_i^\top\right)^{-1} \cdot \left(\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{\boldsymbol{\beta}}_n(\tau_n)^\top \mathbf{x}_i) I(y_i \geq \hat{\boldsymbol{\beta}}_n(\tau_n)^\top \mathbf{x}_i) \cdot \mathbf{x}_i\right). \tag{2.12}$$

Note that (2.12) exhibits an empirical counterpart of (2.9). Then, we derive the predictions of both intermediate conditional quantile and ES by

$$\widehat{Q}_Y(\tau_n|\mathbf{x}) = \hat{\boldsymbol{\beta}}_n(\tau_n)^\top \mathbf{x} \quad \text{and} \quad \widehat{\mathrm{ES}}_Y(\tau_n|\mathbf{x}) = \hat{\boldsymbol{\theta}}_n(\tau_n)^\top \mathbf{x}. \tag{2.13}$$

It is worthy noting that (2.13) are not well-suited for estimation at extreme levels due to the scarcity of observations in the tail regions. We have implemented and compared (2.13) with other proposed extrapolative methods in simulations to highlight the necessity of extrapolation techniques. The asymptotic normality for $\hat{\boldsymbol{\beta}}_n(\tau_n)$ has already been established in Proposition S1. The proposition below provides a stronger result by establishing joint normality for $\hat{\boldsymbol{\beta}}_n(\tau_n)$ and $\hat{\boldsymbol{\theta}}_n(\tau_n)$, which serves as a key step in deriving the asymptotic properties of $\widehat{Q}_Y(\tau_n|\mathbf{x})$ and $\widehat{\mathrm{ES}}_Y(\tau_n|\mathbf{x})$. Denote $U_0 := (1/(1-F_0))^{-1}$ as the tail quantile function of $F_0(\cdot)$.

Proposition 2. *Suppose Assumption 1 holds with $0 < \gamma < 1/2$. Then, we have that as $n \to \infty$,*

$$\sqrt{n(1-\tau_n)}\left(\frac{\hat{\boldsymbol{\beta}}_n(\tau_n)-\boldsymbol{\beta}_0(\tau_n)}{U_0(1/(1-\tau_n))},\frac{\hat{\boldsymbol{\theta}}_n(\tau_n)-\boldsymbol{\theta}_0(\tau_n)}{U_0(1/(1-\tau_n))}\right)^\top \xrightarrow{d} N\left(\begin{pmatrix}\mathbf{0}_p\\ \mathbf{0}_p\end{pmatrix},\begin{pmatrix}\Sigma_{11} & \Sigma_{12}\\ \Sigma_{12} & \Sigma_{22}\end{pmatrix}\right). \tag{2.14}$$

Here, $\mathbf{0}_p$ denotes p-dimensional zero vector and Σ_{11}, Σ_{12}, Σ_{21}, Σ_{22} are all $p \times p$-dimensional matrixes, satisfying

$$\begin{cases}\Sigma_{11} &= \gamma^2\left(E\left(\mathbf{X}\mathbf{X}^\top g\left(\mathbf{X}\right)^{-\gamma}\right)\right)^{-1}E\left(\mathbf{X}\mathbf{X}^\top\right)\left(E\left(\mathbf{X}\mathbf{X}^\top g\left(\mathbf{X}\right)^{-\gamma}\right)\right)^{-1},\\ \Sigma_{12} &= B(2,1/\gamma-1)\left(E\left(\mathbf{X}\mathbf{X}^\top g(\mathbf{X})^{-\gamma}\right)\right)^{-1}E(\mathbf{X}\mathbf{X}^\top g(\mathbf{X})^{\gamma})\left(E(\mathbf{X}\mathbf{X}^\top)\right)^{-1},\\ \Sigma_{21} &= \Sigma_{12}^\top = B(2,1/\gamma-1)\left(E(\mathbf{X}\mathbf{X}^\top)\right)^{-1}E(\mathbf{X}\mathbf{X}^\top g(\mathbf{X})^{\gamma})\left(E\left(\mathbf{X}\mathbf{X}^\top g(\mathbf{X})^{-\gamma}\right)\right)^{-1},\\ \Sigma_{22} &= \frac{B(3,1/\gamma-2)}{\gamma}\left(E(\mathbf{X}\mathbf{X}^\top)\right)^{-1}E(\mathbf{X}\mathbf{X}^\top g(\mathbf{X})^{2\gamma})\left(E(\mathbf{X}\mathbf{X}^\top)\right)^{-1},\end{cases}$$

where $B(\cdot,\cdot)$ denotes the Beta function, $B(a,b)=\int_0^1 t^{a-1}(1-t)^{b-1}\,dt$ with $a>0, b>0$.

Note that the asymptotic properties of $\widehat{Q}_Y(\tau_n|\mathbf{x})$ alone have been well established in previous literature. It is also worth noting that a stronger condition $\gamma \in (0,1/2)$ is required to ensure the existence of the asymptotic variances, involving the calculation of second-order moments, see Lemma S4. Next, we can provide a jointly bivariate asymptotic normality for $\widehat{Q}_Y(\tau_n|\mathbf{x})$ and $\widehat{\mathrm{ES}}_Y(\tau_n|\mathbf{x})$ at an intermediate level τ_n.

Theorem 1. *Suppose Assumption 1 holds with $0 < \gamma < 1/2$. Then, we have that as $n \to \infty$,*

$$\sqrt{n(1-\tau_n)}\left(\frac{\widehat{Q}_Y(\tau_n|\mathbf{x})}{Q_Y(\tau_n|\mathbf{x})}-1, \frac{\widehat{\mathrm{ES}}_Y(\tau_n|\mathbf{x})}{\mathrm{ES}_Y(\tau_n|\mathbf{x})}-1\right)^\top \xrightarrow{d} N(\mathbf{0}_2, \Sigma_0). \tag{2.15}$$

Here, $\Sigma_0 = (\sigma_{ij}^2)$ is a 2×2-dimensional matrix with elements

$$\begin{cases} \sigma_{11}^2 &= \gamma^2 g(\mathbf{x})^{-2\gamma}\mathbf{x}^\top \left(E\left(\mathbf{X}\mathbf{X}^\top g\left(\mathbf{X}\right)^{-\gamma}\right)\right)^{-1} E\left(\mathbf{X}\mathbf{X}^\top\right)\left(E\left(\mathbf{X}\mathbf{X}^\top g\left(\mathbf{X}\right)^{-\gamma}\right)\right)^{-1}\mathbf{x}, \\ \sigma_{12}^2 &= (1-\gamma)B(2,1/\gamma-1)g(\mathbf{x})^{-2\gamma}\mathbf{x}^\top\left(E\left(\mathbf{X}\mathbf{X}^\top g(\mathbf{X})^{-\gamma}\right)\right)^{-1} E(\mathbf{X}\mathbf{X}^\top g(\mathbf{X})^{\gamma})\left(E(\mathbf{X}\mathbf{X}^\top)\right)^{-1}\mathbf{x}, \\ \sigma_{21}^2 &= (1-\gamma)B(2,1/\gamma-1)g(\mathbf{x})^{-2\gamma}\mathbf{x}^\top\left(E(\mathbf{X}\mathbf{X}^\top)\right)^{-1} E(\mathbf{X}\mathbf{X}^\top g(\mathbf{X})^{\gamma})\left(E\left(\mathbf{X}\mathbf{X}^\top g(\mathbf{X})^{-\gamma}\right)\right)^{-1}\mathbf{x}, \\ \sigma_{22}^2 &= \frac{(1-\gamma)^2 B(3,1/\gamma-2)}{\gamma} g(\mathbf{x})^{-2\gamma}\mathbf{x}^\top\left(E(\mathbf{X}\mathbf{X}^\top)\right)^{-1} E(\mathbf{X}\mathbf{X}^\top g(\mathbf{X})^{2\gamma})\left(E(\mathbf{X}\mathbf{X}^\top)\right)^{-1}\mathbf{x}, \end{cases}$$

where $B(\cdot,\cdot)$ denotes the Beta function, $B(a,b) = \int_0^1 t^{a-1}(1-t)^{b-1}\,dt$ with $a > 0, b > 0$.

3. Estimation for Extreme Conditional Expected Shortfall

In this section, we propose three distinct extrapolation methods for estimating conditional ES at an extreme level. After fitting a joint regression model by using two-step procedure at an intermediate level, we first extrapolate an extreme conditional ES based on the intermediate ES estimator; secondly, we extrapolate an extreme conditional ES based on an intermediate conditional quantile regression; additionally, we consider a third approach that extrapolates an extreme conditional ES regression from an intermediate conditional quantile regression, with the quantile level determined through a level selection method.

3.1 Extrapolation based on intermediate quantile and ES

In order to apply the extrapolation technique, we first need to estimate the extreme value index γ. According to Lemmas S1 and S2, $F_Y(\cdot\,|\mathbf{x}_i)$ contains a common γ, which is not affected by the predictor $\mathbf{x}_i$ in the estimators. Therefore, it is natural to use the Hill estimator for heteroscedastic extremes by utilizing only the data $\{y_i\}_{i=1}^n$:

$$\hat{\gamma} = \frac{1}{k}\sum_{j=1}^{k}(\log y_{n-j+1,n} - \log y_{n-k,n}), \tag{3.1}$$

where $y_{1,n} \le y_{2,n} \le \cdots \le y_{n,n}$ are the order statistics of the data $\{y_i\}_{i=1}^n$. We assume the intermediate level $\tau_n := 1 - k/n$, or equivalently $k = n(1 - \tau_n)$. A common method for selecting an appropriate k is to plot the Hill estimator $\hat{\gamma}$ against k and choose a k that corresponds to the first stable segment of the Hill plot. We will employ this method to select k in both our subsequent simulation studies and real data analysis. Moreover, the extrapolative relation between $Q_Y(\tau_n'|\mathbf{x})$ and $Q_Y(\tau_n|\mathbf{x})$ (see Lemma S2 in

Supplement S2), gives

$$\frac{Q_Y(\tau_n'|\mathbf{x})}{Q_Y(1-k/n|\mathbf{x})} = \frac{U_Y(1/(1-\tau_n')|\mathbf{x})}{U_Y(n/k|\mathbf{x})} \sim \left(\frac{k}{n(1-\tau_n')}\right)^{\gamma},$$

which accordingly implies that the extreme conditional quantile can be estimated by:

$$\widetilde{Q}_Y(\tau_n' \,|\, \mathbf{x}) = \left(\frac{k}{n(1-\tau_n')}\right)^{\hat{\gamma}} \widehat{Q}_Y(1-k/n \,|\, \mathbf{x}), \tag{3.2}$$

where $\widehat{Q}_Y(\tau_n \,|\, \mathbf{x})$ is well estimated by the two-step procedure (2.13).

We introduce three extrapolation approaches $\widehat{\mathrm{ES}}_Y^{(i)}(\tau_n'|\mathbf{x})$, $i = 1,2,3$, to estimate $\mathrm{ES}_Y(\tau_n'|\mathbf{x})$ by employing the relationships between $\mathrm{ES}_Y(\tau_n'|\mathbf{x})$ and one of $\mathrm{ES}_Y(\tau_n|\mathbf{x})$, $Q_Y(\tau_n|\mathbf{x})$, and $Q_Y(\tau_n'|\mathbf{x})$, respectively. First, by Proposition 1 and extrapolative relation of extreme quantile, it follows that,

$$\frac{\mathrm{ES}_Y(\tau_n'|\mathbf{x})}{\mathrm{ES}_Y(\tau_n|\mathbf{x})} = \frac{\mathrm{ES}_Y(\tau_n'|\mathbf{x})/Q_Y(\tau_n'|\mathbf{x})}{\mathrm{ES}_Y(\tau_n|\mathbf{x})/Q_Y(\tau_n|\mathbf{x})} \times \frac{Q_Y(\tau_n'|\mathbf{x})}{Q_Y(\tau_n|\mathbf{x})} \sim \left(\frac{1-\tau_n}{1-\tau_n'}\right)^{\gamma}. \tag{3.3}$$

This suggests us the estimator

$$\widehat{\mathrm{ES}}_Y^{(1)}(\tau_n'|\mathbf{x}) = \left(\frac{1-\tau_n}{1-\tau_n'}\right)^{\hat{\gamma}} \widehat{\mathrm{ES}}_Y(\tau_n|\mathbf{x}), \tag{3.4}$$

where $\widehat{\mathrm{ES}}_Y(\tau_n|\mathbf{x})$ is estimated by the two-step procedure (2.13).

Next, we derive the second extrapolation estimator by using Proposition 1 and the extrapolative relation of extreme quantile. Specifically, we have

$$\frac{\mathrm{ES}_Y(\tau_n'|\mathbf{x})}{Q_Y(\tau_n|\mathbf{x})} = \frac{\mathrm{ES}_Y(\tau_n'|\mathbf{x})}{Q_Y(\tau_n'|\mathbf{x})} \times \frac{Q_Y(\tau_n'|\mathbf{x})}{Q_Y(\tau_n|\mathbf{x})} \sim \left(\frac{1}{1-\gamma}\right)\left(\frac{1-\tau_n}{1-\tau_n'}\right)^{\gamma}. \tag{3.5}$$

Statistica Sinica: Preprint
doi:10.5705/ss.202025.0187

We then obtain the second estimator for $\mathrm{ES}_Y(\tau_n'|\mathbf{x})$:

$$\widehat{\mathrm{ES}}_Y^{(2)}(\tau_n'|\mathbf{x}) = \left(\frac{1}{1-\hat{\gamma}}\right)\left(\frac{1-\tau_n}{1-\tau_n'}\right)^{\hat{\gamma}} \widehat{Q}_Y(\tau_n|\mathbf{x}), \tag{3.6}$$

where $\widehat{Q}_Y(\tau_n|\mathbf{x})$ is estimated via (2.13).

Finally, the third extrapolation approach is based on the relationship between $\mathrm{ES}_Y(\tau_n'|\mathbf{x})$ and $Q_Y(\tau_n'|\mathbf{x})$,

$$\frac{\mathrm{ES}_Y(\tau_n'|\mathbf{x})}{Q_Y(\tau_n'|\mathbf{x})} \sim \frac{1}{1-\gamma}, \tag{3.7}$$

which is a direct application of Proposition 1. Thus the third estimator for $\mathrm{ES}_Y(\tau_n'|\mathbf{x})$ can be given by

$$\widehat{\mathrm{ES}}_Y^{(3)}(\tau_n'|\mathbf{x}) = \frac{1}{1-\hat{\gamma}}\widetilde{Q}_Y(\tau_n'|\mathbf{x}) = \left(\frac{1}{1-\hat{\gamma}}\right)\left(\frac{k}{n(1-\tau_n')}\right)^{\hat{\gamma}} \widehat{Q}_Y(1-k/n\,|\,\mathbf{x}). \tag{3.8}$$

It is evident that $\widehat{\mathrm{ES}}_Y^{(2)}(\tau_n'|\mathbf{x})$ and $\widehat{\mathrm{ES}}_Y^{(3)}(\tau_n'|\mathbf{x})$ are essentially the same estimator because of $1-\tau_n = k/n$. Therefore, in simulation and empirical analysis, we only implement $\widehat{\mathrm{ES}}_Y^{(1)}(\tau_n'|\mathbf{x})$ and $\widehat{\mathrm{ES}}_Y^{(3)}(\tau_n'|\mathbf{x})$ for comparison, by choosing a suitable k instead of τ_n. Moreover, we establish the asymptotic relationships between $\widehat{\mathrm{ES}}_Y^{(i)}(\tau_n'|\mathbf{x})$ $(i=1,2,3)$ and $\widetilde{Q}_Y(\tau_n'\,|\,\mathbf{x})$ as follows.

Proposition 3. *Under the conditions of Theorem 1, we have that,*

$$\frac{\widehat{\mathrm{ES}}_Y^{(i)}(\tau_n'|\mathbf{x})}{\widetilde{Q}_Y(\tau_n'\,|\,\mathbf{x})} = \frac{1}{1-\gamma} + o_{\mathbb{P}}\left(\frac{1}{\sqrt{k}}\right), \ \textit{for } i=1,2,3. \tag{3.9}$$

It is worth noting that the ratios $\widehat{\mathrm{ES}}_Y^{(i)}(\tau_n'|\mathbf{x})/\widetilde{Q}_Y(\tau_n'\,|\,\mathbf{x})$ $(i = 1,2,3)$ tend to $1/(1-\gamma)$ in probability as $n \to \infty$, aligning with the limit given in (2.7). In the subsequent theorem, we further analyze the asymptotic properties of $\widehat{\mathrm{ES}}_Y^{(i)}(\tau_n'|\mathbf{x})$ with $i = 1,2,3$.

Theorem 2. *Under the conditions of Theorem 1, we suppose* $d_n := \frac{k}{n(1-\tau_n')} \to \infty$ *and* $\sqrt{k}/\log d_n \to \infty$ *as* $n \to \infty$. *Then, we have that, as* $n \to \infty$,

$$\frac{\sqrt{k}}{\log d_n}\left(\frac{\widehat{\mathrm{ES}}_Y^{(i)}(\tau_n'|\mathbf{x})}{\mathrm{ES}_Y(\tau_n'|\mathbf{x})} - 1\right) \xrightarrow{d} N\left(0,\gamma^2\right), \text{ for } i = 1,2,3. \tag{3.10}$$

At an extreme level, the influence of the intermediate parts $\widehat{\mathrm{ES}}_Y(\tau_n|\mathbf{x})$ and $\widehat{Q}_Y(\tau_n|\mathbf{x})$ diminishes under the standard extreme convergence rate $\sqrt{k}/\log d_n$, despite their differences. Consequently, all the proposed estimators $\widehat{\mathrm{ES}}_Y^{(i)}(\tau_n'|\mathbf{x})$ $(i = 1,2,3)$ share the same asymptotic normality.

3.2 Extrapolation based on level selection method

Alternatively, PELVE proposed by Li and Wang (2023), motivates us to extrapolate an extreme conditional ES estimator via a level selection method. Specifically, we can select an ES level τ and a quantile level ω such that

$$\mathrm{ES}_Y(\tau|\mathbf{x}) = Q_Y(\omega|\mathbf{x}), \tag{3.11}$$

conditional on $\mathbf{x}$. The levels τ and ω are closely related and can be expressed as $\omega = \omega(\tau)$ or $\tau = \tau(\omega)$. Unlike PELVE, (3.11) does not emphasize the uniqueness or size relationship, although it does imply that $\omega \geq \tau$. Rather, it only requires the existence of τ and ω that satisfies (3.11). The proposition below states that the two levels are of the same type. Even if multiple values of τ or ω may satisfy (3.11), they are functionally equivalent in extrapolation, and therefore it suffices to select any one of them.

Proposition 4. *Under Assumption 1 with* $0 < \gamma < 1$ *and* (3.11)*, we have,*

$$\lim_{\tau \to 1} \frac{1-\tau}{1-\omega(\tau)} = \left(\frac{1}{1-\gamma} \right)^{1/\gamma}. \tag{3.12}$$

This implies τ and ω share the same extremeness (intermediate or extreme). We can also interpret (3.12) as a variant of Theorem 3 in Li and Wang (2023), which shows that $c(\varepsilon)$ tends to the same limit as in (3.12). Based on (3.12), we can estimate the level $\omega(\tau)$ by using the Hill estimator:

$$\hat{\omega}(\tau) = 1 - (1-\tau)(1-\hat{\gamma})^{1/\hat{\gamma}}, \tag{3.13}$$

as $\tau \to 1$, or conversely, we can estimate $\tau(\omega)$ by, as $\omega \to 1$,

$$\hat{\tau}(\omega) = 1 - (1-\omega)(1-\hat{\gamma})^{-1/\hat{\gamma}}. \tag{3.14}$$

We now introduce another novel extrapolative method based on (3.13). Specifically, we first select an intermediate ES level τ_n and define the corresponding intermediate quantile level $\hat{\omega}(\tau_n)$ using (3.13). We then extrapolate the extreme ES estimator as follows:

$$\widehat{\mathrm{ES}}_Y^{(4)}(\tau_n'|\mathbf{x}) = \left(\frac{1-\tau_n}{1-\tau_n'}\right)^{\hat{\gamma}} \widehat{\mathrm{ES}}_Y(\tau_n|\mathbf{x}) = \left(\frac{1-\tau_n}{1-\tau_n'}\right)^{\hat{\gamma}} \widehat{Q}_Y(\hat{\omega}(\tau_n)|\mathbf{x}). \quad (3.15)$$

This method provides a unique way to extrapolate the ES estimator by leveraging the relationship between the intermediate levels τ_n and $\hat{\omega}(\tau_n)$, offering a distinct approach from (3.6) and (3.8). As the quantile level is an estimator, the asymptotic normality of $\widehat{\mathrm{ES}}_Y^{(4)}(\tau_n'|\mathbf{x})$ may have complicated uncertainty and limiting distribution. To establish the asymptotic normality, we select two intermediate orders k and $\widetilde{k}$, where k is used for the Hill estimator $\hat{\gamma}$, and $\widetilde{k}$ is used for the intermediate level $\tau_n := 1 - \widetilde{k}/n$ in the quantile regression. It is worth noting that in the first three extrapolation methods (3.4), (3.6), and (3.8), we use $\tau_n = 1-k/n$ as the intermediate level in the regression, while in the fourth extrapolation approach (3.15) based on the level selection method, we use $\tau_n = 1-\widetilde{k}/n$ as the intermediate level. In addition, we study the asymptotic relationships between $\widehat{\mathrm{ES}}_Y^{(4)}(\tau_n'|\mathbf{x})$ and $\widetilde{Q}_Y(\tau_n' \,|\, \mathbf{x})$ as follows.

Proposition 5. *Let $\tau_n = 1 - \widetilde{k}/n$. Suppose $\widetilde{k} \to \infty$, $\widetilde{k}/n \to 0$, $\widetilde{k} = o(k)$ and $\sqrt{\widetilde{k}}A_1(n/\widetilde{k}) = o(1)$, under the conditions of Theorem 1, we have that,*

$$\frac{\widehat{\mathrm{ES}}_Y^{(4)}(\tau_n'|\mathbf{x})}{\widetilde{Q}_Y(\tau_n' \,|\, \mathbf{x})} = \frac{1}{1-\gamma} + O_{\mathbb{P}}\left(\frac{1}{\sqrt{\widetilde{k}}}\right). \quad (3.16)$$

While the ratio converges to $1/(1-\gamma)$ as established in Proposition 1, its convergence rate differs notably from those of $\widehat{\mathrm{ES}}_Y^{(i)}(\tau_n'|\mathbf{x})/\widetilde{Q}_Y(\tau_n'\,|\,\mathbf{x})$ $(i=1,2,3)$ in Proposition 3. The slower convergence of the fourth estimator stems from the additional estimation uncertainty in $\hat{\omega}(\tau_n)$, which reduces the convergence rate from $1/\sqrt{k}$ to $1/\sqrt{\widetilde{k}}$ (see Proposition S2). Subsequently, we delve into a more detailed analysis of the asymptotic properties of $\widehat{\mathrm{ES}}_Y^{(4)}(\tau_n'|\mathbf{x})$.

Theorem 3. *Recall that $\tau_n = 1-\widetilde{k}/n$. Under the conditions of Theorem 1 and Proposition 5, we further suppose $\widetilde{d}_n = \frac{\widetilde{k}}{n(1-\tau_n')} \to \infty$, $(\widetilde{k}/k)^{1/2}\log \widetilde{d}_n \to \infty$, and $\sqrt{k}/\log \widetilde{d}_n \to \infty$ as $n\to\infty$. Then, we have that as $n\to\infty$,*

$$\frac{\sqrt{k}}{\log \widetilde{d}_n}\left(\frac{\widehat{\mathrm{ES}}_Y^{(4)}(\tau_n'|\mathbf{x})}{\mathrm{ES}_Y(\tau_n'|\mathbf{x})}-1\right) \xrightarrow{d} N\left(0,\gamma^2\right). \tag{3.17}$$

The extrapolation method described in (3.15) is also implemented in the subsequent simulation to facilitate practical comparison.

4. Simulation

In this section, we conduct Monte Carlo simulations to compare the performance of our proposed methods (3.4), (3.8), and (3.15) with that of direct estimator at extreme levels. We consider the following multivariate

Statistica Sinica: Preprint
doi:10.5705/ss.202025.0187

predictors as our potential model:

$$y_i = x_{i1} + x_{i2} + (1 + r x_{i1})\sigma_i, \text{ for } i = 1, 2, ..., n.$$

Here, x_{i1} and x_{i2} are independent standard uniform random variables, and the errors σ_i's are also independent random variables, which are also independent of x_{i1} and x_{i2}. These errors are generated from the following three populations with $\gamma = 0.2, 0.3$, and 0.4:

- Student-t distribution with degree of freedom $1/\gamma$;
- Pareto distribution with CDF: $F(x) = 1 - x^{-1/\gamma}$, $x > 1$;
- Fréchet distribution with CDF: $F(x) = \exp\left\{-x^{-1/\gamma}\right\}$, $x > 0$.

In addition, the coefficient r is a constant that controls the degree of heteroscedasticity. We set $r = 0$, 0.5, and 0.9 in this study. Therefore, the τ-th true conditional expected shortfall of Y is given by

$$\mathrm{ES}_Y(\tau|\mathbf{x}_i) = \boldsymbol{\theta}(\tau)^\top \mathbf{x}_i,$$

where $\mathbf{x}_i = (1, x_{i1}, x_{i2})^\top$, $\boldsymbol{\theta}(\tau) = (\theta_0(\tau), \boldsymbol{\theta}_1(\tau)^\top)^\top$ with $\theta_0(\tau) = \mathrm{ES}_\sigma(\tau)$ and $\boldsymbol{\theta}_1(\tau) = (1 + r\mathrm{ES}_\sigma(\tau), 1)^\top$. Here, $\mathrm{ES}_\sigma(\tau)$ denotes the expected shortfall of error variable σ at level τ. In this study, we consider three extreme ES levels $\tau_n' = 0.99, 0.995, 0.999$ and repeat the simulation $m = 500$ times with sample size $n = 1000, 2000$, and 3000. Moreover, we will implement the following four methods to compare their finite-sample performance:

- Method I: a direct estimator $\widehat{\mathrm{ES}}_Y(\tau_n'|\mathbf{x})$ obtained by applying the two-step procedure (2.13) without any extrapolation. This serves as a benchmark to demonstrate the necessity of employing extrapolative techniques in the estimation of extreme conditional ES.

- Method II: our proposed method $\widehat{\mathrm{ES}}_Y^{(1)}(\tau_n'|\mathbf{x})$ in (3.4), which utilizes the first extrapolative approach.

- Method III: our proposed method $\widehat{\mathrm{ES}}_Y^{(3)}(\tau_n'|\mathbf{x})$ in (3.8), which employs the third extrapolative approach. Note that this method is equivalent to $\widehat{\mathrm{ES}}_Y^{(2)}(\tau_n'|\mathbf{x})$ in (3.6).

- Method IV: our proposed method $\widehat{\mathrm{ES}}_Y^{(4)}(\tau_n'|\mathbf{x})$ in (3.15), which incorporates the fourth extrapolative approach via a level selection method.

Both the Hill estimators and the extrapolation techniques necessitate the selection of an intermediate order k (and $\widetilde{k}$). To determine the appropriate k, we plot the Hill estimator $\hat{\gamma}$ against k and identify the first stable segment of the plot, from which we select the corresponding k. Following the approach in Wang et al. (2012), we choose $\widetilde{k} = \left[k/(\log n)^{1/4}\right]$ for the extrapolation techniques employed in Method IV. For each case, we initially use one replication to determine the optimal k, which is then applied consistently across all $m = 500$ replications for extrapolation. To assess the performance of each method, we compute the *integrated square error* (ISE) using $L = 100$ new data points. The ISE serves as a metric to evaluate the average prediction accuracy of each method. Specifically, for the j-th simulation, we estimate $\mathrm{ES}_Y(\tau_n'|\mathbf{x}_l^*)$ and calculate the ISE as follows:

$$\mathrm{ISE}_j = \frac{1}{L}\sum_{l=1}^{L}\left(\frac{\widehat{\mathrm{ES}}_Y^{(j)}(\tau_n'|\mathbf{x}_l^*)}{\mathrm{ES}_Y(\tau_n'|\mathbf{x}_l^*)} - 1\right)^2, \text{ for } j = 1, 2, ..., m,$$

where $\mathbf{x}_1^*, ..., \mathbf{x}_L^*$ are independent random replications of $\mathbf{X}$ for each simulation, and $\mathrm{ES}_Y(\tau_n'|\mathbf{x}_l^*)$ is the true conditional ES derived from the simulation models. Tables S1-S9 (see Supplement S1) provide a comprehensive summary of the mean and standard error of $\{\mathrm{ISE}_j : j = 1, 2, ..., m\}$ for $r = 0$, 0.5, and 0.9 across all models with varying sample sizes and tail indices.

Our analysis reveals several key insights. When the random error term follows a Pareto or a Fréchet distribution, both the mean and variance of the ISE for all four methods consistently increase as the level τ approaches 1. This trend indicates that the estimation becomes more challenging as we move towards more extreme quantiles, due to the sparsity of data in the tail regions. In contrast, when the random error term follows a Student-t distribution, the mean of the ISE for Methods I, III, and IV, as well as the

Statistica Sinica: Preprint
doi:10.5705/ss.202025.0187

variance for all four methods, also increase as τ approaches 1. However, under moderate and heavy tails ($\gamma = 0.3$ and 0.4), an interesting phenomenon emerges: only when the random error term follows the Student-t distribution does the mean of Method II decrease as τ increases. This unique behavior can be attributed to the fact that Method II extrapolates the ES from an intermediate level to an extreme level based on the ES regression, leveraging the heterogeneity information above the intermediate quantiles more effectively. In contrast, Methods III and IV are based on quantile regression at intermediate levels, which only account for threshold values without considering the heterogeneity information from points above the tails. This distinction is crucial, as it highlights the importance of the extrapolation techniques in capturing the tail behavior accurately. Furthermore, while bias and standard deviation generally decrease with increasing sample size, practical constraints often limit our ability to obtain sufficiently large samples. This limitation implies the necessity of employing extrapolation techniques to estimate the extreme conditional ES, especially when data in the tail region are scarce.

In the comparison of the four methods, we observe that, owing to the lack of sufficient sample size for tail data, Method I is generally less effective than the other three methods in estimating the extreme conditional ES and

Statistica Sinica: Preprint
doi:10.5705/ss.202025.0187

the corresponding variance. This further underscores the importance of extrapolation techniques in the practical estimation of extreme conditional ES, consistent with the observations from previous studies.

To compare Methods II, III, and IV under the same level of heavy-tailedness, we present the results in Tables S1–S9. For both Student-t and Pareto-distributed errors, the optimal method exhibits a consistent pattern across varying degrees of tail heaviness. Under the weak tail heaviness ($\gamma = 0.2$, Table S1 and S4), both Methods III and IV demonstrate strong performance. Specifically, Method IV is slightly preferable under low heteroscedasticity ($r = 0, 0.5$), whereas Method III performs better under high heteroscedasticity ($r = 0.9$). Under the moderate tail heaviness ($\gamma = 0.3$, Table S2 and S5), Methods III and IV continue to outperform others, with Method III holding a slight advantage. However, under the strong tail heaviness ($\gamma = 0.4$, Table S3 and S6), Method II clearly emerges as the best performer. For Fréchet-distributed errors, a distinct pattern is observed. Method IV significantly outperforms the other three methods, particularly in scenarios with no data heterogeneity and in cases of high data heterogeneity at more extreme levels for the estimation of conditional ES. In a few instances, where data exhibits heterogeneity but the extreme level is not exceedingly high, Method III performs slightly better than Method IV.

The relative merits of Methods II, III and IV hinge on both heteroscedasticity and sample size. Method III dominates in most heteroscedastic regimes, yet Method II retains its edge under heavy tails even when heteroscedasticity is pronounced. With small samples, the best choice further varies with error distribution, tail weight and heteroscedastic degree. Method IV leads when $\gamma = 0.2$ (Table S7) and stays competitive for $\gamma = 0.3$ (Table S8), especially under no or extreme-level heteroscedasticity; Method III edges it out for intermediate heteroscedasticity. Method II dominates under $\gamma = 0.4$ (Table S9) and records the smallest variance.

Overall, there is no single dominant method, but a clear logic for selection emerges: Method II is the most robust for heavy-tailed data. Method III offers the best overall performance for Student-t and Pareto-distributed errors under light and moderate tails with high heteroscedasticity. Method IV excels for Fréchet-distributed error under light tail and moderate tail with limited sample size.

5. Real Data Analysis

In this section, we apply the four proposed methods to estimate the extreme conditional ES for the weekly market loss (negative return) of the Dow Jones Industrial Average (DJI30). The dataset spans from January 1, 1993, to June 30, 2013, comprising 1,066 observations, sourced from Yahoo Finance. This dataset captures two significant recessions (in 2001 and 2008-2009) as well as several financial crises (in 1994, 1997, 1998, 2000, 2008, and 2011). Given that weekly returns exhibit much weaker autocorrelation, we treat the weekly observations as approximately independent for modeling purposes. The variables included in the dataset are `ret` (weekly market return), `yield3m` (three-month yield change), `credit` (credit spread change), `term` (term spread change), `ted` (short-term TED spread), and `housing` (real estate excess return). Specifically, the variables in our models are defined as follows: 1) Y: the negative weekly return of DJI30 index; 2) x_1: the change in the three-month yield from the Federal Reserve Board's H.15 release; 3) x_2: the change in the credit spread between Moody's Baarated bonds and the ten-year Treasury rate from the Federal Reserve Board's H.15 release; 4) x_3: the change in the slope of the yield curve, measured by the spread between the composite long-term bond yield and the three-month bill rate obtained from the Federal Reserve Board's H.15 release; 5) x_4: a short-term TED spread, defined as the difference between the three-month LIBOR rate and the three-month secondary market treasury bill rate; 6) x_5: the weekly real estate sector return in excess of the market financial sector return (from the real estate companies with SIC code 65-66).

We focus on estimating the extreme conditional ES of the weekly loss of DJI30 at the levels $\tau_n' = 0.99$ and 0.999, utilizing extrapolation techniques. For comparative purposes, we employ the direct estimator $\widehat{\mathrm{ES}}_Y(\tau_n'|\mathbf{x})$ by implementing a two-step procedure, as detailed in (2.13), without any extrapolation. This approach, referred to as Method I in Section 4, serves as our baseline. Additionally, we apply the three proposed Methods II, III, and IV, to estimate the extreme conditional ES, leveraging extrapolation techniques to enhance the accuracy and robustness of our estimators.

First, we perform a visual analysis to assess the validity of the heavy-tail assumption for the data and display the results in Figure 1. The visualizations reveal that the losses exhibit a distribution characterized by a sharp peak and heavy tails, with the upper tail (large losses) being particularly pronounced. The selection method for the thresholds k and $\tilde{k}$ is consistent with the approach detailed in Section 4. We plot the Hill estimator $\hat{\gamma}$ against k in the right panel of Figure 1. Based on this plot, we determine that $k = 45$ is an appropriate choice for the weekly loss of DJI30.

We compare the extreme conditional ES under the different proposed methods and display the results in Figure 2. Figure 2 illustrates the in-sample ES estimators for the weekly loss of DJI30 from January 1, 1993, to June 30, 2013, using Methods I, II, III, and IV at the levels $\tau_n' = 0.99$

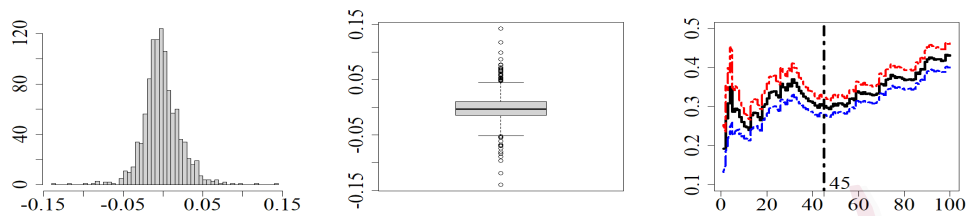

Figure 1: The left and middle panels depict the density histogram and boxplot of the weekly losses of DJI30. The right panel presents the Hill plot for the selection of k: black line is the Hill estimator against k, the upper red and lower blue dashed lines are the 90% confidence bounds, the vertical line shows the selected k.

and 0.999, respectively. As shown in Figure 2, all four methods exhibit a similar trend in estimating the conditional extreme ES. The estimates from Method II are close to those from Method I, whereas the estimates from Methods III and IV are slightly higher than those from Method I. As the level increases, Figure 2 indicates that the estimates derived from Methods II, III, and IV exceed those estimated from Method I. This suggests that, with an increase in the extreme level, estimators without extrapolation may lead to an underestimation of tail risks, while extrapolation-based methods can assist financial institutions in more effectively mitigating tail risk. Additionally, as the extreme level increases, the trajectory of estimates from

Method I tends to be flat, while the extrapolation-based Methods II, III, and IV still maintain volatility. These findings underscore the effectiveness and necessity of extrapolation in estimating conditional extreme ES.

During periods of economic recessions, the estimates of ES obtained from all four methods exhibit a significant increase, indicating a rise in tail risk. This phenomenon is evident during the recession caused by the bursting of the internet bubble in 2001, the instability in the U.S. markets resulting from the Asian financial crisis and the collapse of Long-Term Capital Management in 1998, as well as during the U.S. subprime mortgage crisis from 2008 to 2009. Notably, during these crisis periods, the estimates of extreme conditional ES reached their peaks, which is consistent with actual market conditions. In comparing the three extrapolation-based methods (Methods II, III, and IV), Figures 2 illustrates that the estimates of Method II are slightly lower than those from Methods III and IV. Moreover, while the estimates from Methods III and IV are relatively close, Method IV displays greater volatility during crisis periods. These observations suggest that our proposed extreme conditional ES estimators serve as effective risk measurement tools for measuring extreme tail risks and exhibit robustness.

Furthermore, we employ a rolling approach to estimate the extreme conditional ES and compare it with the in-sample estimators. This rolling

Statistica Sinica: Preprint
doi:10.5705/ss.202025.0187

out-of-sample analysis is conducted using expanding windows. We select a rolling window size of 710 weeks, which constitutes approximately two-thirds of the entire data period, to forecast the remaining 356 weeks, or approximately 7 years, in a single step. Figure S1 (in Supplement S1) illustrates the in-sample and out-of-sample conditional ES estimators of Methods II, III, and IV at levels $\tau_n' = 0.99, 0.999$. Overall, the in-sample and out-of-sample estimators demonstrate similar trends, although the in-sample estimators exhibit greater volatility. Notably, during the crisis period from 2008 to 2009, both out-of-sample estimators, in line with the in-sample estimators, reveal a pronounced upward trend, indicating an escalation in tail risk. Our empirical analysis substantiates that our proposed methods for extreme conditional ES serve as effective risk metrics for crisis forecasting, particularly in relation to loss values within the tail regions.

6. Conclusion

In this paper, we focus on the estimation of extreme conditional ES regression within a joint framework integrating quantile regression, particularly in view of heteroscedastic extremes. Existing studies have primarily developed estimation methods for fixed quantile levels, but have not adequately addressed intermediate or extreme levels, as highlighted in Barendse (2020),

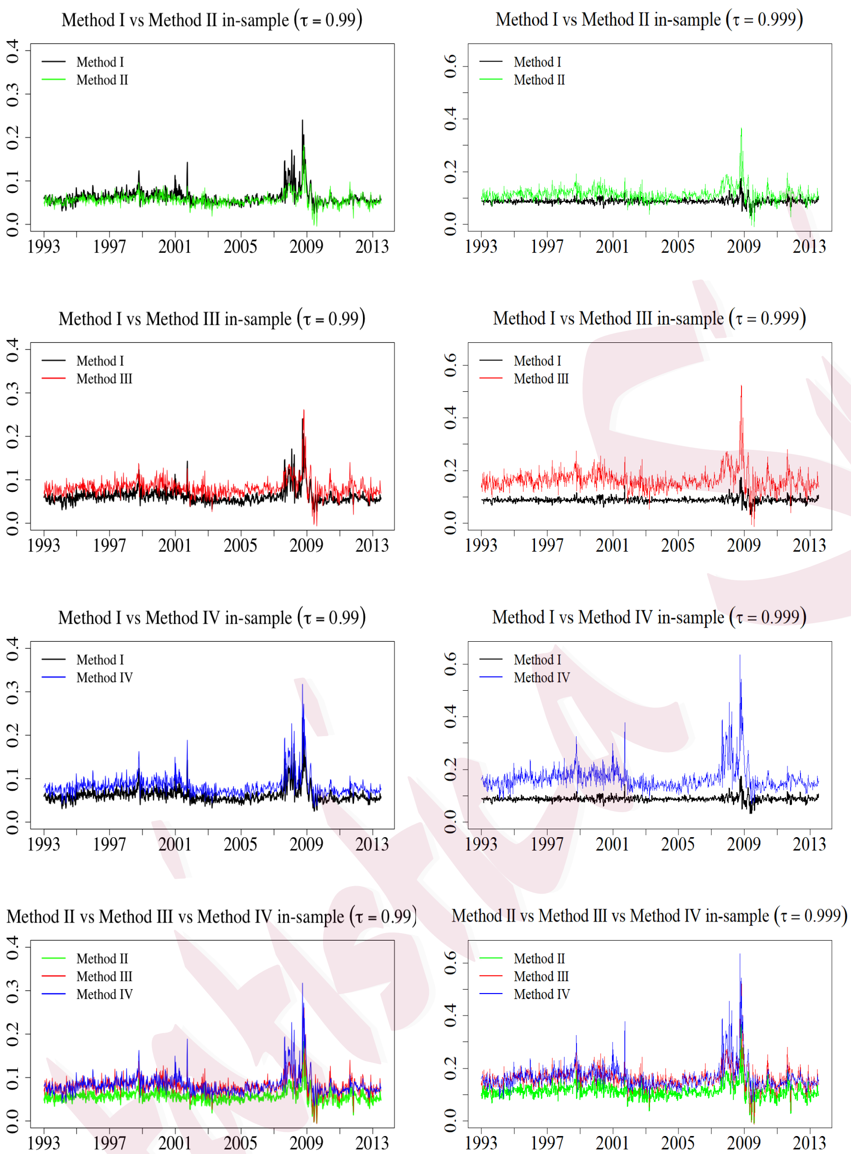

Figure 2: The results of the in-sample conditional ES estimators of Method I (black curve), Method II (green curve), Method III (red curve) and Method IV (blue curve) with $\tau_n' = 0.99$ (left panel) and 0.999 (right panel).

Dimitriadis and Bayer (2019), He et al. (2023). This limitation arises from the scarcity of tail observations, which poses significant challenges for effective estimation. To address this issue, we propose several extrapolative methods for estimating extreme ES regression based on conditional heteroscedastic EVT. Specifically, we first fit intermediate conditional quantile and ES regression models using a two-step procedure recently introduced by Barendse (2020). We then extrapolate an extreme ES by examining various relationships between ES and quantiles. Simulation results demonstrate that all proposed methods outperform the direct application of the two-step method at high levels. A real financial example further highlights the practical advantages of the ES regression model. Although we restrict attention to linear models, our underpinning techniques pave the way for analysing joint non-parameter quantile-ES models and high-dimensional sparse quantile-ES model. We leave these extensions in future research.

Supplementary Material

The online Supplementary Material contains some simulation results, auxiliary results and all technical proofs.

Acknowledgments

Jingyu Ji's research was partially supported by the National Natural Science Foundation of China grant 72403172 and the Young Teacher Foundation from Capital University of Economics and Business XRZ2024050. Liujun Chen's research was partially supported by the National Natural Science Foundation of China grants 12301387 and 12471279. Yanxi Hou's research was partially supported by the National Natural Science Foundation of China grant 72171055. Deyuan Li's research was partially supported by the National Natural Science Foundation of China grant 12471279.

References

Artzner, P., Delbaen, F., Eber, J.M., and Heath, D. (1999). Coherent measures of risk. *Math. Finance* **9**, 203-228.

Barendse, S. (2020). Efficiently weighted estimation of tail and interquantile expectations. *SSRN*

Cai, J.J., Einmahl, J.H., Haan, L., and Zhou, C. (2015). Estimation of the marginal expected shortfall: the mean when a related variable is extreme. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **77**, 417-442.

Chernozhukov, V. (2005). Extremal quantile regression. *Ann. Stat.* **33**, 806-839.

Dimitriadis, T., and Bayer, S. (2019). A joint quantile and expected shortfall regression framework. *Electron. J. Stat.* **13**, 1823-1871.

Einmahl, J.H., Haan, L., and Zhou, C. (2016). Statistics of heteroscedastic extremes. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **78**, 31-51.

Fissler, T., and Ziegel, J.F. (2016). Higher order elicitability and osband's principle. *Ann. Stat.* **44**, 1680-1707.

Girard, S., Stupfler, G., and Usseglio-Carleve, A. (2022). Nonparametric extreme conditional expectile estimation. *Scand. J. Stat.* **49**, 78-115.

Gneiting, T. (2011). Making and evaluating point forecasts. *J. Am. Stat. Assoc.* **106**, 746-762.

Gutenbrunner, C., and Jurecková, J. (1992). Regression rank scores and regression quantiles. *Ann. Stat.* **20**, 305-330.

He, X. (1997). Quantile curves without crossing. *Am. Stat.* **51**, 186-192.

He, X., Tan, K.M., and Zhou, W.X. (2023). Robust estimation and inference for expected shortfall regression with many regressors. *J. R. Stat. Soc. Ser. B* **85**, 1223-1246.

He, Y., Hou, Y., Peng, L., and Shen, H. (2020). Inference for conditional value-at-risk of a predictive regression. *Ann. Stat.* **48**, 3442-3464.

Hoga, Y. (2019). Confidence intervals for conditional tail risk measures in ARMA–GARCH models. *J. Bus. Econ. Stat.* **37**, 613-624.

Hou, Y., Leng, X., Peng, L., and Zhou, Y. (2024). Panel quantile regression for extreme risk. *J. Econometrics* **240**, 105674.

Koenker, R. (2005). Quantile regression. *Cambridge University Press.*

Li, H. and Wang, R. (2023). PELVE: Probability Equivalent Level of VaR and ES. *J. Econometrics* **234**, 353-370.

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

Wang, H.J., and Li, D. (2013). Estimation of extreme conditional quantiles through power transformation. *J. Am. Stat. Assoc.* **108**, 1062-1074.

Wang, H.J., Li, D., and He, X. (2012). Estimation of high conditional quantiles for heavy-tailed distributions. *J. Am. Stat. Assoc.* **107**, 1453-1464.

Xu, W., Hou, Y., and Li, D. (2022). Prediction of extremal expectile based on regression models with heteroscedastic extremes. *J. Bus. Econ. Stat.* **40**, 522-536.

Zhou, Z., and Shao, X. (2013). Inference for linear models with dependent errors. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **75**, 323-343.

School of Data Science, Fudan University, Shanghai, China

E-mail: qzzhong22@m.fudan.edu.cn

School of Statistics and Data Science, Capital University of Economics and Business, Beijing, China

E-mail: jingyuji@cueb.edu.cn

School of Management, University of Science and Technology of China, Hefei, China

E-mail: ljchen22@ustc.edu.cn

School of Data Science, Fudan University, Shanghai, China

E-mail: yxhou@fudan.edu.cn

School of Management, Fudan University, Shanghai, China

E-mail: deyuanli@fudan.edu.cn