

# Tail risk inference via expectiles in heavy-tailed time series

A. C. Davison

Institute of Mathematics, Ecole Polytechnique Fédérale de Lausanne

S. A. Padoan

Department of Decision Sciences, Bocconi University of Milan

and

G. Stupfler

Université de Rennes, Ensai, CNRS, CREST

October 13, 2021

## Abstract

Expectiles define the only law-invariant, coherent and elicitable risk measure apart from the expectation. The popularity of expectile-based risk measures is steadily growing and their properties have been studied for independent data, but further results are needed to use extreme expectiles with dependent time series such as financial data. In this paper we establish a basis for inference on extreme expectiles and expectile-based marginal expected shortfall in a general  $\beta$ -mixing context that encompasses ARMA, ARCH and GARCH models with heavy-tailed innovations. Simulations and applications to financial returns show that the new estimators and confidence intervals greatly improve on existing ones when the data are dependent.

*Keywords:* Asymmetric least squares estimation, Marginal Expected Shortfall, Mixing, Tail copula, Weak dependence.

# 1 Introduction

The quantification of the risk associated to a real-valued profit-loss variable  $Y$  is a key problem in financial econometrics. The best-known financial risk measure, the Value-at-Risk, has serious deficiencies, as it does not generally define a coherent risk measure and thus does not obey the diversification principle ([Artzner et al. 1999](#)), and it depends only on the frequency of tail losses and not on their actual values. A related measure, the expected shortfall, is coherent and takes into account the values of the risk variable, but it is not elicitable ([Gneiting 2011](#)) and thus does not benefit from the existence of a natural backtesting methodology.

A risk measure that addresses these issues is based on the expectile ([Newey & Powell 1987](#)), which extends the mean in the same way that a quantile extends the median, and is defined by minimising an asymmetrically weighted mean squared deviation,

$$\xi_\tau = \arg \min_{\theta \in \mathbb{R}} \mathbb{E}\{\eta_\tau(Y - \theta) - \eta_\tau(Y)\}, \quad (1)$$

where  $Y$  is a real-valued random variable,  $\mathbb{E}$  denotes expectation,  $\eta_\tau(u) = |\mathbb{1}(u \leq 0) - \tau|u^2$  with  $\tau \in (0, 1)$  determining the level of the expectile, and  $\mathbb{1}(\cdot)$  denotes the indicator function. The expectile is related to the gain-loss ratio, which is a popular performance measure in portfolio management (see [Bellini & Di Bernardino 2017](#), and references therein), and it depends on both the realisations of  $Y$  and on their probabilities. Moreover, apart from the expectation it is the only risk measure that is law-invariant, coherent and elicitable. Results of [Bellini & Di Bernardino \(2017\)](#), [Ehm et al. \(2016\)](#) and others indicate that expectiles define sensible alternatives to the Value-at-Risk and expected shortfall, and inference for them has recently burgeoned; see [Holzmann & Klar \(2016\)](#) and [Krättschmer & Zähle \(2017\)](#), for example. The estimation of extreme expectiles, for which  $\tau \approx 1$ , was considered by [Daouia et al. \(2018, 2020\)](#), but for independent and identically distributed data. Financial data are typically dependent, however, so these existing methods cannot safely be applied. Available techniques for expectile estimation with dependent data, such as those in [Kuan et al. \(2009\)](#), [Xie et al. \(2014\)](#) and [Jiang et al. \(2021\)](#), are restricted to central expectiles whose level stays well away from 0 and 1, so extreme expectile estimation for dependent data requires new tools.

In this paper we consider inference for extreme expectile-based risk measures based on  $\beta$ -mixing heavy-tailed strictly stationary time series, which encompass some widely-used time series models. To the best of our knowledge, we are the first to develop inference for extreme marginal expectiles in a time series context. We start by considering the estimation of a tail expectile at the intermediate order, for which  $\tau_n \rightarrow 1$  such that  $n(1 - \tau_n) \rightarrow \infty$  as  $n \rightarrow \infty$ . Although a large expectile,  $\xi_{\tau_n}$  is expected to typically fall within the range of the data. We then discuss the prediction of a tail expectile of order  $\tau_n = 1 - p_n \rightarrow 1$  as the sample size  $n \rightarrow \infty$  such that  $np_n$  is bounded; typically  $p_n \leq 1/n$ , and then the expectile is expected to fall in a neighbourhood of or above the largest observations available. We consider the least asymmetrically weighted squares estimator, which is the empirical counterpart of (1), and a quantile-based estimator obtained from an asymptotic relationship between high expectiles and high quantiles. We then expand our statistical model to bivariate  $\beta$ -mixing time series  $\{(X_t, Y_t), t = 1, 2, \dots\}$  and establish estimation results for the marginal expected shortfall, which is important in assessing systemic risk (Acharya et al. 2017, Brownlees & Engle 2017). In this context,  $X = X_1$  and  $Y = Y_1$  represent the marginal losses on the returns of a company and of the entire market, and the marginal expected shortfall (MES) is  $\mathbb{E}(X \mid Y > u)$ , where  $u$  is large to reflect a substantial market downturn. Our results apply both to the expectile-based form of the extreme marginal expected shortfall, where  $u$  is taken to be a high expectile of  $Y$ , and to its quantile-based counterpart. We discuss the construction of confidence intervals for extreme expectile-based risk measures and compare them with intervals obtained by assuming independent observations.

Our methods and data have been incorporated into the R package **ExtremeRisks**, freely available on CRAN. The online Supplementary Material contains technical results, further discussion and additional numerical work. The layout of the paper is the following. Section 2 explains in detail our statistical context. Section 3 considers intermediate and extreme expectile estimation. Section 4 introduces a more general bivariate time series context within which general MES estimators at extreme levels are investigated. Section 5 provides a finite sample procedure for constructing confidence intervals and discusses the selection of the expectile level in practice. The finite-sample performance of the methods is examined on simulated data in Section 6 and on real financial data in Section 7.

## 2 Statistical model and time series framework

Let  $(Y_t, t = 1, 2, \dots)$  be a strictly stationary time series whose continuous marginal distribution  $F$  has survival function  $\bar{F} = 1 - F$  and tail quantile function  $U : s \mapsto \inf\{y \in \mathbb{R} : 1/\bar{F}(y) \geq s\}$ . Throughout the paper,  $Y$  should be seen as the negative of a generic financial position, so large positive values of  $Y$  represent extreme losses associated to this position. Our goal is to estimate extreme expectiles of  $Y$ , so we focus on heavy-tailed distributions, which are ubiquitous in the modelling of extreme actuarial and financial losses ([Embrechts et al. 1997](#)). We therefore assume for any  $y > 0$  that as  $s \rightarrow \infty$ ,  $\bar{F}(sy)/\bar{F}(s) \rightarrow y^{-1/\gamma}$  or equivalently  $U(sy)/U(s) \rightarrow y^\gamma$ . The tail index  $\gamma > 0$  specifies the tail weight of  $\bar{F}$ , whose tail becomes heavier as  $\gamma$  increases. Together with the condition  $\mathbb{E}(|Y_-|) < \infty$ , where  $Y_- = \min(Y, 0)$ , the assumption  $\gamma < 1$  ensures that the first moment of  $Y$  exists, and then the expectile  $\xi_\tau$  is well-defined by [\(1\)](#) for any  $\tau$ .

Existing results on extreme expectile estimation require independent data ([Daouia et al. 2018](#)). Here we instead suppose that the time series is  $\beta$ -mixing. For any  $m = 1, 2, \dots$ , let  $\mathcal{F}_{1,m} = \sigma(Y_1, \dots, Y_m)$  and  $\mathcal{F}_{m,\infty} = \sigma(Y_m, Y_{m+1}, \dots)$  denote the past and future  $\sigma$ -fields generated by  $(Y_t)$ . Then  $(Y_t)$  is said to be  $\beta$ -mixing if the coefficients  $\beta(l) = \sup_{m \geq 1} \beta_m(l)$  satisfy  $\beta(l) \rightarrow 0$  as  $l \rightarrow \infty$ , where

$$\beta_m(l) = \mathbb{E}[\sup\{|\mathbb{P}(B \mid \mathcal{F}_{1,m}) - \mathbb{P}(B)| : B \in \mathcal{F}_{m+l,\infty}\}], \quad l = 1, 2, \dots$$

[Doukhan \(1994, Section 2.4\)](#) shows that many Markov processes, including ARMA processes, nonlinear autoregressive processes and GARCH models, are geometrically  $\beta$ -mixing: under mild conditions there exists  $a \in (0, 1)$  such that  $\beta(l) \leq a^l$  for large enough  $l$ . The good balance between theoretical applicability and modelling strength offered by  $\beta$ -mixing motivates the following basic assumption.

**Condition A.** The time series  $(Y_t, t = 1, 2, \dots)$  is strictly stationary and  $\beta$ -mixing and its one-dimensional marginal distribution function  $F$  is continuous and heavy-tailed.

We henceforth suppose that our observations  $Y_1, \dots, Y_n$  are from a time series satisfying Condition [A](#) and aim to estimate an unconditional marginal extreme expectile  $\xi_{\tau'_n}$  of  $Y = Y_1$ , where  $\tau'_n \rightarrow 1$  as  $n \rightarrow \infty$ . The initial step is to estimate extreme but intermediate, i.e., in-sample, expectiles  $\xi_{\tau_n}$ , which we now discuss.

### 3 Extreme expectile estimation in time series

#### 3.1 Background

Let  $\tau_n$  represent a sequence of probabilities such that  $\tau_n \rightarrow 1$  and  $n(1 - \tau_n) \rightarrow \infty$  as  $n \rightarrow \infty$ . A natural estimator of the expectile  $\xi_{\tau_n}$  of the marginal distribution  $F$  is the empirical counterpart of (1), i.e.,

$$\tilde{\xi}_{\tau_n} = \arg \min_{\theta \in \mathbb{R}} \sum_{t=1}^n \eta_{\tau_n}(Y_t - \theta);$$

this is readily computed, for example by iterative reweighted least squares. An alternative estimator exploits an asymptotic proportionality relationship between the high expectile  $\xi_\tau$  and the corresponding quantile  $q_\tau$ : within our heavy-tailed model (Bellini et al. 2014),

$$\lim_{\tau \rightarrow 1} \frac{\xi_\tau}{q_\tau} = (\gamma^{-1} - 1)^{-\gamma}, \quad \gamma < 1. \quad (2)$$

Thus a quantile-based estimator of  $\xi_{\tau_n}$  is  $\hat{\xi}_{\tau_n} = (\hat{\gamma}_n^{-1} - 1)^{-\hat{\gamma}_n} \hat{q}_{\tau_n}$ , where  $\hat{q}_{\tau_n} = Y_{n - \lfloor n(1 - \tau_n) \rfloor, n}$  is the empirical counterpart of  $q_{\tau_n}$ ,  $Y_{1,n} \leq \dots \leq Y_{n,n}$  are the ascending order statistics of  $Y_1, \dots, Y_n$ , and  $\hat{\gamma}_n$  is a consistent estimator of  $\gamma$ . Here  $\lfloor x \rfloor$  is the floor function, that is, the largest integer smaller than or equal to  $x$ . We focus on the following two simple examples. The first, the Hill (1975) estimator,

$$\hat{\gamma}_n^H = \frac{1}{\lfloor n(1 - \tau_n) \rfloor} \sum_{i=1}^{\lfloor n(1 - \tau_n) \rfloor} \log \left( \frac{Y_{n-i+1,n}}{Y_{n - \lfloor n(1 - \tau_n) \rfloor, n}} \right),$$

is the maximum likelihood estimator in a purely Pareto model, and thus is a natural estimator of  $\gamma$ . The second estimator exploits the relationship (2), which can be rephrased as  $\bar{F}(\xi_\tau)/(1 - \tau) \rightarrow \gamma^{-1} - 1$  as  $\tau \uparrow 1$ , which implies that

$$\gamma = \lim_{\tau \uparrow 1} \left( 1 + \frac{\bar{F}(\xi_\tau)}{1 - \tau} \right)^{-1}.$$

Taking  $\tau = \tau_n \rightarrow 1$ , and estimating  $\bar{F}(\xi_\tau)$  by  $\hat{\bar{F}}_n(\tilde{\xi}_{\tau_n})$ , where  $\hat{\bar{F}}_n(u) = n^{-1} \sum_{t=1}^n \mathbb{1}\{Y_t > u\}$  is the empirical survival function, suggests the expectile-based (EB) estimator

$$\hat{\gamma}_n^E = \left( 1 + \frac{\hat{\bar{F}}_n(\tilde{\xi}_{\tau_n})}{1 - \tau_n} \right)^{-1}.$$

Our first main results rely on the following conditions concerning dependence within the time series  $(Y_t)$  and the gap between the right tail of  $\bar{F}$  and a purely Pareto tail.

**Condition B.** The time series  $(Y_t, t = 1, 2, \dots)$  has the following properties:

(i) there exist sequences of integers  $(l_n)$  and  $(r_n)$  such that  $l_n \rightarrow \infty$ ,  $r_n \rightarrow \infty$ ,  $l_n/r_n \rightarrow 0$ ,  $r_n/n \rightarrow 0$  and  $n\beta(l_n)/r_n \rightarrow 0$  as  $n \rightarrow \infty$ ;

(ii) for any  $t = 1, 2, \dots$  there exists a function  $R_t$  on  $\mathcal{E}_2 = [0, \infty]^2 \setminus \{(\infty, \infty)\}$  such that

$$\lim_{s \rightarrow \infty} s \mathbb{P} \left\{ \overline{F}(Y_1) \leq \frac{x}{s}, \overline{F}(Y_{t+1}) \leq \frac{y}{s} \right\} = R_t(x, y), \quad (x, y) \in \mathcal{E}_2;$$

(iii) there exist  $D \geq 0$  and a nonnegative sequence  $\rho(t)$  satisfying  $\sum_{t=1}^{\infty} \rho(t) < \infty$  and such that for  $s$  large enough, any  $t = 1, 2, \dots$ , and all  $u, u', v, v' \in [0, 1]$  such that  $u' < u$  and  $v' < v$ ,

$$s \mathbb{P} \left\{ \frac{u'}{s} < \overline{F}(Y_1) \leq \frac{u}{s}, \frac{v'}{s} < \overline{F}(Y_{t+1}) \leq \frac{v}{s} \right\} \leq \rho(t) \{(u - u')(v - v')\}^{1/2} + \frac{D}{s} (u - u')(v - v').$$

Conditions **B(i)** and **B(ii)** have been employed previously in extreme value analysis with mixing conditions (e.g., [Drees 2003](#), [Drees & Rootzén 2010](#)). The sequences  $(l_n)$  and  $(r_n)$  in Condition **B(i)** are small-block and big-block sequences used for standard arguments in the literature on mixing time series. Condition **B(iii)** is slightly more precise than condition (C3) in [Drees \(2003\)](#). Condition **B** is discussed further in the Supplementary Material.

**Condition C.** There exist  $\rho \leq 0$  and a measurable function  $A(\cdot)$  having constant sign and converging to zero at infinity such that

$$\lim_{s \rightarrow \infty} \frac{1}{A\{1/\overline{F}(s)\}} \left\{ \frac{\overline{F}(sy)}{\overline{F}(s)} - y^{-1/\gamma} \right\} = y^{-1/\gamma} \frac{y^{\rho/\gamma} - 1}{\gamma\rho}, \quad y > 0;$$

when  $\rho = 0$ , the right-hand side should be read as  $\gamma^{-2} y^{-1/\gamma} \log y$ .

[Beirlant et al. \(2004\)](#) and [de Haan & Ferreira \(2006\)](#) give numerous examples of continuous distributions satisfying Condition **C**. We start by a general result on the least asymmetrically weighted squares estimator  $\tilde{\xi}_{\tau_n}$ .

**Theorem 3.1.** *Assume that Conditions **A** and **B** are satisfied. Assume further that there exists a  $\delta > 0$  such that  $\mathbb{E}|Y_-|^{2+\delta} < \infty$ ,  $0 < \gamma < 1/(2 + \delta)$  and  $\sum_{l \geq 1} [\beta(l)]^{\delta/(2+\delta)} < \infty$ .*

Let  $\tau_n \uparrow 1$  be such that  $n(1 - \tau_n) \rightarrow \infty$ ,  $r_n(1 - \tau_n) \rightarrow 0$  and  $r_n\{r_n/\sqrt{n(1 - \tau_n)}\}^\delta \rightarrow 0$  as  $n \rightarrow \infty$ . Then,

$$\sqrt{n(1 - \tau_n)} \left( \frac{\tilde{\xi}_{\tau_n}}{\xi_{\tau_n}} - 1 \right) \xrightarrow{d} \mathcal{N} \left( 0, \frac{2\gamma^3}{1 - 2\gamma} (1 + \sigma^2(\gamma, \mathbf{R})) \right),$$

$$\text{with } \sigma^2(\gamma, \mathbf{R}) := \frac{(1 - \gamma)(1 - 2\gamma)}{\gamma^2} \iint_{[1, \infty)^2} \sum_{t=1}^{\infty} R_t(x^{-1/\gamma}, y^{-1/\gamma}) dx dy.$$

The family of functions  $R_t$  specifies the extremal dependence within the time series between different time points; when  $R_t \equiv 0$  for any  $t \geq 1$ , which is for instance the case when  $(Y_t)$  is an i.i.d. sequence, the asymptotic variance is  $2\gamma^3/(1 - 2\gamma)$ . The quantity  $\sigma^2(\gamma, \mathbf{R})$  represents the proportion of increase of this asymptotic variance due to the mixing setting. The conditions  $\mathbb{E}|Y_-|^{2+\delta} < \infty$  and  $0 < \gamma < 1/(2 + \delta)$  essentially amount to assuming that the time series  $Y_t$  has a finite variance. Theorem 3.1 above represents a substantial theoretical step compared to available results on extreme expectile estimation, because its proof is based on rather delicate arguments involving a tailored central limit theory for tail array sums in the time-dependent setting. Our assumptions are somewhat involved, but are very mild when  $\beta(l)$  converges to 0 geometrically fast as  $l \rightarrow \infty$ . In that case, condition  $\sum_{l \geq 1} [\beta(l)]^{\delta/(2+\delta)} < \infty$  is satisfied for any  $\delta > 0$ , and one may choose  $l_n = \lfloor C \log n \rfloor$ ,  $r_n = \lfloor \log^2(n) \rfloor$  and  $\tau_n = 1 - n^{-\tau}$ , for any  $\tau \in (0, 1)$  and sufficiently large  $C$ . As we highlighted in Section 2, geometrically strong  $\beta$ -mixing covers many widely-used financial time series models, including ARMA processes, ARCH/GARCH processes and solutions of stochastic difference equations (see Doukhan 1994, Drees 2000, 2003, Francq & Zakoïan 2006, Boussama et al. 2011). The following corollary provides a rigorous statement in this geometrically mixing case.

**Corollary 3.2.** *Assume that Conditions A and B(ii)-(iii) are satisfied, and that  $\beta(l) = O(a^l)$  for some  $a \in (0, 1)$ . Assume further that there exists  $\delta > 0$  such that  $\mathbb{E}|Y|^{2+\delta} < \infty$ . Let  $\tau_n = 1 - n^{-\tau}$ , for some  $\tau \in (0, 1)$ . Then as  $n \rightarrow \infty$ ,*

$$\sqrt{n(1 - \tau_n)} \left( \frac{\tilde{\xi}_{\tau_n}}{\xi_{\tau_n}} - 1 \right) \xrightarrow{d} \mathcal{N} \left( 0, \frac{2\gamma^3}{1 - 2\gamma} (1 + \sigma^2(\gamma, \mathbf{R})) \right),$$

with the notation of Theorem 3.1.

We turn to the analysis of the estimators  $\widehat{\xi}_{\tau_n}$  based on  $\widehat{\gamma}_n^H$  and  $\widehat{\gamma}_n^E$ . The crucial result in this case, which is of interest in its own right, consists in a joint Gaussian approximation of the processes  $s \mapsto \widehat{q}_{1-(1-\tau_n)s} = Y_{n-\lfloor n(1-\tau_n)s \rfloor, n}$  and  $s \mapsto \log \widehat{q}_{1-(1-\tau_n)s}$  (that is, the tail empirical quantile process and its logarithm) in our mixing framework. See also Theorem 2.4.8 in [de Haan & Ferreira \(2006\)](#) for a result restricted to the independent case, as well as Proposition A.1 in [de Haan et al. \(2016\)](#) and Proposition 1 in [Chavez-Demoulin & Guillou \(2018\)](#) for related but different statements that do not apply in our setup.

**Theorem 3.3.** *Assume that Conditions [A](#), [B](#) and [C](#) are satisfied, and that  $\tau_n \uparrow 1$ ,  $n(1 - \tau_n) \rightarrow \infty$ ,  $r_n(1 - \tau_n) \rightarrow 0$ ,  $r_n \log^2(n(1 - \tau_n))/\sqrt{n(1 - \tau_n)} \rightarrow 0$  and  $\sqrt{n(1 - \tau_n)}A((1 - \tau_n)^{-1}) = O(1)$  as  $n \rightarrow \infty$ . Suppose finally that  $n(1 - \tau_n)$  is a sequence of integers and pick  $s_0 > 0$ . Then there exist appropriate versions of the process  $s \mapsto \widehat{q}_{1-(1-\tau_n)s}$  and a continuous, centred Gaussian process  $W$  having covariance function*

$$r(x, y) := \min(x, y) + \sum_{t=1}^{\infty} R_t(x, y) + R_t(y, x)$$

such that, for any  $\varepsilon > 0$  sufficiently small, we have, uniformly in  $s \in (0, s_0]$ ,

$$\frac{\widehat{q}_{1-(1-\tau_n)s}}{q_{\tau_n}} = s^{-\gamma} \left( 1 + \frac{1}{\sqrt{n(1 - \tau_n)}} \gamma s^{-1} W(s) + \frac{s^{-\rho} - 1}{\rho} A((1 - \tau_n)^{-1}) + o_{\mathbb{P}} \left( \frac{s^{-1/2-\varepsilon}}{\sqrt{n(1 - \tau_n)}} \right) \right)$$

and

$$\log \frac{\widehat{q}_{1-(1-\tau_n)s}}{q_{\tau_n}} = -\gamma \log s + \frac{1}{\sqrt{n(1 - \tau_n)}} \gamma s^{-1} W(s) + \frac{s^{-\rho} - 1}{\rho} A((1 - \tau_n)^{-1}) + o_{\mathbb{P}} \left( \frac{s^{-1/2-\varepsilon}}{\sqrt{n(1 - \tau_n)}} \right).$$

Condition [B](#) is assumed in Theorem [3.3](#) for consistency with our framework; an inspection of the proof shows that Condition [B\(iii\)](#) can be replaced by its version with  $u = u' = 0$ . As a consequence of Theorem [3.3](#), one may determine the asymptotic behaviour of the pair  $(\widehat{\gamma}_n^H, \widehat{q}_{\tau_n})$ , which we then use in the following corollary to establish the limiting distribution of the estimator  $\widehat{\xi}_{\tau_n}$  constructed using  $\widehat{\gamma}_n^H$  as the tail index estimator.

**Corollary 3.4.** *Assume that Conditions [A](#), [B](#) and [C](#) are satisfied, with  $\mathbb{E}|Y_-| < \infty$  and  $0 < \gamma < 1$ . Let  $\tau_n \uparrow 1$  be such that  $n(1 - \tau_n) \rightarrow \infty$ ,  $r_n(1 - \tau_n) \rightarrow 0$ ,  $r_n \log^2(n(1 - \tau_n))/\sqrt{n(1 - \tau_n)} \rightarrow 0$ ,  $\sqrt{n(1 - \tau_n)}A((1 - \tau_n)^{-1}) \rightarrow \lambda_1 \in \mathbb{R}$  and  $\sqrt{n(1 - \tau_n)}q_{\tau_n}^{-1} \rightarrow \lambda_2 \in \mathbb{R}$  as  $n \rightarrow \infty$ . Then, for  $\widehat{\gamma}_n = \widehat{\gamma}_n^H$  in the estimator  $\widehat{\xi}_{\tau_n}$ , one has*

$$\sqrt{n(1 - \tau_n)} \left( \frac{\widehat{\xi}_{\tau_n}}{\xi_{\tau_n}} - 1 \right) \xrightarrow{d} \mathcal{N} \left( \frac{m(\gamma)}{1 - \rho} \lambda_1 - \lambda, \gamma^2 v^H(\gamma, \mathbf{R}) \right)$$



with

$$\begin{aligned} \lambda &:= \left( \frac{(\gamma^{-1} - 1)^{-\rho}}{1 - \gamma - \rho} + \frac{(\gamma^{-1} - 1)^{-\rho} - 1}{\rho} \right) \lambda_1 + \gamma(\gamma^{-1} - 1)^\gamma \mathbb{E}(Y) \lambda_2 \\ \text{and } v^H(\gamma, \mathbf{R}) &:= (1 + [m(\gamma)]^2) \left( 1 + 2 \sum_{t=1}^{\infty} R_t(1, 1) \right) \\ &\quad + 2m(\gamma) \int_0^1 \sum_{t=1}^{\infty} \left[ \frac{R_t(s, 1) + R_t(1, s)}{s} - 2R_t(1, 1) \right] ds. \end{aligned}$$

Corollary 3.4 is, to the best of our knowledge, the first result on the estimator  $\widehat{\xi}_{\tau_n}$  at intermediate levels under weak dependence assumptions (i.e., in the  $\beta$ -mixing framework). The asymptotic properties of  $\widehat{\xi}_{\tau_n}$  using the estimator  $\widehat{\gamma}_n^E$  are given in Theorem C.5 in the Supplementary Material, where it can be seen that, due to asymptotic variance considerations, this estimator will tend to be less variable than the Hill estimator (in the i.i.d. case, when  $\gamma < 0.38$ ). This may make it valuable in constructing confidence intervals requiring an estimate of  $\gamma$ , as we illustrate in Sections B.1 and B.2 of the Supplementary Material when constructing confidence intervals for intermediate expectiles.

## 3.2 Extrapolation

We now consider estimating extreme expectiles  $\xi_{\tau'_n}$  whose level  $\tau'_n \rightarrow 1$  satisfies  $n(1 - \tau'_n) \rightarrow c \in [0, \infty)$  as  $n \rightarrow \infty$ . A typical choice in applications is  $\tau'_n = 1 - p_n$  for an exceedance probability  $p_n$  not greater than  $1/n$  (e.g., Cai et al. 2015). Our semiparametric approach is motivated by a combination of the heavy-tailed assumption with Equation (2) and defines an extreme expectile estimator through a Weissman (1978)-type construction, whereby

$$\frac{\xi_{\tau'_n}}{\xi_{\tau_n}} \approx \frac{q_{\tau'_n}}{q_{\tau_n}} = \frac{U\{(1 - \tau'_n)^{-1}\}}{U\{(1 - \tau_n)^{-1}\}} \approx \left( \frac{1 - \tau'_n}{1 - \tau_n} \right)^{-\gamma}, \quad (3)$$

for  $n$  large. This suggests the following class of plug-in estimators of  $\xi_{\tau'_n}$ ,

$$\bar{\xi}_{\tau'_n}^* \equiv \bar{\xi}_{\tau'_n}^*(\tau_n) = \left( \frac{1 - \tau'_n}{1 - \tau_n} \right)^{-\widehat{\gamma}_n} \bar{\xi}_{\tau_n},$$

where  $\widehat{\gamma}_n$  and  $\bar{\xi}_{\tau_n}$  are consistent estimators of  $\gamma$  and of the expectile  $\xi_{\tau_n}$ . We call  $\bar{\xi}_{\tau'_n}^*$  the extrapolating least asymmetrically weighted squares (LAWS) estimator when  $\bar{\xi}_{\tau_n} = \widetilde{\xi}_{\tau_n}$ , then denoting it by  $\widetilde{\xi}_{\tau'_n}^*$ , and call it the extrapolating quantile-based (QB) estimator when  $\bar{\xi}_{\tau_n} = \widehat{\xi}_{\tau_n}$ , then denoting it by  $\widehat{\xi}_{\tau'_n}^*$ . We can then prove the following result.

**Theorem 3.5.** Suppose that  $\mathbb{E}|Y_-| < \infty$  and that Conditions [A](#), [B](#) and [C](#) are satisfied with  $0 < \gamma < 1$  and  $\rho < 0$ . Let  $\tau_n, \tau'_n \uparrow 1$  with  $n(1 - \tau_n) \rightarrow \infty$ ,  $n(1 - \tau'_n) \rightarrow c \in [0, \infty)$  and  $\{n(1 - \tau_n)\}^{1/2} / \log\{(1 - \tau_n)/(1 - \tau'_n)\} \rightarrow \infty$  as  $n \rightarrow \infty$ . Assume also that  $r_n(1 - \tau_n) \rightarrow 0$ ,  $r_n\{n(1 - \tau_n)\}^{-1/2} \log^2\{n(1 - \tau_n)\} \rightarrow 0$ ,  $\{n(1 - \tau_n)\}^{1/2} A\{(1 - \tau_n)^{-1}\} \rightarrow \lambda_1 \in \mathbb{R}$  and  $\{n(1 - \tau_n)\}^{1/2} q_{\tau_n}^{-1} \rightarrow \lambda_2 \in \mathbb{R}$  as  $n \rightarrow \infty$ . Suppose finally that as  $n \rightarrow \infty$ ,

$$\{n(1 - \tau_n)\}^{1/2} \left( \frac{\bar{\xi}_{\tau_n}}{\xi_{\tau_n}} - 1 \right) \rightarrow \Delta$$

weakly, where  $\Delta$  is a nondegenerate random variable. Then, for  $\hat{\gamma}_n = \hat{\gamma}_n^H$ ,

$$\frac{\{n(1 - \tau_n)\}^{1/2}}{\log\{(1 - \tau_n)/(1 - \tau'_n)\}} \left( \frac{\bar{\xi}_{\tau'_n}^*}{\xi_{\tau'_n}} - 1 \right) \rightarrow \mathcal{N} \left[ \frac{\lambda_1}{1 - \rho}, \gamma^2 \left\{ 1 + 2 \sum_{t=1}^{\infty} R_t(1, 1) \right\} \right]$$

weakly as  $n \rightarrow \infty$ .

Theorem [3.5](#) enables the construction of confidence intervals for extreme expectiles. The quantity  $2 \sum_{t=1}^{\infty} R_t(1, 1)$  represents the relative increase in the asymptotic variance due to the temporal dependence. When  $R_t \equiv 0$  for any  $t \geq 1$ , the asymptotic variance is  $\gamma^2$ . An analogous result may be obtained for the estimator  $\hat{\gamma}_n^E$ , although we shall not pursue this for the sake of brevity.

## 4 Marginal expected shortfall estimation

It is important to assess overall, or systemic, risk when working with actuarial and financial data, for instance by simultaneously considering several lines of business of an insurance company or several stock market indices. A prominent way to measure such risk is via the marginal expected shortfall, defined as the propensity of a financial institution to be undercapitalised when the financial system as a whole is undercapitalised ([Acharya et al. 2017](#), [Brownlees & Engle 2017](#)). The contribution that an individual firm with loss return  $X$  makes to systemic risk, represented by a loss  $Y$  in the aggregated return of the market, can be measured by the quantile-based marginal expected shortfall

$$\text{QMES}_{X,\tau} = \mathbb{E}(X \mid Y > q_{Y,\tau}), \quad \tau \in (0, 1), \quad (4)$$

where  $q_{Y,\tau}$  is the  $\tau$  quantile of  $Y$ . A systemic crisis typically corresponds to a situation in which  $\tau$  is close to or exceeds  $1 - 1/n$ , where  $n$  is the historical sample size. An alternative to (4) is the expectile-based marginal expected shortfall,

$$\text{XMES}_{X,\tau} = \mathbb{E}(X \mid Y > \xi_{Y,\tau}), \quad \tau \in (0, 1), \quad (5)$$

with  $\xi_{Y,\tau}$  the  $\tau$  expectile of  $Y$ . Estimation of  $\text{QMES}_{X,\tau}$  and  $\text{XMES}_{X,\tau}$  at extreme levels is considered by [Cai et al. \(2015\)](#) and [Daouia et al. \(2018\)](#), but only for independent data.

We now extend the results of Section 3 to inference for these definitions of marginal expected shortfall at extreme levels in our weakly-dependent setting. Suppose that the data come from a strictly stationary bivariate time series  $\{(X_t, Y_t), t = 1, 2, \dots\}$ ; for instance,  $X_t$  and  $Y_t$  could be the daily loss returns on a specific stock and on a market index. For any  $m \geq 1$ , let  $\mathcal{F}_{1,m} = \sigma(X_1, Y_1, \dots, X_m, Y_m)$  and  $\mathcal{F}_{m,\infty} = \sigma(X_m, Y_m, X_{m+1}, Y_{m+1}, \dots)$  denote the past and future  $\sigma$ -fields generated by  $(X_t, Y_t)$ . Then the corresponding  $\beta$ -mixing coefficients can be defined as  $b(l) = \sup_{m \geq 1} b_m(l)$ , where

$$b_m(l) = \mathbb{E}[\sup\{|\mathbb{P}(B \mid \mathcal{F}_{1,m}) - \mathbb{P}(B)| : B \in \mathcal{F}_{m+l,\infty}\}], \quad l = 1, 2, \dots$$

The sequence  $\{(X_t, Y_t), t = 1, 2, \dots\}$  is then said to be  $\beta$ -mixing if  $b(l) \rightarrow 0$  as  $l \rightarrow \infty$ . If it is  $\beta$ -mixing, then  $(Y_t, t = 1, 2, \dots)$  is also  $\beta$ -mixing in the sense of Section 2. Our condition below similarly extends Conditions A and B in a natural way.

- Condition D.** (i) The time series  $\{(X_t, Y_t), t = 1, 2, \dots\}$  is strictly stationary,  $\beta$ -mixing and the one-dimensional marginal distribution functions  $F_X$  and  $F_Y$  of  $(X_t, t = 1, 2, \dots)$  and  $(Y_t, t = 1, 2, \dots)$  are continuous and heavy-tailed with respective tail indices  $\gamma_X$  and  $\gamma_Y$ ;
- (ii) there exist sequences of integers  $(l_n)$  and  $(r_n)$  such that  $l_n \rightarrow \infty$ ,  $r_n \rightarrow \infty$ ,  $l_n/r_n \rightarrow 0$ ,  $r_n/n \rightarrow 0$  and  $nb(l_n)/r_n \rightarrow 0$  as  $n \rightarrow \infty$ ;
- (iii) for any  $t = 1, 2, \dots$  there exists a function  $r_t$  on  $\mathcal{E}_4 = [0, \infty]^4 \setminus \{(\infty, \infty, \infty, \infty)\}$  such that

$$\lim_{s \rightarrow \infty} s \mathbb{P} \left\{ \overline{F}_X(X_1) \leq \frac{x_1}{s}, \overline{F}_X(X_{t+1}) \leq \frac{x_{t+1}}{s}, \overline{F}_Y(Y_1) \leq \frac{y_1}{s}, \overline{F}_Y(Y_{t+1}) \leq \frac{y_{t+1}}{s} \right\}$$

equals  $r_t(x_1, x_{t+1}, y_1, y_{t+1})$  for any  $(x_1, x_{t+1}, y_1, y_{t+1}) \in \mathcal{E}_4$ ; and

- (iv) there exist  $D \geq 0$  and a nonnegative sequence  $\rho(t)$  satisfying  $\sum_{t=1}^{\infty} \rho(t) < \infty$  such that if  $s$  is large enough, for any  $t = 1, 2, \dots$ , all  $x_1, x_{t+1} \in [0, \infty]$ , and all  $u, u', v, v' \in [0, 1]$  with  $u' < u$  and  $v' < v$ ,

$$s \mathbb{P} \left\{ \bar{F}_X(X_1) \leq \frac{x_1}{s}, \bar{F}_X(X_{t+1}) \leq \frac{x_{t+1}}{s}, \frac{u'}{s} < \bar{F}_Y(Y_1) \leq \frac{u}{s}, \frac{v'}{s} < \bar{F}_Y(Y_{t+1}) \leq \frac{v}{s} \right\}$$

is no greater than

$$\rho(t) \{ \min(x_1, u - u') \min(x_{t+1}, v - v') \}^{1/2} + \frac{D}{s} \min(x_1, u - u') \min(x_{t+1}, v - v').$$

If  $\{(X_t, Y_t), t = 1, 2, \dots\}$  satisfies Condition **D**, then the univariate time series  $(Y_t, t = 1, 2, \dots)$  automatically satisfies Conditions **A** and **B**, since if we set  $x_1 = x_{t+1} = \infty$  then the function  $R_t \equiv R_{Y,t}$  is given by  $R_{Y,t}(y_1, y_{t+1}) = r_t(\infty, \infty, y_1, y_{t+1})$ .

We now embed (4) and (5) in a more general marginal expected shortfall framework. Define  $\text{MES}_{X,\tau} = \mathbb{E}(X \mid Y > z_{Y,\tau})$  for  $\tau \in (0, 1)$ , where  $z_{Y,\tau}$  is a risk measure on  $Y$  such that  $\bar{F}(z_{Y,\tau})/(1 - \tau) \rightarrow z = z(\gamma_Y) \in (0, \infty)$  as  $\tau \uparrow 1$ . If an  $\{n(1 - \tau_n)\}^{1/2}$ -relatively consistent estimator  $\bar{z}_{Y,\tau_n}$  of  $z_{Y,\tau_n}$  is available at the intermediate level  $\tau_n$ , then an extrapolation relationship similar to (3) suggests defining an estimator of  $\text{MES}_{X,\tau'_n} = \mathbb{E}(X \mid Y > z_{Y,\tau'_n})$  as

$$\overline{\text{MES}}_{X,\tau'_n}^* \equiv \overline{\text{MES}}_{X,\tau'_n}^*(\tau_n) = \left( \frac{1 - \tau'_n}{1 - \tau_n} \right)^{-\hat{\gamma}_{X,n}} \frac{\sum_{t=1}^n X_t \mathbb{1}(X_t > 0, Y_t > \bar{z}_{Y,\tau_n})}{\sum_{t=1}^n \mathbb{1}(Y_t > \bar{z}_{Y,\tau_n})}.$$

Replacing  $\bar{z}_{Y,\tau_n}$  by  $\hat{q}_{\tau_n}$ ,  $\tilde{\xi}_{\tau_n}$  or  $\hat{\xi}_{\tau_n}$  yields estimators  $\widehat{\text{QMES}}_{X,\tau'_n}^*$ ,  $\widehat{\text{XMES}}_{X,\tau'_n}^*$  or  $\widehat{\text{XMES}}_{X,\tau'_n}^*$  of (4) or (5). The following second-order condition is needed to quantify the bias incurred in constructing these estimators. Below we use the classical asymptotic notation  $o(\cdot)$ ,  $O(\cdot)$  and  $O_{\mathbb{P}}(\cdot)$  (Dodge & Commenges 2006).

**Condition E.** Under Conditions **D(i)** and **D(iii)**, suppose that there exist  $\beta > \gamma_X$  and  $\kappa < 0$  such that  $R_{(X,Y)}(x, y) = r_1(x, \infty, y, \infty)$  satisfies  $R_{(X,Y)}(1, 1) > 0$  and

$$\sup_{x \in (0, \infty)} \left| \frac{s \mathbb{P}\{\bar{F}_X(X_1) \leq x/s, \bar{F}_Y(Y_1) \leq y/s\} - R_{(X,Y)}(x, y)}{\min(x^\beta, 1)} \right| = O(s^\kappa)$$

locally uniformly in  $y \in (0, \infty)$  as  $s \rightarrow \infty$ .

The following theorem gives the asymptotic distribution of  $\overline{\text{MES}}_{X,\tau'_n}^*$ .

**Theorem 4.1.** Suppose that  $X = X_1$  and  $Y = Y_1$  satisfy Condition **C** with respective parameters  $(\gamma_X, \rho_X, A_X)$  and  $(\gamma_Y, \rho_Y, A_Y)$ , and that Conditions **D** and **E** hold. Suppose also that  $\rho_X < 0$ , that there exists  $\delta > 0$  such that  $0 < \gamma_X < 1/(2 + \delta)$ , and that

(i)  $\tau_n, \tau'_n \uparrow 1$ , with  $n(1 - \tau_n) \rightarrow \infty$ ,  $n(1 - \tau'_n) \rightarrow c < \infty$  and  $\{n(1 - \tau_n)\}^{1/2} / \log\{(1 - \tau_n)/(1 - \tau'_n)\} \rightarrow \infty$  as  $n \rightarrow \infty$ ;

(ii)  $r_n(1 - \tau_n) \rightarrow 0$  and  $r_n[r_n\{n(1 - \tau_n)\}^{-1/2}]^\delta \rightarrow 0$  as  $n \rightarrow \infty$ ;

(iii) there exists  $\varepsilon > 0$  such that  $\{n(1 - \tau_n)\}^{1/2} |A_X\{(1 - \tau_n)^{-1}\}|^{\gamma_X/(1 - \rho_X) - \varepsilon} \rightarrow 0$ , and  $n(1 - \tau_n)^{1 - 2\kappa} \rightarrow 0$  as  $n \rightarrow \infty$ ;

(iv)  $\mathbb{E}(|X_-|^{1/\gamma_X}) < \infty$  and  $n(1 - \tau_n) \times (1 - \tau'_n)^{-2\kappa(1 - \gamma_X)} \rightarrow 0$  as  $n \rightarrow \infty$ ;

(v) the bias conditions

$$\{n(1 - \tau_n)\}^{1/2} \left\{ \frac{\bar{F}_Y(z_{Y, \tau_n})}{1 - \tau_n} - z \right\} = o(1), \quad \{n(1 - \tau_n)\}^{1/2} \left\{ \frac{\bar{F}_Y(z_{Y, \tau'_n})}{1 - \tau'_n} - z \right\} = o(1),$$

hold as  $n \rightarrow \infty$ ; and

(vi) the weak convergence  $\{n(1 - \tau_n)\}^{1/2}(\hat{\gamma}_{X,n} - \gamma_X) \rightarrow \Gamma$  holds, where  $\Gamma$  is a nondegenerate random variable, and

$$\{n(1 - \tau_n)\}^{1/2} \left( \frac{\bar{z}_{Y, \tau_n}}{z_{Y, \tau_n}} - 1 \right) = O_{\mathbb{P}}(1).$$

If also  $\{n(1 - \tau_n)\}^{1/2} A_Y\{(1 - \tau_n)^{-1}\} \rightarrow 0$  as  $n \rightarrow \infty$ , then

$$\frac{\{n(1 - \tau_n)\}^{1/2}}{\log\{(1 - \tau_n)/(1 - \tau'_n)\}} \left( \frac{\overline{\text{MES}}_{X, \tau'_n}^*}{\text{MES}_{X, \tau'_n}} - 1 \right) \rightarrow \Gamma$$

weakly. Condition (iv) is unnecessary if  $X > 0$  with probability one.

Condition (i) was used in Theorem 3.5. Condition (ii) is used to deal with serial dependence. Condition (iii) is slightly weaker than condition (d) of Cai et al. (2015), and Condition (iv) is taken from Theorem 2 therein. Condition (v) is used to control the error made in using the extrapolation relationship, and Condition (vi) ensures that all quantities used in the construction of the estimators converge at the appropriate rate. In this condition,  $\Gamma$  is typically a normal distribution, as is for instance the case when  $\hat{\gamma}_{X,n}$  is a Hill estimator. Using Theorem 4.1, we obtain the following important corollary.

**Corollary 4.2.** *Under the conditions of Theorem 4.1 (apart from (v) and (vi)), assume also that  $r_n \log^2(n(1-\tau_n))/\sqrt{n(1-\tau_n)} \rightarrow 0$  as  $n \rightarrow \infty$  and let  $\hat{\gamma}_{X,n} = \hat{\gamma}_{X,n}^H$ . Let  $R_{X,t}(x_1, x_{t+1}) := r_t(x_1, x_{t+1}, \infty, \infty)$ . Then as  $n \rightarrow \infty$  the weak convergence*

$$\frac{\sqrt{n(1-\tau_n)}}{\log\{(1-\tau_n)/(1-\tau'_n)\}} \left( \frac{\overline{\text{MES}}_{X,\tau'_n}^*}{\text{MES}_{X,\tau'_n}} - 1 \right) \rightarrow \mathcal{N} \left[ 0, \gamma_X^2 \left\{ 1 + 2 \sum_{t=1}^{\infty} R_{X,t}(1, 1) \right\} \right]$$

holds for

- $\overline{\text{MES}}_{X,\tau'_n}^* = \widehat{\text{QMES}}_{X,\tau'_n}^*$ , without further assumptions,
- $\overline{\text{MES}}_{X,\tau'_n}^* = \widehat{\text{XMES}}_{X,\tau'_n}^*$ , assuming further with the notation of Theorem 4.1 that  $\mathbb{E}|Y_-|^{2+\delta} < \infty$ ,  $0 < \gamma_Y < 1/(2+\delta)$ ,  $\sum_{l \geq 1} [b(l)]^{\delta/(2+\delta)} < \infty$  and that  $\sqrt{n(1-\tau_n)} q_{Y,\tau_n}^{-1} \rightarrow 0$  as  $n \rightarrow \infty$ , and
- $\overline{\text{MES}}_{X,\tau'_n}^* = \widehat{\text{XMES}}_{X,\tau'_n}^*$ , assuming further with the notation of Theorem 4.1 that  $\mathbb{E}|Y_-| < \infty$ ,  $0 < \gamma_Y < 1$ , and that  $\sqrt{n(1-\tau_n)} q_{Y,\tau_n}^{-1} \rightarrow 0$  and  $\sqrt{n(1-\tau_n)}(\hat{\gamma}_{Y,n} - \gamma_Y) = O_{\mathbb{P}}(1)$  as  $n \rightarrow \infty$ .

## 5 Finite-sample inference and expectile level selection

The theory developed in Sections 3 and 4 allows us to obtain confidence intervals for extreme expectiles and the expectile-based marginal expected shortfall, provided we can estimate the asymptotic variance  $w(\gamma, \mathbf{R}) = \gamma^2 \{1 + 2 \sum_{t=1}^{\infty} R_t(1, 1)\}$  in Theorem 3.5. By Proposition 2.1 in Drees (2003) we have, when  $n(1-\tau_n) \rightarrow \infty$ ,  $r_n \rightarrow \infty$  and  $r_n(1-\tau_n) \rightarrow 0$ ,

$$\frac{1}{r_n(1-\tau_n)} \text{Var} \left[ \sum_{i=1}^{r_n} \mathbb{1}\{F(Y_i) > \tau_n\} \right] \rightarrow 1 + 2 \sum_{t=1}^{\infty} R_t(1, 1), \quad n \rightarrow \infty.$$

We split the data into big blocks of length  $r_n$  separated by small blocks of length  $l_n$  and define

$$Z_j = \sum_{t=1+j\ell_n}^{r_n+j\ell_n} \mathbb{1}\{\hat{F}_n(Y_t) > \tau_n\}, \quad j = 0, 1, \dots, m_n - 1,$$

where  $m_n = \lfloor n/\ell_n \rfloor$ ,  $\ell_n = r_n + l_n$ , and  $\hat{F}_n$  is the empirical distribution function of all the data. We then compute the sample variance  $\Sigma_n$  of  $Z_0, \dots, Z_{m_n-1}$  and thus obtain an estimator

$$\hat{w}_n(\gamma, \mathbf{R}) = \frac{(\hat{\gamma}_n^H)^2}{r_n(1-\tau_n)} \Sigma_n \tag{6}$$

of  $w(\gamma, \mathbf{R})$ . Theorem 3.5 yields as  $n \rightarrow \infty$  the weak convergence

$$\frac{\{n(1 - \tau_n)\}^{1/2}}{\log\{(1 - \tau_n)/(1 - \tau'_n)\}} \log \frac{\bar{\xi}_{\tau'_n}^*}{\xi_{\tau'_n}} \rightarrow \mathcal{N} \left\{ \frac{\lambda_1}{1 - \rho}, w(\gamma, \mathbf{R}) \right\}.$$

Thus a  $(1 - \alpha)$ -equi-tailed confidence interval for  $\xi_{\tau'_n}$  has limits

$$\bar{\xi}_{\tau'_n}^* \left( \frac{1 - \tau_n}{1 - \tau'_n} \right)^{-\hat{b} \pm z_{1-\alpha/2} [\hat{w}_n(\gamma, \mathbf{R}) / \{n(1 - \tau_n)\}]^{1/2}}, \quad (7)$$

where  $\bar{\xi}_{\tau'_n}^*$  is the least asymmetrically weighted squares or quantile-based extrapolating estimator,  $\hat{b} = \hat{\gamma}_n^H \hat{\beta}_n (1 - \tau_n)^{-\hat{\rho}_n} / (1 - \hat{\rho}_n)$  estimates the bias  $\lambda_1 / (1 - \rho)$ , with the estimates  $\hat{\beta}_n$  and  $\hat{\rho}_n$  obtained from the R package `evt0` (Manjunath & Caeiro 2013), and  $z_p$  the  $p$ -quantile of the standard normal distribution. The intervals given by (7) can be constructed with  $\bar{\xi}_{\tau'_n}^*$  either the least asymmetrically weighted squares or the quantile-based extrapolating estimator, and ignoring the bias, i.e., setting  $\hat{b} = 0$ . The asymptotic variance in the independent case,  $\gamma^2$ , is estimated by  $(\hat{\gamma}_n^H)^2$ . Similar confidence intervals can be constructed for (5), by substituting  $X_t$  for  $Y_t$  in both the  $Z_j$  and the Hill estimator, and by replacing  $\bar{\xi}_{\tau'_n}^*$  either by  $\widehat{\text{XMES}}_{X, \tau'_n}^*$  or by  $\widehat{\text{XMES}}_{X, \tau'_n}^*$ .

A crucial issue in risk management is the choice of the prudentiality level  $\tau'_n$ . If expectiles are considered for their financial interpretation in terms of the gain-loss ratio, then  $\tau'_n$  should be selected to achieve a certain gain-loss value; otherwise, it seems reasonable to select  $\tau'_n$  so that  $\xi_{\tau'_n} \equiv q_{\alpha_n}$  for a given  $\alpha_n$ . In our context this level  $\tau'_n = \tau'_n(\alpha_n)$  satisfies  $\{1 - \tau'_n(\alpha_n)\} / (1 - \alpha_n) \rightarrow \gamma / (1 - \gamma)$  as  $n \rightarrow \infty$ . One can then define a natural estimator of  $\tau'_n(\alpha_n)$ , i.e.,

$$\hat{\tau}'_n(\alpha_n) = 1 - (1 - \alpha_n) \frac{\hat{\gamma}_n^H}{1 - \hat{\gamma}_n^H}.$$

Using this estimator in place of  $\tau'_n$  in the extrapolating least asymmetrically weighted squares and quantile-based estimators yields composite estimators of  $\xi_{\tau'_n(\alpha_n)} \equiv q_{\alpha_n}$ . If one uses the same  $\tau_n$  in the extrapolation step and in the calculation of  $\hat{\tau}'_n(\alpha_n)$ , the composite extrapolating least asymmetrically weighted squares estimator is

$$\tilde{\xi}_{\hat{\tau}'_n(\alpha_n)}^* = \left( \frac{1 - \hat{\tau}'_n(\alpha_n)}{1 - \tau_n} \right)^{-\hat{\gamma}_n^H} \tilde{\xi}_{\tau_n} = ((\hat{\gamma}_n^H)^{-1} - 1)^{\hat{\gamma}_n^H} \tilde{\xi}_{\alpha_n}^*.$$

In other words, rewriting Equation (2) as  $q_{\alpha_n} \equiv \xi_{\tau'_n(\alpha_n)} \approx (\gamma^{-1} - 1)^\gamma \xi_{\alpha_n}$ , the composite extrapolating least asymmetrically weighted squares estimator can be constructed by

inserting the estimator  $\hat{\gamma}_n^H$  and the extrapolating least asymmetrically weighted squares estimator at level  $\alpha_n$  into the right-hand side of this approximation. This has not, to the best of our knowledge, previously been noted in the literature. When combining this with our marginal expected shortfall estimators, this gives two estimators of  $\text{QMES}_{X,\alpha_n}$ . The asymptotic properties of the estimators discussed so far are established next.

**Theorem 5.1.** *Suppose the conditions of Theorem 3.5 hold with  $\alpha_n$  in place of  $\tau'_n$ . Then, if  $\hat{\gamma}_n = \hat{\gamma}_n^H$  and  $\bar{\xi}^*$  is either  $\hat{\xi}^*$  or  $\tilde{\xi}^*$ , we have*

$$\frac{\sqrt{n(1-\tau_n)}}{\log[(1-\tau_n)/(1-\alpha_n)]} \left( \frac{\bar{\xi}_{\hat{\tau}'_n(\alpha_n)}^*}{q_{\alpha_n}} - 1 \right) \rightarrow \mathcal{N} \left( \frac{\lambda_1}{1-\rho}, \gamma^2 \left[ 1 + 2 \sum_{t=1}^{\infty} R_t(1,1) \right] \right),$$

in the sense of weak convergence as  $n \rightarrow \infty$ .

An analogous asymptotic normality result holds for the composite estimators  $\widetilde{\text{XMES}}_{X,\hat{\tau}'_n(\alpha_n)}^*$  and  $\widehat{\text{XMES}}_{X,\hat{\tau}'_n(\alpha_n)}^*$  of  $\text{QMES}_{X,\alpha_n}$ .

**Theorem 5.2.** *Suppose the conditions of Corollary 4.2 hold with  $\alpha_n$  in place of  $\tau'_n$ . Then, if  $\hat{\gamma}_{X,n} = \hat{\gamma}_{X,n}^H$  and  $\overline{\text{XMES}}^* = \widetilde{\text{XMES}}^*$  (resp.  $\overline{\text{XMES}}^* = \widehat{\text{XMES}}^*$ ), then*

$$\frac{\sqrt{n(1-\tau_n)}}{\log[(1-\tau_n)/(1-\alpha_n)]} \left( \frac{\overline{\text{XMES}}_{X,\hat{\tau}'_n(\alpha_n)}^*}{\text{QMES}_{X,\alpha_n}} - 1 \right) \rightarrow \mathcal{N} \left( 0, \gamma_X^2 \left[ 1 + 2 \sum_{t=1}^{\infty} R_{X,t}(1,1) \right] \right),$$

in the sense of weak convergence as  $n \rightarrow \infty$ .

## 6 Simulation experiments

### 6.1 Extreme expectile estimation

Here we summarise a numerical study of the finite-sample performance of the point and interval expectile estimators at extreme levels. More details and results on a similar study carried out for the estimators at the intermediate level may be found in Section B of the Supplementary Material. We consider simulations from

- (a) the AR(1) model  $Y_{t+1} = 0.8Y_t + \varepsilon_{t+1}$ , where the innovations  $\varepsilon_t$  are independent and have a  $t_3$  distribution;



- (b) the ARMA(1,1) model  $Y_{t+1} = 0.95 Y_t + \varepsilon_{t+1} + 0.9 \varepsilon_t$ , where the innovations  $\varepsilon_t$  are independent and have a symmetric Pareto distribution with shape parameter  $\zeta = 3$ ;
- (c) the ARCH(1) model  $Y_{t+1} = \sigma_{t+1} \varepsilon_{t+1}$ , where  $\sigma_{t+1}^2 = 0.4 + 0.6 Y_t^2$ , and the  $\varepsilon_t$  are independent standard Gaussian innovations; and
- (d) the GARCH(1,1) model  $Y_{t+1} = \sigma_{t+1} \varepsilon_{t+1}$ , where  $\sigma_{t+1}^2 = 0.1 + 0.4 Y_t^2 + 0.4 \sigma_t^2$ , and the  $\varepsilon_t$  are independent standard Gaussian innovations.

The first two models have strong linear serial dependence, and the last two have quadratic serial dependence leading to heteroskedasticity. If standard linear time series have heavy-tailed innovations with balanced tails then the tail index of  $(Y_t)$  is that of the innovations, i.e.,  $1/3$  in models (a) and (b). The marginal distribution in (c) and (d) is heavy-tailed, and the tail indices, calculated using Theorem 2.1 in Mikosch & Stărică (2000), equal 0.262 and 0.239 respectively.

We simulate  $10^4$  samples of size  $n = 2500$  for each model and consider the extreme level  $\tau'_n = 0.9995 \approx 1 - 1/n$ ; values of  $\xi_{\tau'_n}$  are in the top part of Table B.5 in the Supplementary Material. We calculate the extrapolating least asymmetrically weighted squares and quantile-based estimators using the intermediate level  $\tau_n = 1 - k/n$  for  $k \in \{6, 8, \dots, 700\}$  for each simulated dataset and obtain various 95% confidence intervals based on (7). The big- and small-block sequences are chosen as  $r_n = \lfloor \log^2(n) \rfloor$  and  $l_n = \lfloor C \log n \rfloor$ , where  $C$  is selected such that  $l_n$  is greater than or equal to a lag after which the largest of the sample autocorrelations of the data from the time series  $(X_t)$  and  $(X_t^2)$  are smaller than 0.1. Checking whether the confidence intervals contain the true expectile allows us to compute the empirical coverage probability.

The top row of Figure 1 shows that our confidence intervals have much better coverages than those that assume the data to be independent, and the coverages for intervals based on least asymmetrically weighted squares outperform those for the quantile-based estimators over a wider range of the intermediate sequence  $k$ . Coverages for the heteroskedastic models are satisfactory for a shorter range of the intermediate sequence  $k$  and the intervals for these models tend to be slightly permissive. The bias-corrected versions behave well in heteroskedastic models and are slightly conservative for linear time series. Overall,

according to a rule of thumb to select a  $k$  that falls in  $[50, 200]$ , our proposed confidence intervals perform very well for such a value of  $n$ . Similar results at level  $\tau'_n = 0.9999 > 1 - 1/n$  are found in Figure B.2 of the Supplementary Material.

## 6.2 Extreme marginal expected shortfall estimation

In order to assess the finite-sample performance of the marginal expected shortfall estimators at extreme levels we simulate data from four models in which the sequences of bivariate innovations are independent and identically distributed:

- (e)  $X_{t+1} = 0.8 X_t + \varepsilon_{X,t+1}$  and  $Y_{t+1} = 0.8 Y_t + \varepsilon_{Y,t+1}$ . For any  $t$ , the innovation  $\varepsilon_{X,t}$  is distributed as  $Z \mathbb{1}(Z > 0) - (-Z)^{1/2} \mathbb{1}(Z < 0)$ ,  $Z$  and  $\varepsilon_{Y,t}$  have  $t_3$  distributions, and the dependence of  $(\varepsilon_{X,t}, \varepsilon_{Y,t})$  is given by a Student- $t$  copula with correlation parameter  $\rho = 0.8$  and three degrees of freedom;
- (f)  $X_{t+1} = 0.95 X_t + \varepsilon_{X,t+1} + 0.9 \varepsilon_{X,t}$ , and  $Y_{t+1} = 0.95 Y_t + \varepsilon_{Y,t+1} + 0.9 \varepsilon_{Y,t}$ . For any  $t$ , the innovation  $\varepsilon_{X,t}$  is distributed as  $Z \mathbb{1}(Z > 0) - (-Z)^{1/2} \mathbb{1}(Z < 0)$ , where  $Z$  and  $\varepsilon_{Y,t}$  have symmetric Pareto distributions with shape parameter  $\zeta = 3$ , and  $(\varepsilon_{X,t}, \varepsilon_{Y,t})$  have dependence given by a Gumbel copula with parameter  $\theta = 2$ ;
- (g)  $X_{t+1} = \sigma_{X,t+1} \varepsilon_{X,t+1}$  and  $Y_{t+1} = \sigma_{Y,t+1} \varepsilon_{Y,t+1}$ , with  $\sigma_{X,t+1}^2 = 0.4 + 0.6 X_t^2$  and  $\sigma_{Y,t+1}^2 = 0.4 + 0.6 Y_t^2$ . Each  $\varepsilon_{X,t}$  has density  $h(z) = 0.5 \mathbb{1}(-1 < z \leq 0) + 0.5 e^{-z} \mathbb{1}(z > 0)$ ,  $\varepsilon_{Y,t}$  is standard Gaussian, and the dependence of  $(\varepsilon_{X,t}, \varepsilon_{Y,t})$  is as in (a); and
- (h)  $X_{t+1} = \sigma_{X,t+1} \varepsilon_{X,t+1}$ , with  $\sigma_{X,t+1}^2 = 0.1 + 0.4 X_t^2 + 0.4 \sigma_{X,t}^2$ , and  $Y_{t+1} = \sigma_{Y,t+1} \varepsilon_{Y,t+1}$ , with  $\sigma_{Y,t+1}^2 = 0.1 + 0.4 Y_t^2 + 0.4 \sigma_{Y,t}^2$ . Each  $\varepsilon_{X,t}$  has density  $h(z) = 0.5 \mathbb{1}(-1 < z \leq 0) + 0.5 e^{-z} \mathbb{1}(z > 0)$ ,  $\varepsilon_{Y,t}$  is standard Gaussian, and the dependence of  $(\varepsilon_{X,t}, \varepsilon_{Y,t})$  is given by a Gumbel copula with parameter  $\theta = 5$ .

The  $Y_t$  components of models (e), (f), (g) and (h) are distributed according to models (a), (b), (c) and (d) respectively, so the models considered here extend those of Section 6.1.

We simulate  $10^4$  samples of size  $n = 2500$  for each model, and consider estimating  $\text{QMES}(\alpha_n)$  at levels  $\alpha_n$  such that  $\tau'_n(\alpha_n) = 0.9995$  and  $0.9999$ , using our composite estimators  $\widetilde{\text{XMES}}_{X, \hat{\tau}'_n(\alpha_n)}^*$  and  $\widehat{\text{XMES}}_{X, \hat{\tau}'_n(\alpha_n)}^*$ ; recall that  $(1 - \alpha_n) \approx (\gamma_Y^{-1} - 1)\{1 - \tau'_n(\alpha_n)\}$  and the

construction of  $\hat{\tau}'_n(\alpha_n)$  in Section 5. The true values of  $\text{QMES}(\alpha_n) \equiv \text{XMES}\{\tau'_n(\alpha_n)\}$ , found by simulation, are in the bottom part of Table B.5 in Supplementary Material. Estimation at this level  $\alpha_n$  gives us an idea of the performance of our composite estimation method, in a problem comparable in difficulty to that of Section 6.1, at least as far as the  $Y$  component is concerned. The confidence intervals are constructed as in Section 6.1, applied to the  $X_t$  component. In the quantile-based composite estimator  $\widehat{\text{XMES}}_{X, \hat{\tau}'_n(\alpha_n)}^*$ , the Hill estimators of  $\gamma_X$  and  $\gamma_Y$  use the same  $k$  for simplicity.

The bottom row of Figure 1 gives results at level  $\tau'_n(\alpha_n) = 0.9995$ . The intervals for the linear time series have the same broad properties as in the upper panels. The coverages of our proposed confidence intervals are very good for linear time series, for a wide range of values of the intermediate sequence  $k$  and for the bias-corrected and unadjusted versions. The coverages with the heteroskedastic models (g) and (h) are less satisfactory, with our confidence intervals estimators providing reasonable results for a narrower range of values of  $k$ . Nevertheless, our intervals represent a significant improvement with respect to those relying on the theory for i.i.d. observations. Also in this case, with the aforementioned rule of thumb (selecting  $k \in [50, 200]$ ), our proposed confidence intervals provide reliable results overall. Results at the level  $\tau'_n(\alpha_n) = 0.9999$  may be found in Figure B.3 of the Supplementary Material.

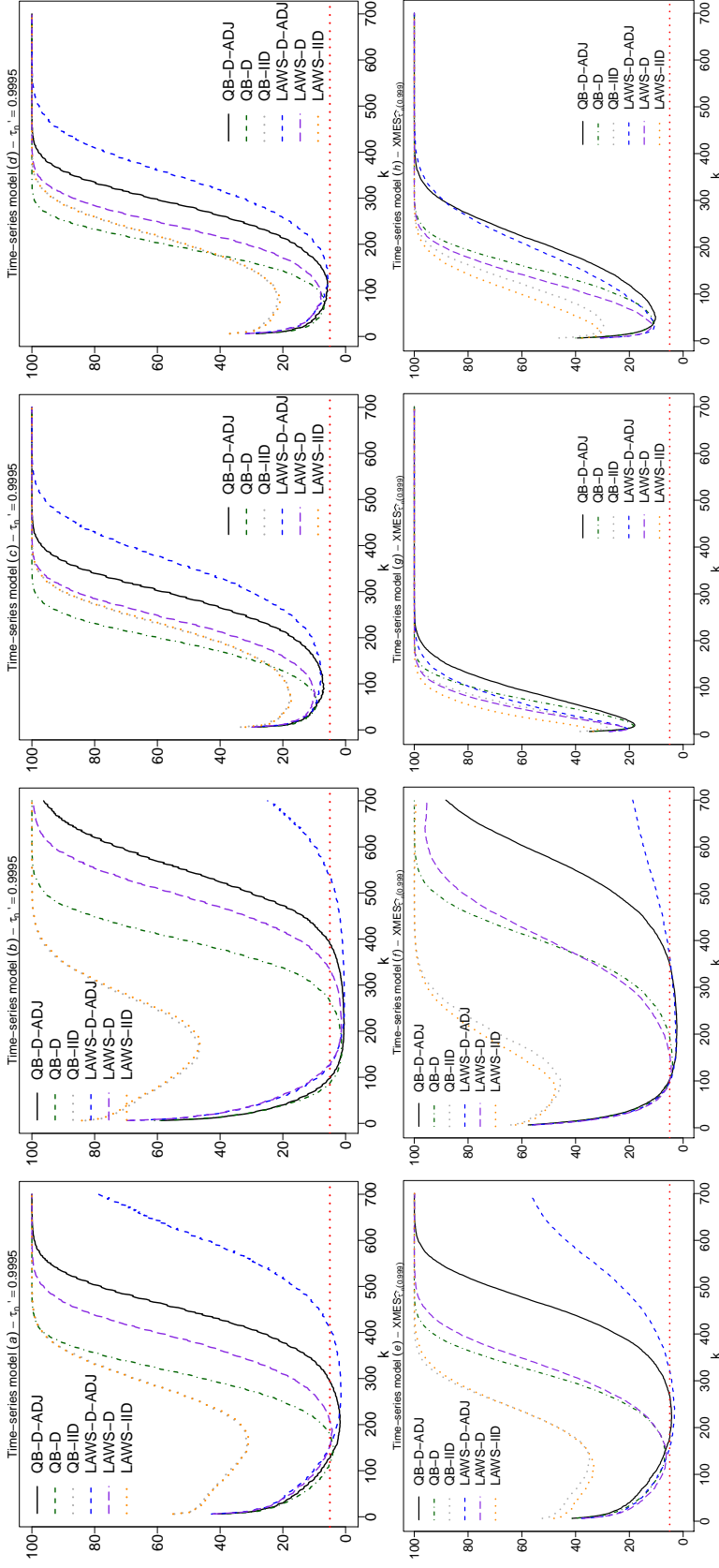


Figure 1: Empirical error rate (%) for nominal 95% confidence intervals for the expectile  $\xi_{\tau_n}'$  (top row, with  $\tau_n' = 0.9995$ ) and the marginal expected shortfall  $\text{QMES}(\alpha_n)$  (bottom row, with level  $\alpha_n$  such that  $\tau_n'(\alpha_n) = 0.9995$ ). The true model is indicated above each panel. The horizontal dotted red line represents the 5% nominal error rate. The quantile-based methods are QB-D-ADJ (using the extrapolating estimator), QB-D (setting  $\hat{b} = 0$ ), QB-IID (ignoring the dependence), and likewise for the least asymmetric weighted square estimators.

## 7 Applications

### 7.1 Stock market index data

The rightmost panels of the top two rows in Figure 2 show  $n = 8,785$  daily negative log-returns of the S&P 500 and Dow Jones Industrial Average indices from 29 January 1985 to 12 December 2019. The data illustrate stylised facts such as the heteroskedasticity and fat tails of financial time series (Embrechts et al. 1997). For these two series, the Hill, maximum likelihood and moment-based estimators of the tail index (e.g., de Haan & Ferreira 2006, Chapter 3) are fairly stable for  $k \in [100, 300]$  and suggest that these series have heavy right tails with  $\gamma \approx 0.35$  (see top panels of Figure B.4 in the Supplementary Material). Below we use the Hill estimator to construct our estimates and confidence intervals.

The analysis of tail risk of loss returns is typically based on estimated Value-at-Risk at the 99.9% level (e.g., Drees 2003, de Haan et al. 2016) or at some level  $\alpha_n = 1 - p_n$ , for  $p_n \leq 1/n$ . Bellini & Di Bernardino (2017) showed that such estimates yield capital requirements similar to expectile-based forecasts, if the level  $\tau'_n$  of the expectile is chosen carefully, at a higher level than that of the Value-at-Risk. Here we fix  $p_n = 1/n$  and  $\alpha_n = 1 - p_n = 0.9998862$ , and we estimate  $\tau'_n(\alpha_n)$  by  $\hat{\tau}'_n(\alpha_n)$ . Then we estimate the expectile at the extreme level  $\hat{\tau}'_n(\alpha_n)$  using the composite extrapolating least asymmetrically weighted squares estimator  $\tilde{\xi}_{\tau'_n}^*$  and the corresponding quantile-based estimator  $\hat{\xi}_{\tau'_n}^*$  for  $\tau'_n = \hat{\tau}'_n(\alpha_n)$ . This produces estimators of  $\xi_{\tau'_n(\alpha_n)}$ , which is also the Value-at-Risk  $q_{\alpha_n} = q_{1-1/n}$ . The leftmost panels of the first and second row in Figure 2 plot  $\hat{\tau}'_n(\alpha_n)$  against  $k$  for  $k \leq 700$ , where  $\tau_n = 1 - k/n$  as before. These estimates fluctuate initially, then stabilise around a common value, and finally drift away due to the inclusion of data from the centre of the distribution; taking  $k = 200$  seems reasonable. We check this by calculating the composite extrapolating least asymmetrically weighted squares and composite extrapolating quantile-based estimators and the various confidence intervals of Section 5 at level  $\tau'_n = \hat{\tau}'_n(\alpha_n)$ . The estimates and confidence intervals appear to be fairly stable provided  $k$  is not too small, the choice  $k = 200$  again seems sensible. As expected, the confidence intervals that take the dependence into account are wider than those that do not. With  $k = 200$ , we find  $\hat{\tau}'_n(\alpha_n) \approx 0.9999423$  for the S&P 500 data, and 0.9999402 for the Dow Jones data, both

rather higher than the original  $\alpha_n = 1 - p_n = 0.9998862$ .

Figure 2 also compares our extrapolating least asymmetrically weighted squares and quantile-based estimators with the Weissman extreme quantile estimator at level  $\alpha_n$ , i.e.,

$$\hat{q}_{\alpha_n}^* = \left( \frac{1 - \alpha_n}{1 - \tau_n} \right)^{-\hat{\gamma}_n^H} \hat{q}_{\tau_n} = \left( \frac{1 - \alpha_n}{1 - \tau_n} \right)^{-\hat{\gamma}_n^H} X_{n - \lfloor n(1 - \tau_n) \rfloor, n}. \quad (8)$$

Confidence intervals for the extreme quantile  $q_{\alpha_n}$  can also be constructed using this estimator: here we use a method of Drees (2003) for the estimation of the variance component  $w(\gamma, \mathbf{R})$  in Section 5. Unlike our estimator, this method does not rely on a big-block/small-block argument, see Formula (33) of Drees (2003). The estimates and 95% confidence intervals are reported in Table 1 and are shown in the rightmost panels of the top two rows in Figure 2. The point estimates are reassuringly similar, but the third and fourth panels suggest that the confidence intervals based on  $\hat{q}_{\alpha_n}^*$  are generally much more volatile than the least asymmetrically weighted squares interval; moreover, in a neighbourhood of  $k \approx 200$ , they are very close to the intervals based on independent observations. For our selected  $k = 200$ , the fourth column of Figure 2 shows that these intervals do not contain the largest sample value, despite estimating  $q_{\alpha_n} = q_{1-1/n}$ , whereas the least asymmetrically weighted squares intervals do contain the sample maximum.

## 7.2 Financial returns of individual banks

We now analyse the financial returns of Goldman Sachs and Morgan Stanley in the context of systemic risk. We consider the daily negative log-returns  $(X_t)$  on their equity prices from 3 July 2000 to 30 June 2010, alongside daily loss returns  $(Y_t)$  of a value-weighted market index aggregating the New York Stock Exchange, the American Express Stock Exchange and the National Association of Securities Dealers Automated Quotation system for the same period. The corresponding tail index estimates again indicate heavy right tails (see bottom panels of Figure B.4 in the Supplementary Material, we have represented the same type of estimates for the  $(Y_t)$  series in Figure B.5 of the Supplementary Material). The leftmost panels of the first and second row of Figure 2 display values of  $\hat{\tau}_n'(\alpha_n)$  against  $k$  with again  $\alpha_n = 1 - 1/n = 0.9996021$  and  $\tau_n = 1 - k/n$ .

Table 1: Estimates and 95% confidence intervals for the daily log-returns for the S&P 500 and Dow Jones data with  $k = 200$  and  $\alpha_n = 1 - 1/n = 0.9998862$ .

Estimate	S&P 500	Dow Jones
$\hat{\gamma}_n^H$	0.336 [0.220, 0.453]	0.344 [0.222, 0.467]
$\tilde{\xi}_{\hat{\tau}_n(\alpha_n)}^*$	0.136 [0.064, 0.259]	0.136 [0.061, 0.263]
$\hat{\xi}_{\hat{\tau}_n(\alpha_n)}^*$	0.140 [0.064, 0.258]	0.139 [0.061, 0.260]
$\hat{q}_{\alpha_n}^*$	0.140 [0.112, 0.174]	0.139 [0.103, 0.190]

Table 2: Estimates and 95% confidence intervals for the loss returns for Goldman Sachs and Morgan Stanley data with  $k = 150$  and  $\alpha_n = 1 - 1/n = 0.9996021$ .

Estimate	Goldman Sachs	Morgan Stanley
$\hat{\gamma}_{X,n}^H$	0.410 [0.172, 0.648]	0.459 [0.279, 0.639]
$\widehat{\text{XMES}}_{X,\hat{\tau}'_n(\alpha_n)}^*$	0.342 [0.079, 1.046]	0.590 [0.185, 1.303]
$\widehat{\text{XMES}}_{X,\hat{\tau}'_n(\alpha_n)}^*$	0.345 [0.080, 1.055]	0.603 [0.189, 1.332]
$\widehat{\text{QMES}}_{X,\alpha_n}^*$	0.345 [0.216, 0.551]	0.603 [0.294, 1.236]

These estimates slightly fluctuate initially, then stabilise around a common value, and finally drift away due to the inclusion of data from the centre of the distribution; taking  $k = 150$  seems reasonable. With the same setup, the estimates obtained with the composite extrapolating least asymmetrically weighted squares and composite extrapolating quantile-based estimators of  $\text{QMES}_{X,\alpha_n}$ , and the various confidence intervals at level  $\tau'_n = \hat{\tau}'_n(\alpha_n)$  are reported in first three panels of Figure 3. The estimates and confidence intervals are more stable for Goldman Sachs than for Morgan Stanley and for the least asymmetric squares method than for the quantile-based method. According to these results, taking  $k = 150$  seems reasonable, and we then find  $\hat{\tau}'_n(\alpha_n) \approx 0.9997239$  for the Goldman Sachs data and 0.9996626 for the Morgan Stanley data, which we compare to the Weissman

estimate  $\widehat{\text{QMES}}_{X, \alpha_n}^*$ . All estimates are reported in Table 2 and the least asymmetrically weighted squares estimates are shown in the rightmost panels of the bottom two rows of Figure 2; they lead to conclusions similar to those in Section 7.1. For  $k = 150$  they give similar results on the Goldman Sachs data, and somewhat shorter confidence intervals on the Morgan Stanley data. Like on the stock market index data, the confidence intervals constructed with the method of Drees (2003) are more volatile than ours.

In an application either to stock market index data or to an individual bank, it would be wise to inspect the analogue of Figure 2, in order to identify stable regions for the estimates and confidence intervals.

## Acknowledgements

G. Stupfler was previously based at the University of Nottingham, and the support of the Nottingham PEF Fund, of the French National Research Agency (grant ANR-19-CE40-0013) and of the AXA Research Fund is gratefully acknowledged. S.A. Padoan is supported by the Bocconi Institute for Data Science and Analytics. A. C. Davison is supported by the Swiss National Science Foundation.

## Supplementary material

Supplementary material available at *Journal of Business & Economic Statistics* online provides a detailed discussion of the technical conditions, further numerical details and results, and all proofs.

## References

- Acharya, V., Pedersen, L., Philippon, T. & Richardson, M. (2017), ‘Measuring systemic risk’, *The Review of Financial Studies* **30**, 2–47.
- Artzner, P., Delbaen, F., Eber, J.-M. & Heath, D. (1999), ‘Coherent measures of risk’, *Mathematical Finance* **9**, 203–228.



- Beirlant, J., Goegebeur, Y., Segers, J. & Teugels, J. (2004), *Statistics of Extremes: Theory and Applications*, John Wiley & Sons, Chichester.
- Bellini, F. & Di Bernardino, E. (2017), ‘Risk management with expectiles’, *The European Journal of Finance* **23**, 487–506.
- Bellini, F., Klar, B., Müller, A. & Gianin, E. R. (2014), ‘Generalized quantiles as risk measures’, *Insurance: Mathematics and Economics* **54**, 41–48.
- Boussama, F., Fuchs, F. & Stelzer, R. (2011), ‘Stationarity and geometric ergodicity of BEKK multivariate GARCH models’, *Stochastic Processes and their Applications* **121**, 2331–2360.
- Brownlees, C. T. & Engle, R. (2017), ‘SRISK: A conditional capital shortfall measure of systemic risk’, *The Review of Financial Studies* **30**, 48–79.
- Cai, J.-J., Einmahl, J. H. J., de Haan, L. & Zhou, C. (2015), ‘Estimation of the marginal expected shortfall: the mean when a related variable is extreme’, *Journal of the Royal Statistical Society: Series B* **77**, 417–442.
- Chavez-Demoulin, V. & Guillou, A. (2018), ‘Extreme quantile estimation for  $\beta$ -mixing time series and applications’, *Insurance: Mathematics and Economics* **83**, 59–74.
- 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.
- de Haan, L. & Ferreira, A. (2006), *Extreme Value Theory: An Introduction*, Springer-Verlag, New York.
- de Haan, L., Mercadier, C. & Zhou, C. (2016), ‘Adapting extreme value statistics to financial time series: dealing with bias and serial dependence’, *Finance and Stochastics* **20**, 321–354.

- Dodge, Y. & Commenges, D. (2006), *The Oxford dictionary of statistical terms*, Oxford University Press on Demand.
- Doukhan, P. (1994), *Mixing: Properties and Examples*, Springer, New York.
- Drees, H. (2000), ‘Weighted approximations of tail processes for  $\beta$ -mixing random variables’, *Annals of Applied Probability* **10**, 1274–1301.
- Drees, H. (2003), ‘Extreme quantile estimation for dependent data, with applications to finance’, *Bernoulli* **9**, 617–657.
- Drees, H. & Rootzén, H. (2010), ‘Limit theorems for empirical processes of cluster functionals’, *Annals of Statistics* **38**, 2145–2186.
- Ehm, W., Gneiting, T., Jordan, A. & Krüger, F. (2016), ‘Of quantiles and expectiles: consistent scoring functions, Choquet representations and forecast rankings’, *Journal of the Royal Statistical Society: Series B* **78**, 505–562.
- Embrechts, P., Klüppelberg, C. & Mikosch, T. (1997), *Modelling Extremal Events for Insurance and Finance*, Springer-Verlag, Berlin.
- Francq, C. & Zakoïan, J.-M. (2006), ‘Mixing properties of a general class of GARCH(1,1) models without moment assumptions on the observed process’, *Econometric Theory* **22**, 815–834.
- Gneiting, T. (2011), ‘Making and evaluating point forecasts’, *Journal of the American Statistical Association* **106**, 746–762.
- Hill, B. M. (1975), ‘A simple general approach to inference about the tail of a distribution’, *Annals of Statistics* **3**, 1163–1174.
- Holzmann, H. & Klar, B. (2016), ‘Expectile asymptotics’, *Electronic Journal of Statistics* **10**, 2355–2371.
- Jiang, R., Hu, X. & Yu, K. (2021), ‘Single-index expectile models for estimating conditional Value at Risk and Expected Shortfall’, *Journal of Financial Econometrics*, to appear. <https://doi.org/10.1093/jjfinec/nbaa016>.

- Krätschmer, V. & Zähle, H. (2017), ‘Statistical inference for expectile-based risk measures’, *Scandinavian Journal of Statistics* **44**, 425–454.
- Kuan, C.-M., Yeh, J.-H. & Hsu, Y.-C. (2009), ‘Assessing value at risk with CARE, the Conditional Autoregressive Expectile models’, *Journal of Econometrics* **150**, 261–270.
- Manjunath, B. & Caeiro, F. (2013), ‘`evt0`: Mean of order  $p$ , peaks over random threshold Hill and high quantile estimates’, *R package version* pp. 1–1.
- Mikosch, T. & Stărică, C. (2000), ‘Limit theory for the sample autocorrelations and extremes of a GARCH(1,1) process’, *Annals of Statistics* **28**, 1427–1451.
- Newey, W. K. & Powell, J. L. (1987), ‘Asymmetric least squares estimation and testing’, *Econometrica* **55**, 819–847.
- Weissman, I. (1978), ‘Estimation of parameters and large quantiles based on the  $k$  largest observations’, *Journal of the American Statistical Association* **73**, 812–815.
- Xie, S., Zhou, Y. & Wan, A. T. K. (2014), ‘A varying-coefficient expectile model for estimating Value at Risk’, *Journal of Business & Economic Statistics* **32**, 576–592.

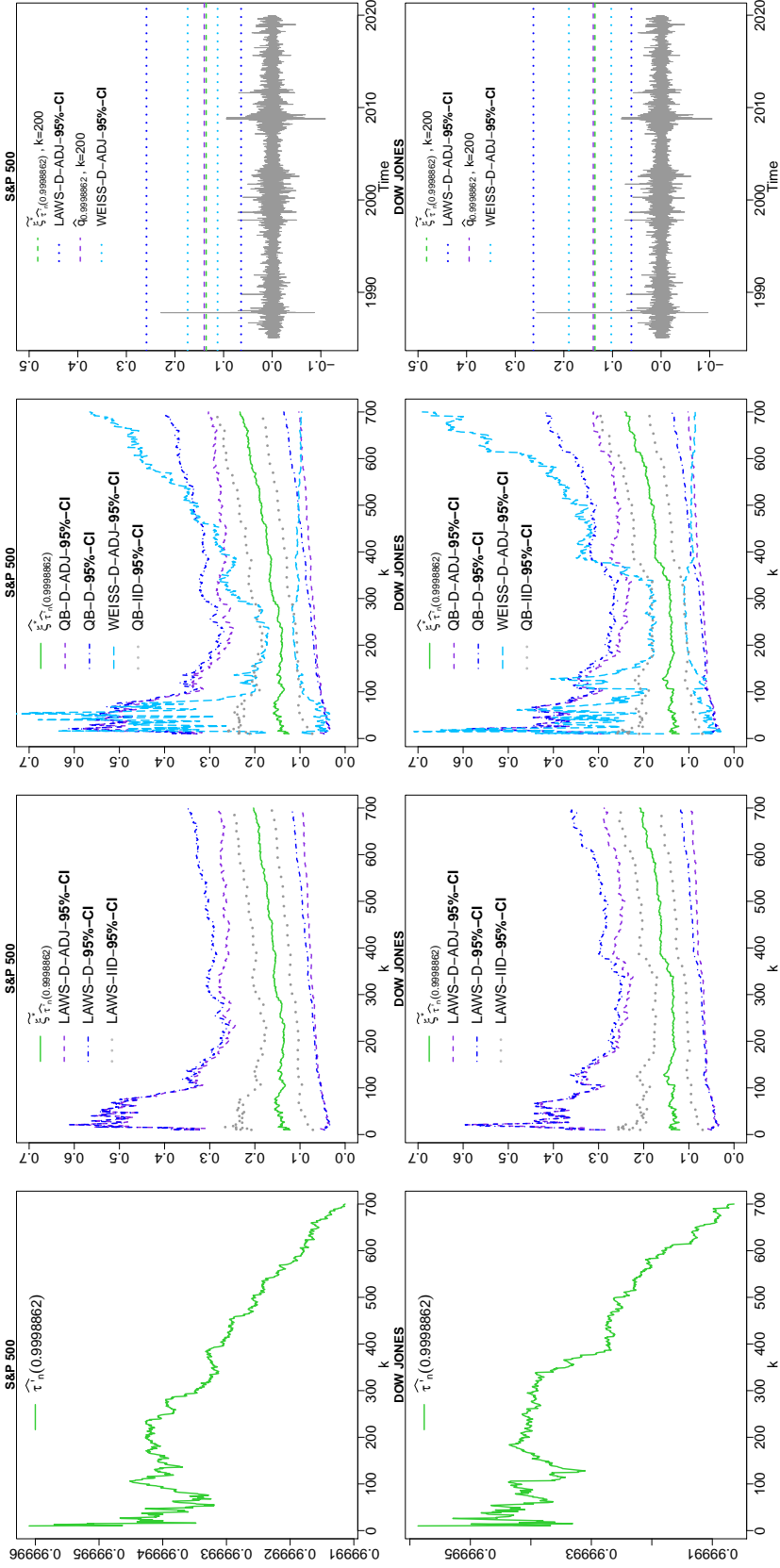


Figure 2: Financial data analysis. Analysis of S&P 500 and Dow Jones industrial average data with  $\alpha_n = 0.9998862$ . From left to right: estimate  $\hat{\tau}_n(\alpha_n)$ ; composite extrapolated least asymmetrically weighted squares estimate at level  $\hat{\tau}_n(\alpha_n)$ , with 95% LAWS-IID, LAWS-D and LAWS-D-ADJ confidence intervals; composite extrapolated quantile-based estimate at level  $\alpha_n$ , with 95% QB-IID, QB-D, QB-D-ADJ and WEISS-D-ADJ confidence intervals; the data, with composite extrapolated least asymmetrically weighted squares estimate at level  $\hat{\tau}_n(\alpha_n)$  and the Weissman quantile estimate  $\hat{q}_{\alpha_n}^*$ , both with 95% confidence intervals.

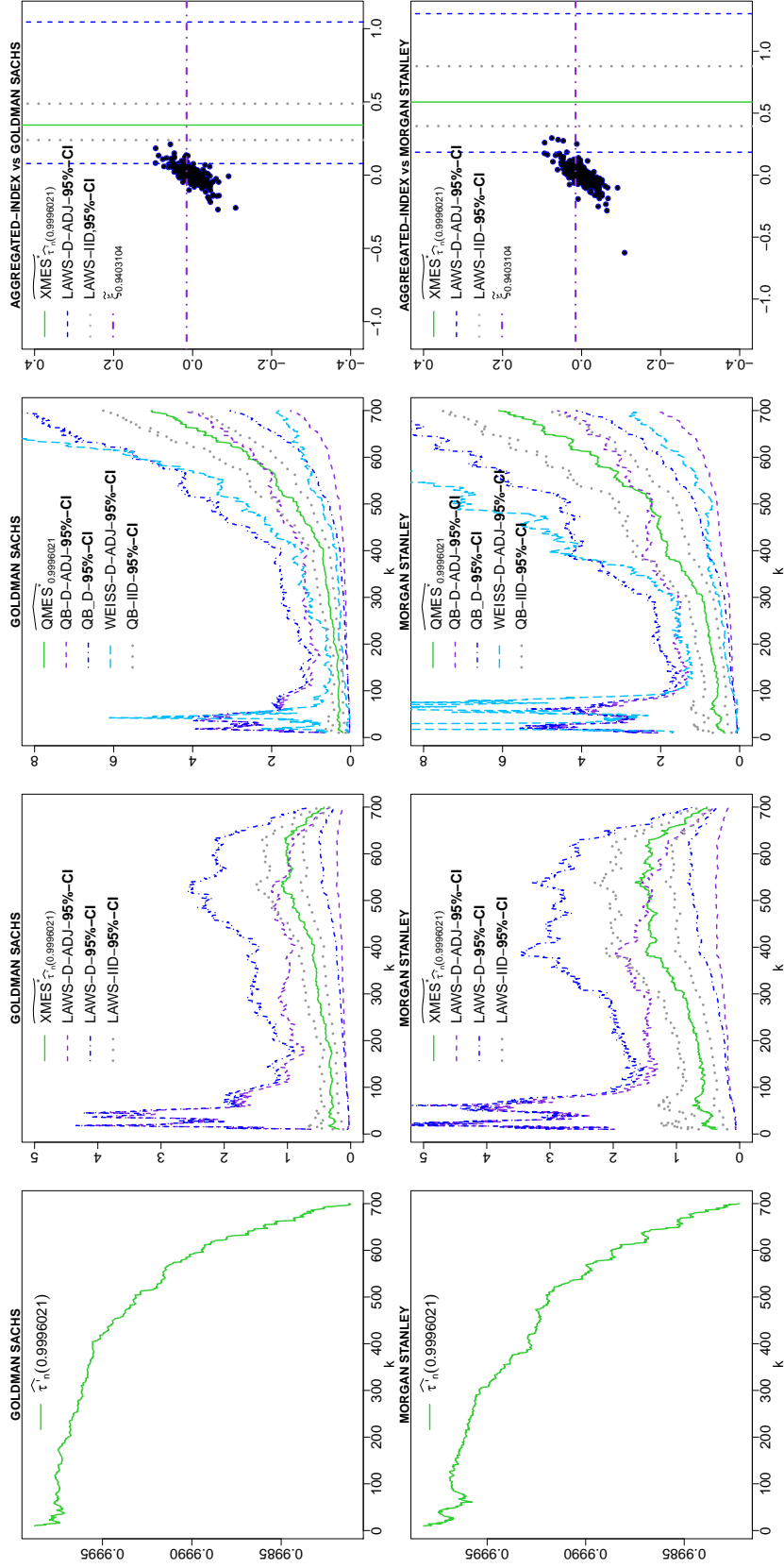


Figure 3: Financial data analysis. Analysis of Goldman Sachs and Morgan Stanley data with  $\alpha_n = 0.9996021$ . From left to right: estimate  $\hat{\tau}_n(\alpha_n)$ ; composite extrapolated least asymmetrically weighted squares estimate of (5) at level  $\hat{\tau}_n(\alpha_n)$ , with 95% LAWS-IID, LAWS-D and LAWS-D-ADJ confidence intervals; least asymmetrically weighted squares estimate of (5) at level  $\alpha_n$ , with 95% QB-IID, QB-D, QB-D-ADJ and WEISS-D-ADJ confidence intervals; scatterplot of the data, with the composite extrapolated least asymmetrically weighted squares estimate of (5) at level  $\hat{\tau}_n(\alpha_n)$  with 95% confidence intervals. The purple horizontal line is the intermediate level  $\tau_n$ . The WEISS-D-ADJ confidence intervals are based on (8); see the caption to Figure 1 for the other abbreviations.