Extreme expectile estimation for heavy-tailed time series

Simone A. Padoan, Gilles Stupfler

Extreme expectile estimation for heavy-tailed time series

Simone A. Padoan[*] and Gilles Stupfler[†]

April 28, 2020

Abstract

Expectiles are a least squares analogue of quantiles which have lately received substantial attention in actuarial and financial risk management contexts. Unlike quantiles, expectiles define coherent risk measures and are determined by tail expectations rather than tail probabilities; unlike the Expected Shortfall, they define elicitable risk measures. This has motivated recent studies of the behaviour and estimation of extreme expectile-based risk measures. The case of stationary but weakly dependent observations has, however, been left largely untouched, even though correctly accounting for the uncertainty present in typical financial applications requires the consideration of dependent data. We investigate the estimation of, and construction of accurate confidence intervals for, extreme expectiles and expectile-based Marginal Expected Shortfall in a general $\beta-$mixing context, containing the classes of ARMA, ARCH and GARCH models with heavy-tailed innovations that are of interest in financial applications. The methods are showcased in a numerical simulation study and on real financial data.

Keywords: Asymmetric least squares, Expectiles, Extremal dependence, Heavy tails, Marginal Expected Shortfall, Mixing, Tail copula, Weak dependence.

1 Introduction

A major problem in econometrics and statistical finance is to quantify the risk associated to a real-valued profit-loss variable Y. The best-known risk measure in the financial sector is arguably Value-at-Risk (VaR) at a confidence level $\tau \in (0,1)$, defined as the negative τth quantile $-q_\tau$ of the real-valued profit-loss distribution, with τ being close to zero representing the situations carrying the greatest risk. The quantile can be obtained by minimising asymmetrically weighted mean absolute deviations (Koenker and Bassett, 1978):

$$q_\tau \in \underset{q\in\mathbb{R}}{\arg\min}\, \mathbb{E}(\rho_\tau(Y-q) - \rho_\tau(Y)), \tag{1}$$

[*]Simone A. Padoan works at the Department of Decision Sciences, Bocconi University of Milan, via Roentgen 1, 20136 Milano, Italy, E-mail: `simone.padoan@unibocconi.it`.

[†]Gilles Stupfler works at Univ Rennes, Ensai, CNRS, CREST - UMR 9194, F-35000 Rennes, France, E-mail: `gilles.stupfler@ensai.fr`.

where $\rho_\tau(u) = |\tau - \mathbb{1}\{u \leq 0\}||u|$ is the quantile check function and $\mathbb{1}\{\cdot\}$ the indicator function. The VaR suffers from certain serious drawbacks. It does not, in general, define a coherent risk measure in the sense of Artzner et al. (1999), and thus does not abide by the intuitive diversification principle. Besides, quantiles are often criticised for missing out on important information since they only depend on the frequency of tail losses and not on their actual values. Unlike the VaR, the popular quantile-based Expected Shortfall (ES, or Conditional VaR), is coherent, takes into account the actual values of the risk variable on the tail event, but is not elicitable in the sense of Gneiting (2011), and as such does not benefit from the existence of a natural backtesting methodology.

An alternative risk measure which addresses these issues is given by the concept of expectiles, introduced by Newey and Powell (1987). Expectiles extend the mean in the way quantiles extend the median and are found by substituting the absolute deviations in (1) with squared deviations:

$$\xi_\tau = \arg\min_{\theta \in \mathbb{R}} \mathbb{E}(\eta_\tau(Y - \theta) - \eta_\tau(Y)), \tag{2}$$

where $\eta_\tau(u) = |\tau - \mathbb{1}\{u \leq 0\}|u^2$. For each $\tau \in (0, 1)$, the τth expectile exists, is uniquely defined by its convex problem, and satisfies $\tau = \mathbb{E}[|Y - \xi_\tau|\mathbb{1}\{Y \leq \xi_\tau\}]/\mathbb{E}|Y - \xi_\tau|$. This property of expectiles is intimately connected to the notion of gain-loss ratio, which is a popular performance measure in portfolio management (see Bellini and Di Bernardino, 2017, and references therein). It implies that, unlike the VaR, the τth expectile is determined by tail expectations rather than tail probabilities and thus depends on tail realisations of the loss variable and their probability. This motivated Kuan et al. (2009) to introduce a notion of expectile-based Value-at-Risk as $-\xi_\tau$ for real-valued profit-loss distributions. The advantages of the expectile include that it induces a law-invariant, coherent and elicitable risk measure, see Bellini et al. (2014) and Ziegel (2016). It is actually the only risk measure, apart from the simple expectation, satisfying these three properties. Further results, both theoretical and numerical, obtained by Ehm et al. (2016) and Bellini and Di Bernardino (2017) among others, indicate that expectiles define sensible alternatives to the standard VaR and ES.

Expectile estimation has recently regained interest in a wide range of complex models, see for example Sobotka and Kneib (2012) and the references therein, as well as Holzmann and Klar (2016) and Krätschmer and Zähle (2017) for advanced theoretical results. However, a theory for extreme expectiles is still in full development. Probabilistic aspects of extreme expectiles, with $\tau \uparrow 1$, have been examined by Bellini et al. (2014) and Bellini and Di Bernardino (2017). Inference on extreme expectiles has been considered even more recently in Daouia et al. (2018, 2019, 2020). This literature on extreme expectile estimation has been restricted to independent and identically distributed (i.i.d.) data or, in Daouia et al. (2019), to strictly stationary $\phi-$mixing time series. This extension is, in fact, only of minor interest because standard processes in financial and econometric modelling, such as ARCH and GARCH processes, are not in general $\phi-$mixing. Asymptotic results currently available on extreme expectile estimation thus cannot reasonably be used in typical financial contexts to produce accurate confidence intervals. This is a serious gap, to be addressed if expectiles are to be used widely in financial risk management.

This paper contributes to filling that gap in the following way. In a general framework of $\beta-$mixing and heavy-tailed strictly stationary time series, we first rigorously investigate the convergence of several estimation techniques for extreme expectile-based risk measures. We start by, given a strictly stationary, $\beta-$mixing time series (Y_t) having a one-dimensional marginal heavy-tailed distribution that satisfies general conditions, considering the estimation of an intermediate tail expectile of order $\tau_n \to 1$ such that $n(1-\tau_n) \to \infty$ as $n \to \infty$, where n is the available sample size. This framework makes it possible to consider a wide class of processes, such as ARMA, ARCH and GARCH processes under reasonably mild conditions. We focus on two estimation methods: the Least Asymmetrically Weighted Squares (LAWS) estimator, defined as the direct empirical counterpart of the expectile in (2), and the indirect Quantile-Based (QB) estimator obtained using an asymptotic proportionality relationship that links high expectiles to their quantile counterparts. The resulting estimates are extrapolated to proper extreme levels τ_n' converging to 1 at an arbitrarily fast rate in the sense that $n(1-\tau_n') \to c$ as $n \to \infty$, for some nonnegative constant c. We then expand our statistical model to include the bivariate $\beta-$mixing time series framework (X_t, Y_t), with the goal of establishing a general theory for the estimation of the Marginal Expected Shortfall (MES). This risk measure has been argued to be important in the assessment of systemic financial risk by, among others, Engle et al. (2015), Acharya et al. (2017) and Brownlees and Engle (2017). In this context, $X := X_1$ and $Y := Y_1$ respectively stand for the marginal losses on the return of a financial firm and that of the entire market, and the MES is $\mathbb{E}(X|Y > u)$, where u is large to reflect a substantial market downturn. Our results apply not only to the estimation of the expectile-based form of the extreme MES, where u is taken to be a high expectile of Y, but also to its quantile-based counterpart. As an application of our results, we then discuss the important problem of constructing accurate yet readily implementable asymptotic confidence intervals for extreme expectile-based risk measure, that take into account the dependence between observations. They are compared, on simulated and real data, to the naive intervals obtained via the asymptotic theory of i.i.d. observations, to show the importance of accounting for the dependence structure of typical financial time series.

The outline of the paper is the following. Section 2 explains in detail our statistical context. Section 3 considers first intermediate and then extreme expectile estimation. Section 4 then introduces a more general bivariate time series context within which general MES estimators at extreme levels are investigated. Section 5 explores the implications of our results on asymptotic confidence interval construction. Section 6 discusses the important question of the selection of expectile level in practice. The finite-sample performance of the methods is examined on simulated data sets in Section 7 and on real financial data in Section 8. The methods and data considered in this article have been incorporated into the `R` package `ExtremeRisks`, freely available on CRAN. The `R` code for the simulation study and real data analysis is available in the "Software" Section of S. Padoan's personal webpage at the URL `http://mypage.unibocconi.it/simonepadoan/`. A Supplementary Material document discusses in more depth our technical conditions, gives all necessary mathematical proofs, and contains further finite-sample results.

2 Statistical model and time series framework

Let $(Y_t,\, t \geq 1)$ be a strictly stationary time series having a continuous marginal heavy-tailed distribution F: in other words, F is the distribution function of $Y := Y_1$ and of each Y_t. Let $\overline{F} := 1 - F$ and $U : s \mapsto \inf\{y \in \mathbb{R} \,|\, 1/\overline{F}(y) \geq s\}$ be its survival and tail quantile functions. Throughout, Y should be seen as the negative of a generic financial position, so that large positive values of Y represent extreme losses associated to this position.

Motivated by applications to financial risk management, our target is the estimation of extreme expectiles of Y, having order tending to 1. To this end, we focus on heavy-tailed distributions, which are found to represent the tail structure of many financial data examples quite well and therefore are ubiquitous in the modelling of extreme actuarial and financial losses, as argued on p.9 of Embrechts et al. (1997). Mathematically, we assume that

$$\forall y > 0,\ \lim_{s\to\infty} \frac{\overline{F}(sy)}{\overline{F}(s)} = y^{-1/\gamma} \ \text{ or equivalently } \ \lim_{s\to\infty} \frac{U(sy)}{U(s)} = y^{\gamma}. \tag{3}$$

The tail index γ specifies the tail heaviness of $\overline{F}$: the tail of $\overline{F}$ gets heavier as γ increases. Together with condition $\mathbb{E}|Y_-| < \infty$, where $Y_- := \min(Y, 0)$, the assumption $\gamma < 1$ then ensures that the first moment of Y exists, which entails that expectiles of Y of any order are well-defined. These conditions shall therefore be part of our minimal assumptions.

An extension of the results of Daouia et al. (2018), on extreme expectile estimation for heavy tails in the i.i.d. setup, is given in Daouia et al. (2019) in a $\phi-$mixing dependence framework. To the best of our knowledge, the latter is the only work in the literature considering the estimation of extreme expectiles in heavy-tailed models for weakly dependent data. The $\phi-$mixing framework is the following. For any $m \geq 1$, let $\mathcal{F}_{1,m} = \sigma(Y_1, \ldots, Y_m)$ and $\mathcal{F}_{m,\infty} = \sigma(Y_m, Y_{m+1}, \ldots)$ denote the past and future σ-fields generated by the sequence (Y_t). The $\phi-$mixing coefficients of this sequence are then defined by

$$\forall\, l \geq 1,\ \phi(l) = \sup_{m\geq 1} \sup_{A\in\mathcal{F}_{1,m}} \sup_{B\in\mathcal{F}_{m+l,\infty}} |\mathbb{P}(B|A) - \mathbb{P}(B)|.$$

The time series (Y_t) is said to be $\phi-$mixing (or uniformly strongly mixing) if $\phi(l) \to 0$ as $l \to \infty$. This is in fact a very stringent assumption. For instance, even the simple AR(1) process with heavy-tailed innovations is never $\phi-$mixing (see the Introduction of Rio, 2017). We work here in the more general context of $\beta-$mixing, defined through the coefficients

$$\forall\, l \geq 1,\ \beta(l) = \sup_{m\geq 1} \mathbb{E}\left(\sup_{B\in\mathcal{F}_{m+l,\infty}} |\mathbb{P}(B|\mathcal{F}_{1,m}) - \mathbb{P}(B)| \right).$$

The time series (Y_t) is then said to be $\beta-$mixing (or absolutely regular) if $\beta(l) \to 0$ as $l \to \infty$. Roughly speaking, the $\beta-$mixing property brings a form of memorylessness much weaker than its $\phi-$mixing version: the $\beta-$mixing property is written in an L^1 sense, while the $\phi-$mixing property is written in the much stronger L^∞ sense. That $\beta-$mixing is weaker than $\phi-$mixing can be seen by noting that $\beta(l) \leq \phi(l)$ for any l, see Doukhan (1994, Section 1.1).

Our motivation for making the $\beta-$mixing assumption is twofold. On the one hand, $\beta-$mixing is satisfied in a much wider class of models than $\phi-$mixing: for instance, Doukhan (1994, Section 2.4) shows that a large class of Markov processes, among which ARMA processes, nonlinear autoregressive processes, ARCH and GARCH models are in fact geometrically β-mixing (*i.e.* there is $a < 1$ such that $\beta(l) \leq a^l$ for l large enough) under reasonably mild conditions. On the other hand, there is a general theory of extremes for strictly stationary and $\beta-$mixing processes, developed in a series of papers by Drees (2000, 2002, 2003). This important portion of the extreme value literature provides probabilistic tools for the statistical analysis of extremes of strictly stationary and $\beta-$mixing observations through a powerful approximation result for the tail quantile process by a Gaussian process. The $\beta-$mixing condition has more generally played a substantial role in contemporary research on the extremes of a time series: see among others de Haan et al. (2016) and Chavez-Demoulin and Guillou (2018) for the development of bias-reduced estimators in the dependent setting, and Drees and Rootzén (2010) for further theoretical contributions. The $\beta-$mixing assumption thus strikes a good balance between theoretical applicability and modelling strength, and as such constitutes a reasonable framework for our objective of estimating extreme expectiles in heavy-tailed time series. This motivates our basic modelling assumption below.

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

Condition A will be key to our development of an asymptotic theory for extreme expectile estimation, which we provide in the next section.

3 Extreme expectile estimation in time series

Suppose that we observe a random sample $(Y_1, \ldots, Y_n)$ extracted from a time series (Y_t) satisfying Condition A. The objective in this section is to estimate a marginal, unconditional extreme expectile ξ_{τ_n} of the random variable $Y = Y_1$, where $\tau_n \to 1$ as $n \to \infty$. Note that this estimation problem is different from the prediction of extreme expectiles in dynamic time series models, where the interest is in estimating conditional extreme expectile levels for tomorrow given our knowledge of today.

We shall start by the case of an intermediate level τ_n, meaning that $\tau_n \to 1$ and $n(1-\tau_n) \to \infty$ as $n \to \infty$. Intermediate expectile estimates will then be extrapolated to estimate expectiles at properly extreme levels τ_n', satisfying $n(1-\tau_n') \to c > 0$ as $n \to \infty$, using a semiparametric approach warranted by the heavy-tailed assumption.

3.1 At intermediate levels

Direct asymmetric least squares estimator Let τ_n be an intermediate level. We first consider estimating the expectile ξ_{τ_n} of the marginal distribution F by its empirical estimator

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

This LAWS estimator can easily be computed, for example using an iteratively reweighted least squares minimisation procedure. To find the asymptotic distribution of $\widetilde{\xi}_{\tau_n}$, we make the following assumption on the dependence within the time series (Y_t).

Condition B. For the time series (Y_t), assume that

(i) There are sequences of integers (l_n) and (r_n) such that

$$l_n \to \infty,\ r_n \to \infty,\ \frac{l_n}{r_n} \to 0,\ \frac{r_n}{n} \to 0 \text{ and } \frac{n\,\beta(l_n)}{r_n} \to 0 \quad \text{as } n \to \infty.$$

(ii) For any $t \geq 1$, there is a function R_t on $[0,\infty]^2 \setminus \{(\infty,\infty)\}$ such that

$$\forall (x,y) \in [0,\infty]^2 \setminus \{(\infty,\infty)\},\ \lim_{s\to\infty} s\,\mathbb{P}\left(\overline{F}(Y_1) \leq \frac{x}{s},\ \overline{F}(Y_{t+1}) \leq \frac{y}{s}\right) = R_t(x,y).$$

(iii) There exist $D \geq 0$ and a nonnegative sequence $\rho(t)$ satisfying $\sum_{t\geq 1}\rho(t) < \infty$ and such that, for s large enough, we have

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

for any $t \geq 1$ and all $u, u', v, v' \in [0,1]$ with $u' < u$ and $v' < v$.

Condition B(i) and B(ii) are standard in the emerging literature on extreme value analysis with mixing conditions, see *e.g.* Drees (2002, 2003) and Drees and Rootzén (2010) (we thank Holger Drees for pointing out in private communication that there is a typo in the first term of condition (C1) in Drees, 2003). In Assumption B(i), the sequences (l_n) and (r_n) are small-block and big-block sequences used to develop the kind of "big blocks separated by small blocks" arguments that are succesfully employed in the literature on mixing time series. Condition B(iii) is slightly more precise than condition (C3) in Drees (2003).

We now state our first main result on the asymptotic normality of the estimator $\widetilde{\xi}_{\tau_n}$.

Theorem 3.1. *Assume that Conditions A and B are satisfied. Assume further that there is* $\delta > 0$ *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* $n(1-\tau_n) \to \infty$, $r_n(1-\tau_n) \to 0$ *and* $r_n(r_n/\sqrt{n(1-\tau_n)})^\delta \to 0$ *as* $n \to \infty$. *Then,*

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

with $\sigma^2(\gamma,\boldsymbol{R}) := \dfrac{(1-\gamma)(1-2\gamma)}{\gamma^2}\displaystyle\iint_{[1,\infty)^2}\sum_{t=1}^{\infty} R_t(x^{-1/\gamma}, y^{-1/\gamma})\,\mathrm{d}x\,\mathrm{d}y.$

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 $\sigma^2(\gamma, \boldsymbol{R})$ represents the proportion of increase of this asymptotic variance due to the mixing setting. The conditions $\mathbb{E}|Y_-|^{2+\delta} < \infty$ and $0 < \gamma < 1/(2+\delta)$ already appear in Theorem 2 of Daouia et al. (2018) for the i.i.d. case, of which the present result can be considered a generalisation. In addition, our Theorem 3.1 represents a substantial theoretical step compared to Theorem 1 of Daouia et al. (2019) in the $\phi-$mixing case. While the latter is shown by simply updating a couple of correlation calculations in the proof of Theorem 2 of Daouia et al. (2018), the proof of Theorem 3.1 uses some rather delicate arguments involving a tailored central limit theory for tail array sums in the time-dependent setting, developed by Rootzén et al. (1998).

An exhaustive discussion of our hypotheses is somewhat involved. However, our assumptions are very mild when $\beta(l)$ converges to 0 geometrically fast as $l \to \infty$. 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 (where $\lfloor \cdot \rfloor$ denotes the floor function). The case of geometrically strong $\beta-$mixing covers many cases widely used in the modelling of financial time series, such as ARMA processes, ARCH/GARCH processes and solutions of stochastic difference equations (see Doukhan, 1994; Drees, 2000, 2003; Francq et al., 2006; Boussama et al., 2011). The following corollary summarises that discussion in this important geometrically mixing case.

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

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

with the notation of Theorem 3.1.

We turn to the consideration of a different estimator, built on the asymptotic proportionality between high expectiles and their quantile counterparts.

Indirect quantile-based estimator A competitor is obtained by exploiting an asymptotic proportionality relationship between high expectiles and quantiles: within our heavy-tailed model,

$$\frac{\xi_\tau}{q_\tau} \to (\gamma^{-1} - 1)^{-\gamma} \quad \text{as} \quad \tau \uparrow 1. \tag{4}$$

This was first noted by Bellini et al. (2014). An indirect QB estimator of ξ_{τ_n} can then be obtained through the asymptotic proportionality relationship (4):

$$\widehat{\xi}_{\tau_n} = (\widehat{\gamma}_n^{-1} - 1)^{-\widehat{\gamma}_n} \widehat{q}_{\tau_n},$$

where $\widehat{q}_{\tau_n} = Y_{n-\lfloor n(1-\tau_n)\rfloor,n}$ is the empirical counterpart of q_{τ_n} (where $Y_{1,n} \leq \cdots \leq Y_{n,n}$ are the ascending order statistics of $(Y_1, \ldots, Y_n)$) and $\widehat{\gamma}_n$ is a consistent estimator of γ.

Our next main contribution is to give, in our mixing time series framework, two examples of estimators $\widehat{\gamma}_n$ for which the asymptotic distribution of $\widehat{\xi}_{\tau_n}$ can rigorously be established. We start by the Hill estimator (Hill, 1975), which is also the maximum likelihood estimator in a purely Pareto model and is arguably the most popular semiparametric estimator in the analysis of heavy tails:

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

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. We do so under the classical second-order condition below, which controls the gap between the right tail of $\overline{F}$ and a purely Pareto tail.

Condition C. The function $\overline{F}$ is second-order regularly varying in a neighbourhood of $+\infty$ with index $-1/\gamma < 0$, second-order parameter $\rho \leq 0$ and an auxiliary measurable function A having constant sign and converging to 0 at infinity. Precisely,

$$\forall\, y > 0,\ \lim_{s\to\infty} \frac{1}{A(1/\overline{F}(s))}\left[\frac{\overline{F}(sy)}{\overline{F}(s)} - y^{-1/\gamma}\right] = y^{-1/\gamma}\frac{y^{\rho/\gamma}-1}{\gamma\rho},$$

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

Further interpretation of this assumption can be found in Beirlant et al. (2004) and de Haan and Ferreira (2006) along with numerous examples of commonly used continuous distributions satisfying this condition. We are now ready to state our general result on the process $\widehat{q}$. See also Theorem 2.4.8 in de Haan and Ferreira (2006) for an analogue result in the independent case, as well as Proposition A.1 in de Haan et al. (2016) and Proposition 1 in Chavez-Demoulin and Guillou (2018) for related statements.

Theorem 3.3. *Assume that Conditions A, B and C are satisfied. Assume that $\tau_n \uparrow 1$, $n(1-\tau_n) \to \infty$, $r_n(1-\tau_n) \to 0$, $r_n \log^2(n(1-\tau_n))/\sqrt{n(1-\tau_n)} \to 0$ and $\sqrt{n(1-\tau_n)}A((1-\tau_n)^{-1}) = \mathrm{O}(1)$ as $n \to \infty$. 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 $s \in (0, s_0]$,

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

and

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

Note that Condition B is assumed in Theorem 3.3 for the sake of 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 $n(1-\tau_n) \to \infty$, $r_n(1-\tau_n) \to 0$, $r_n \log^2(n(1-\tau_n))/\sqrt{n(1-\tau_n)} \to 0$, $\sqrt{n(1-\tau_n)} A((1-\tau_n)^{-1}) \to \lambda_1 \in \mathbb{R}$ and $\sqrt{n(1-\tau_n)} q_{\tau_n}^{-1} \to \lambda_2 \in \mathbb{R}$ as $n \to \infty$. Then, for $\widehat{\gamma}_n = \widehat{\gamma}_n^H$ in the estimator $\widehat{\xi}_{\tau_n}$, one has*

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

with

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

Corollary 3.4 is, to the best of our knowledge, the first result on the QB estimator at intermediate levels under weak dependence assumptions. This result contains Corollary 2 in Daouia et al. (2018), restricted to the i.i.d. setup, in which case the asymptotic variance of $\widehat{\xi}_{\tau_n}$ is $\gamma^2(1+[m(\gamma)]^2)$, as our result indeed shows by taking $R_t \equiv 0$ for any $t \geq 1$.

There are of course many other ways to estimate the tail index γ. We briefly present here an alternative estimator based on the use of intermediate expectiles. The asymptotic proportionality relationship (4) can be equivalently rephrased as $\overline{F}(\xi_\tau)/(1-\tau) \to \gamma^{-1} - 1$ as $\tau \uparrow 1$, which implies

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

Taking $\tau = \tau_n \to 1$, and estimating $\overline{F}(\xi_\tau)$ by $\widehat{\overline{F}}_n(\widetilde{\xi}_{\tau_n})$, where $\widehat{\overline{F}}_n(u) = n^{-1}\sum_{t=1}^n \mathbb{1}\{Y_t > u\}$ is the empirical survival function, suggests the Expectile-Based (EB) estimator

$$\widehat{\gamma}_n^E = \left(1 + \frac{\widehat{\overline{F}}_n(\widetilde{\xi}_{\tau_n})}{1-\tau_n}\right)^{-1}.$$

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 a valuable device in the construction of confidence intervals requiring an estimate of γ.

3.2 At extreme levels

We now consider the important problem of the estimation of extreme expectiles $\xi_{\tau_n'}$, whose level $\tau_n' \to 1$ satisfies $n(1-\tau_n') \to c \in [0,\infty)$ as $n \to \infty$. A typical choice in applications is $\tau_n' = 1 - p_n$ for an exceedance probability p_n not greater than $1/n$, see *e.g.* Cai et al. (2015). The idea of the semiparametric approach we present here is to define an estimator of an extreme expectile through a Weissman-type construction (Weissman, 1978). This is motivated by a combination of the heavy-tailed assumption with Equation (4), resulting in

$$\frac{\xi_{\tau_n'}}{\xi_{\tau_n}} \approx \frac{q_{\tau_n'}}{q_{\tau_n}} = \frac{U((1-\tau_n')^{-1})}{U((1-\tau_n)^{-1})} \approx \left(\frac{1-\tau_n'}{1-\tau_n}\right)^{-\gamma} \quad \text{as } n \to \infty. \tag{5}$$

This suggests to consider the following class of plug-in estimators of $\xi_{\tau_n'}$:

$$\overline{\xi}^{\star}_{\tau_n'} \equiv \overline{\xi}^{\star}_{\tau_n'}(\tau_n) := \left(\frac{1-\tau_n'}{1-\tau_n}\right)^{-\widehat{\gamma}_n} \overline{\xi}_{\tau_n}$$

where $\widehat{\gamma}_n$ and $\overline{\xi}_{\tau_n}$ are consistent estimators of γ and of the intermediate expectile ξ_{τ_n}, respectively. We say that $\overline{\xi}^{\star}_{\tau_n'}$ is the extrapolating LAWS estimator when $\overline{\xi}_{\tau_n} = \widetilde{\xi}_{\tau_n}$, and we denote it by $\widetilde{\xi}^{\star}_{\tau_n'}$. We call it the extrapolating QB estimator when $\overline{\xi}_{\tau_n} = \widehat{\xi}_{\tau_n}$, and we denote it by $\widehat{\xi}^{\star}_{\tau_n'}$.

When $\widehat{\gamma}_n$ is chosen to be the Hill estimator $\widehat{\gamma}_n^H$, we have the following asymptotic normality result.

Theorem 3.5. *Assume 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 $n(1-\tau_n) \to \infty$, $n(1-\tau_n') \to c \in [0,\infty)$ and $\sqrt{n(1-\tau_n)}/\log[(1-\tau_n)/(1-\tau_n')] \to \infty$ as $n \to \infty$. Assume also that $r_n(1-\tau_n) \to 0$, $r_n \log^2(n(1-\tau_n))/\sqrt{n(1-\tau_n)} \to 0$, $\sqrt{n(1-\tau_n)}A((1-\tau_n)^{-1}) \to \lambda_1 \in \mathbb{R}$ and $\sqrt{n(1-\tau_n)}q_{\tau_n}^{-1} \to \lambda_2 \in \mathbb{R}$ as $n \to \infty$. Suppose finally that*

$$\sqrt{n(1-\tau_n)}\left(\frac{\overline{\xi}_{\tau_n}}{\xi_{\tau_n}} - 1\right) \xrightarrow{d} \Delta.$$

Then, if $\widehat{\gamma}_n = \widehat{\gamma}_n^H$ in $\overline{\xi}^{\star}_{\tau_n'}$, one has

$$\frac{\sqrt{n(1-\tau_n)}}{\log[(1-\tau_n)/(1-\tau_n')]}\left(\frac{\overline{\xi}^{\star}_{\tau_n'}}{\xi_{\tau_n'}} - 1\right) \xrightarrow{d} \mathcal{N}\left(\frac{\lambda_1}{1-\rho}, \gamma^2\left[1 + 2\sum_{t=1}^{\infty} R_t(1,1)\right]\right).$$

Theorem 3.5 makes it possible to construct confidence intervals for our extreme expectile estimators, and, in view of Theorem 3.1 and Corollary 3.2, applies indifferently to the estimators extrapolated from the direct and indirect intermediate expectile estimators. When

$R_t \equiv 0$ for any $t \geq 1$, the asymptotic variance is γ^2. The quantity $2\sum_{t=1}^{\infty} R_t(1,1)$ represents the proportion of increase of this asymptotic variance, compared to the i.i.d. case, due to the temporal dependence. An analogue result is of course possible for the EB estimator $\widehat{\gamma}_n^E$, although we shall not pursue this for the sake of brevity.

4 Marginal expected shortfall estimation

When working with actuarial and financial data, it is also important to assess a global form of risk, for instance by considering several lines of business of an insurance company or several stock market indices simultaneously. The theory we have developed so far is written for univariate time series. As such, it cannot account for dependence between multiple risk variables and therefore cannot be used to produce more insightful global risk estimates.

A prominent way of measuring systemic risk is via the Marginal Expected Shortfall (MES), defined in the econometric literature by, among others, Acharya et al. (2017), Engle et al. (2015) and Brownlees and Engle (2017) as the propensity of a financial institution to be undercapitalised when the financial system as a whole is undercapitalised. These authors measure the contribution of an individual firm, with loss return X, to systemic risk, represented by a loss Y in the aggregated return of the market, using the quantile-based MES

$$\mathrm{QMES}_{X,\tau} = \mathbb{E}(X|Y > q_{Y,\tau}),\ \tau \in (0,1), \tag{6}$$

where $q_{Y,\tau}$ is the τth quantile of Y. A systemic crisis typically corresponds to the case when τ is extremely high and potentially larger than $1-1/n$, where n is the sample size of available historical data. Daouia et al. (2018) study instead the expectile-based MES

$$\mathrm{XMES}_{X,\tau} = \mathbb{E}(X|Y > \xi_{Y,\tau}),\ \tau \in (0,1), \tag{7}$$

with $\xi_{Y,\tau}$ denoting the τth expectile of Y. The estimation of $\mathrm{QMES}_{X,\tau}$ and $\mathrm{XMES}_{X,\tau}$ at extreme levels $\tau = \tau_n \uparrow 1$ is considered in Cai et al. (2015) and Daouia et al. (2018), respectively, without recourse to the parametric specification and limited time horizon of Acharya et al. (2017), or the restrictions of the methods of Engle et al. (2015) and Brownlees and Engle (2017) that cannot handle extreme events with $1-\tau = \mathrm{O}(1/n)$. The results of Cai et al. (2015) and Daouia et al. (2018) are, however, limited to i.i.d. data, making inference about the QMES and XMES difficult in financial settings unless one works with low-frequency data.

Our goal here is to extend the theory of Section 3 to derive an inferential framework for these notions of MES at extreme levels in the weakly dependent setting. Suppose that the data comes from a strictly stationary bivariate time series $(X_t, Y_t,\ t \geq 1)$; for instance, X_t and Y_t could be respectively the daily loss returns on a specific stock and on an aggregated market index. For any $m \geq 1$, let $\mathcal{F}_{1,m} = \sigma(X_1, Y_1, \ldots, X_m, Y_m)$ and $\mathcal{F}_{m,\infty} = \sigma(X_m, Y_m, X_{m+1}, Y_{m+1}, \ldots)$ denote the past and future σ-fields generated by the sequence (X_t, Y_t). Then, the $\beta-$mixing coefficients of the sequence (X_t, Y_t) can be defined as

$$\forall\, l \geq 1,\ b(l) = \sup_{m\geq 1} \mathbb{E}\left(\sup_{B\in\mathcal{F}_{m+l,\infty}} |\mathbb{P}(B|\mathcal{F}_{1,m}) - \mathbb{P}(B)| \right).$$

The sequence (X_t, Y_t) is then said to be $\beta-$mixing if $b(l) \to 0$ as $l \to \infty$. Note that if (X_t, Y_t) is $\beta-$mixing in this sense, then (Y_t) is also $\beta-$mixing in the sense of Section 2. Our modelling condition below on (X_t, Y_t) similarly extends Conditions A and B.

Condition D. For the bivariate time series (X_t, Y_t), assume that

(i) The time series (X_t, Y_t) is strictly stationary, $\beta-$mixing and the one-dimensional marginal distribution functions F_X and F_Y of (X_t) and (Y_t) are both continuous and heavy-tailed with respective tail indices γ_X and γ_Y.

(ii) There are sequences of integers (l_n) and (r_n) such that

$$l_n \to \infty,\ r_n \to \infty,\ \frac{l_n}{r_n} \to 0,\ \frac{r_n}{n} \to 0 \text{ and } \frac{n\, b(l_n)}{r_n} \to 0 \quad \text{as } n \to \infty.$$

(iii) For any $t \geq 1$, there is a function r_t on $[0,\infty]^4 \setminus \{(\infty,\infty,\infty,\infty)\}$ such that

$$\begin{aligned} &\lim_{s\to\infty} s\, \mathbb{P}\left(\overline{F}_X(X_1) \leq \frac{x_1}{s},\ \overline{F}_X(X_{t+1}) \leq \frac{x_{t+1}}{s},\ \overline{F}_Y(Y_1) \leq \frac{y_1}{s},\ \overline{F}_Y(Y_{t+1}) \leq \frac{y_{t+1}}{s} \right) \\ = \;& r_t(x_1, x_{t+1}, y_1, y_{t+1}) \ \text{ for any } (x_1, x_{t+1}, y_1, y_{t+1}) \in [0,\infty]^4 \setminus \{(\infty,\infty,\infty,\infty)\}. \end{aligned}$$

(iv) There exist $D \geq 0$ and a nonnegative sequence $\rho(t)$ satisfying $\sum_{t\geq 1} \rho(t) < \infty$ such that we have, if s is large enough,

$$\begin{aligned} & s\, \mathbb{P}\left(\overline{F}_X(X_1) \leq \frac{x_1}{s},\ \overline{F}_X(X_{t+1}) \leq \frac{x_{t+1}}{s},\ \frac{u'}{s} < \overline{F}_Y(Y_1) \leq \frac{u}{s},\ \frac{v'}{s} < \overline{F}_Y(Y_{t+1}) \leq \frac{v}{s} \right) \\ \leq \;& \rho(t)\sqrt{\min(x_1, u-u')\min(x_{t+1}, v-v')} + \frac{D}{s}\min(x_1, u-u')\min(x_{t+1}, v-v'), \end{aligned}$$

for any $t \geq 1$, all $x_1, x_{t+1} \in [0,\infty]$, and all $u, u', v, v' \in [0,1]$ with $u' < u$ and $v' < v$.

If the bivariate time series (X_t, Y_t) satisfies Condition D, then the univariate time series (Y_t) automatically satisfies Conditions A and B (by taking $x_1 = x_{t+1} = \infty$, the function $R_t \equiv R_{Y,t}$ is obtained as $R_{Y,t}(y_1, y_{t+1}) := r_t(\infty,\infty,y_1,y_{t+1})$). In this sense, Condition D is a sensible yet novel generalisation of our setup of Section 3 to the bivariate time series case.

We now introduce two estimators of the QMES and XMES at an extreme level $\tau_n' \to 1$ satisfying $n(1-\tau_n') \to c \in [0,\infty)$ as $n \to \infty$. The estimators are defined by

$$\begin{aligned} \widehat{\text{QMES}}^{\star}_{X,\tau_n'} \equiv \widehat{\text{QMES}}^{\star}_{X,\tau_n'}(\tau_n) &:= \left(\frac{1-\tau_n'}{1-\tau_n}\right)^{-\widehat{\gamma}_{X,n}} \frac{\sum_{t=1}^n X_t \mathbb{1}\{X_t > 0, Y_t > \widehat{q}_{Y,\tau_n}\}}{\sum_{t=1}^n \mathbb{1}\{Y_t > \widehat{q}_{Y,\tau_n}\}} \\ \text{and } \ \widetilde{\text{XMES}}^{\star}_{X,\tau_n'} \equiv \widetilde{\text{XMES}}^{\star}_{X,\tau_n'}(\tau_n) &:= \left(\frac{1-\tau_n'}{1-\tau_n}\right)^{-\widehat{\gamma}_{X,n}} \frac{\sum_{t=1}^n X_t \mathbb{1}\{X_t > 0, Y_t > \widetilde{\xi}_{Y,\tau_n}\}}{\sum_{t=1}^n \mathbb{1}\{Y_t > \widetilde{\xi}_{Y,\tau_n}\}}. \end{aligned}$$

The construction of $\widehat{\mathrm{QMES}}^{\star}_{X,\tau_n'}$ and $\widetilde{\mathrm{XMES}}^{\star}_{X,\tau_n'}$ is based on the extrapolation relationships

$$\frac{\mathrm{QMES}_{X,\tau_n'}}{\mathrm{QMES}_{X,\tau_n}} \approx \frac{U_X((1-\tau_n')^{-1})}{U_X((1-\tau_n)^{-1})} \quad \text{and} \quad \frac{\mathrm{XMES}_{X,\tau_n'}}{\mathrm{XMES}_{X,\tau_n}} \approx \frac{U_X(1/\overline{F}_Y(\xi_{Y,\tau_n'}))}{U_X(1/\overline{F}_Y(\xi_{Y,\tau_n}))}$$

and the positive extremal dependence between X and Y. The QMES and XMES of (6) and (7) can actually be embedded in a more general MES framework. Define

$$\mathrm{MES}_{X,\tau} = \mathbb{E}(X|Y > z_{Y,\tau}), \ \tau \in (0,1),$$

where $z_{Y,\tau}$ is a risk measure on Y such that $\overline{F}(z_{Y,\tau})/(1-\tau) \to z = z(\gamma_Y) \in (0,\infty)$ as $\tau \uparrow 1$. Assuming that a $\sqrt{n(1-\tau_n)}$−relatively consistent estimator $\overline{z}_{Y,\tau_n}$ of z_{Y,τ_n} is available (at the intermediate level τ_n), then one may define an estimator of $\mathrm{MES}_{X,\tau_n'} = \mathbb{E}(X|Y > z_{Y,\tau_n'})$ by

$$\overline{\mathrm{MES}}^{\star}_{X,\tau_n'} \equiv \overline{\mathrm{MES}}^{\star}_{X,\tau_n'}(\tau_n) := \left(\frac{1-\tau_n'}{1-\tau_n}\right)^{-\widehat{\gamma}_{X,n}} \frac{\sum_{t=1}^n X_t \mathbb{1}\{X_t > 0, Y_t > \overline{z}_{Y,\tau_n}\}}{\sum_{t=1}^n \mathbb{1}\{Y_t > \overline{z}_{Y,\tau_n}\}}.$$

This does not seem to have been appreciated in earlier literature. To quantify the bias incurred in the construction of this estimator, the following bias condition is key.

Condition E. Assume that Conditions D(i) and D(iii) hold, and that there exist $\beta > \gamma_X$ and $\kappa < 0$ such that the function $R_{(X,Y)}$ defined by $R_{(X,Y)}(x,y) := r_1(x,\infty,y,\infty)$ satisfies $R_{(X,Y)}(1,1) > 0$ and

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

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

This is another version, compatible with our assumptions, of condition (a) in Cai et al. (2015) and condition $\mathcal{JC}_2(R,\beta,\kappa)$ in Daouia et al. (2018), under which extrapolating estimators of $\mathrm{QMES}_{X,\tau_n'}$ and $\mathrm{XMES}_{X,\tau_n'}$ converge to a Gaussian distribution in the i.i.d. data setting. The following generic theorem gives high-level conditions to obtain the asymptotic distribution of $\overline{\mathrm{MES}}^{\star}_{X,\tau_n'}$.

Theorem 4.1. *Suppose that $X := X_1$ and $Y := Y_1$ satisfy Condition C with respective parameters (γ_X, ρ_X, A_X) and (γ_Y, ρ_Y, A_Y), and that Conditions D and E hold. Suppose also that $\rho_X < 0$, and that there is $\delta > 0$ such that $0 < \gamma_X < 1/(2+\delta)$. Assume further that*

(i) $\tau_n, \tau_n' \uparrow 1$, with $n(1-\tau_n) \to \infty$, $n(1-\tau_n') \to c < \infty$ and $\sqrt{n(1-\tau_n)}/\log[(1-\tau_n)/(1-\tau_n')] \to \infty$ as $n \to \infty$;

(ii) $r_n(1-\tau_n) \to 0$ and $r_n(r_n/\sqrt{n(1-\tau_n)})^\delta \to 0$ as $n \to \infty$;

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

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

(v) The following bias conditions hold:

$$\sqrt{n(1-\tau_n)}\left(\frac{\overline{F}_Y(z_{Y,\tau_n})}{1-\tau_n} - z\right) = \mathrm{o}(1) \quad \textit{and} \quad \sqrt{n(1-\tau_n)}\left(\frac{\overline{F}_Y(z_{Y,\tau_n'})}{1-\tau_n'} - z\right) = \mathrm{o}(1);$$

(vi) One has

$$\sqrt{n(1-\tau_n)}\left(\frac{\overline{z}_{Y,\tau_n}}{z_{Y,\tau_n}} - 1\right) = \mathrm{O}_{\mathbb{P}}(1) \quad \textit{and} \quad \sqrt{n(1-\tau_n)}(\widehat{\gamma}_{X,n} - \gamma_X) \xrightarrow{d} \Gamma,$$

where Γ *is a nondegenerate random variable.*

If in addition $\sqrt{n(1-\tau_n)}A_Y((1-\tau_n)^{-1}) \to 0$ *as* $n \to \infty$*, then*

$$\frac{\sqrt{n(1-\tau_n)}}{\log[(1-\tau_n)/(1-\tau_n')]}\left(\frac{\overline{\mathrm{MES}}^{\star}_{X,\tau_n'}}{\mathrm{MES}_{X,\tau_n'}} - 1\right) \xrightarrow{d} \Gamma.$$

[Condition (iv) above is unnecessary if $X > 0$ *with probability 1.]*

Condition (i) is already standard in the i.i.d. case and is assumed in Corollary 3.2, for instance. Condition (ii) is used to deal with serial dependence, as in Theorem 3.1. Condition (iii) is a slightly weaker assumption than condition (d) of Cai et al. (2015), while Condition (iv) is taken from Theorem 2 therein. Condition (v) is a bias condition used to control the error made in the use of the extrapolation relationship for the construction of $\overline{\mathrm{MES}}^{\star}_{X,\tau_n'}$, while Condition (vi) ensures that all estimators appearing in the construction of this estimator converge at the appropriate rate.

Using Theorem 4.1, we obtain the following three important corollaries on our QMES and XMES estimators. We state first a corollary on QMES estimation at extreme levels.

Corollary 4.2. *Work under the conditions of Theorem 4.1 (apart from (v) and (vi)), assume also that* $r_n \log^2(n(1-\tau_n))/\sqrt{n(1-\tau_n)} \to 0$ *as* $n \to \infty$ *and let* $\widehat{\gamma}_{X,n} = \widehat{\gamma}^H_{X,n}$ *be the Hill estimator of* X*. Then, letting* $R_{X,t}(x_1, x_{t+1}) := r_t(x_1, x_{t+1}, \infty, \infty)$*,*

$$\frac{\sqrt{n(1-\tau_n)}}{\log[(1-\tau_n)/(1-\tau_n')]}\left(\frac{\widehat{\mathrm{QMES}}^{\star}_{X,\tau_n'}}{\mathrm{QMES}_{X,\tau_n'}} - 1\right) \xrightarrow{d} \mathcal{N}\left(0, \gamma_X^2\left[1 + 2\sum_{t=1}^{\infty} R_{X,t}(1,1)\right]\right).$$

Under further assumptions that ensure the asymptotic normality of the LAWS extreme expectile estimator, we also obtain a corollary on the asymptotic normality of the LAWS-based extreme expectile estimator $\widetilde{\mathrm{XMES}}^{\star}_{X,\tau_n'}$.

Corollary 4.3. *Work under the conditions of Theorem 4.1 (apart from (v) and (vi)) and let $\widehat{\gamma}_{X,n} = \widehat{\gamma}^H_{X,n}$ be the Hill estimator of X. Suppose further that, with the notation of Theorem 4.1, $\mathbb{E}|Y_-|^{2+\delta} < \infty$ and $0 < \gamma_Y < 1/(2+\delta)$, that $\sum_{l\geq 1}[b(l)]^{\delta/(2+\delta)} < \infty$, and that $r_n \log^2(n(1-\tau_n))/\sqrt{n(1-\tau_n)} \to 0$ and $\sqrt{n(1-\tau_n)}q^{-1}_{Y,\tau_n} \to 0$ as $n \to \infty$. Then, letting $R_{X,t}(x_1, x_{t+1}) := r_t(x_1, x_{t+1}, \infty, \infty)$,*

$$\frac{\sqrt{n(1-\tau_n)}}{\log[(1-\tau_n)/(1-\tau_n')]}\left(\frac{\widetilde{\mathrm{XMES}}^{\star}_{X,\tau_n'}}{\mathrm{XMES}_{X,\tau_n'}} - 1\right) \xrightarrow{d} \mathcal{N}\left(0, \gamma_X^2\left[1 + 2\sum_{t=1}^{\infty} R_{X,t}(1,1)\right]\right).$$

We finally have the following asymptotic normality result on the indirect, QB estimator of the XMES defined as

$$\widehat{\mathrm{XMES}}^{\star}_{X,\tau_n'} := (\widehat{\gamma}^{-1}_{Y,n} - 1)^{-\widehat{\gamma}_{X,n}}\widehat{\mathrm{QMES}}^{\star}_{X,\tau_n'}$$

where $\widehat{\gamma}_{Y,n}$ is a $\sqrt{n(1-\tau_n)}$−consistent estimator of γ_Y.

Corollary 4.4. *Work under the conditions of Theorem 4.1 (apart from (v) and (vi)) and let $\widehat{\gamma}_{X,n} = \widehat{\gamma}^H_{X,n}$ be the Hill estimator of X. Suppose further that, with the notation of Theorem 4.1, $\mathbb{E}|Y_-| < \infty$ and $0 < \gamma_Y < 1$, and that $r_n \log^2(n(1-\tau_n))/\sqrt{n(1-\tau_n)} \to 0$ and $\sqrt{n(1-\tau_n)}q^{-1}_{Y,\tau_n} \to 0$ as $n \to \infty$. Assume finally that $\sqrt{n(1-\tau_n)}(\widehat{\gamma}_{Y,n} - \gamma_Y) = \mathrm{O}_{\mathbb{P}}(1)$. Then, letting $R_{X,t}(x_1, x_{t+1}) := r_t(x_1, x_{t+1}, \infty, \infty)$,*

$$\frac{\sqrt{n(1-\tau_n)}}{\log[(1-\tau_n)/(1-\tau_n')]}\left(\frac{\widehat{\mathrm{XMES}}^{\star}_{X,\tau_n'}}{\mathrm{XMES}_{X,\tau_n'}} - 1\right) \xrightarrow{d} \mathcal{N}\left(0, \gamma_X^2\left[1 + 2\sum_{t=1}^{\infty} R_{X,t}(1,1)\right]\right).$$

These results extend Theorems 1 and 2 of Cai et al. (2015) and Theorems 4 and 5 of Daouia et al. (2018) to our time series context. Like Corollary 3.2, they provide the appropriate theory for the construction of asymptotic confidence intervals about extreme MES, which take into account the serial dependence of some of the typical time series contexts in financial applications.

5 Asymptotic confidence interval construction

On the basis of the theory developed in Sections 3 and 4, we propose here asymptotic confidence interval estimators for inferring extreme expectiles and XMES.

First, we recall that with the notation of Section 3, an estimator of the expectile of Y_1 at the extreme level τ_n' is given by

$$\overline{\xi}^{\star}_{\tau_n'} = \left(\frac{1-\tau_n'}{1-\tau_n}\right)^{-\widehat{\gamma}^H_n} \overline{\xi}_{\tau_n}, \tag{8}$$

where $\widehat{\gamma}^H_n$ is the Hill estimator of Y_1 and $\overline{\xi}_{\tau_n}$ is an estimator of the expectile of Y_1 at the intermediate level τ_n. Second, for the asymptotic variance of the estimator in (8), deduced in Theorem 3.5 and denoted hereafter by $w(\gamma, \boldsymbol{R})$, we propose the following estimator. By Proposition 2.1 in Drees (2003) we have, when $n(1-\tau_n) \to \infty$, $r_n \to \infty$ and $r_n(1-\tau_n) \to 0$,

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

We then adopt a "big-block/small-block" technique where we split the data into big blocks of size r_n separated by small blocks of size l_n, and we define $Z_j = \sum_{t=1+j\ell_n}^{r_n+j\ell_n} \mathbb{1}\{\widehat{F}_n(Y_t) > \tau_n\}$ for $j = 0, 1, \ldots, m_n - 1$, where $m_n = \lfloor n/\ell_n \rfloor$ and $\ell_n = r_n + l_n$, and $\widehat{F}_n$ is the empirical distribution function of all the observations. We compute the sample variance Σ_n of the sequence $(Z_0, \ldots, Z_{m_n-1})$ and obtain an empirical estimator of the asymptotic variance $w(\gamma, \boldsymbol{R}) := \gamma^2 \left[1 + 2\sum_{t=1}^{\infty} R_t(1,1)\right]$ as

$$\widehat{w}_n(\gamma, \boldsymbol{R}) = \frac{(\widehat{\gamma}_n^H)^2}{r_n(1-\tau_n)} \Sigma_n. \tag{9}$$

Next, note that Theorem 3.5 is equivalent to its (in practice more accurate) log-scale version

$$\frac{\sqrt{n(1-\tau_n)}}{\log[(1-\tau_n)/(1-\tau_n')]} \log \frac{\overline{\xi}^{\star}_{\tau_n'}}{\xi_{\tau_n'}} \xrightarrow{d} \mathcal{N}\left(\frac{\lambda_1}{1-\rho}, w(\gamma, \boldsymbol{R})\right) \quad \text{as } n \to \infty.$$

Hence, we propose the following interval estimator

$$\left[\overline{\xi}^{\star}_{\tau_n'} \left(\frac{1-\tau_n}{1-\tau_n'}\right)^{z_{\alpha/2}\sqrt{\widehat{w}_n(\gamma,\boldsymbol{R})/[n(1-\tau_n)]}}, \; \overline{\xi}^{\star}_{\tau_n'} \left(\frac{1-\tau_n}{1-\tau_n'}\right)^{z_{1-\alpha/2}\sqrt{\widehat{w}_n(\gamma,\boldsymbol{R})/[n(1-\tau_n)]}}\right], \tag{10}$$

where $\overline{\xi}^{\star}_{\tau_n'}$ is the estimator in (8), $\widehat{w}_n(\gamma, \boldsymbol{R})$ is the estimator in (9), $z_{\alpha/2}$ and $z_{1-\alpha/2}$ are the $(\alpha/2)$th and $(1-\alpha/2)$th quantiles of the standard normal distribution, with $\alpha \in (0,1)$. For simplicity we have ignored the bias term $\lambda_1/(1-\rho)$. We call the estimator in (10) the LAWS-D-based estimator when $\overline{\xi}_{\tau_n} = \widetilde{\xi}_{\tau_n}$ in (8), and QB-D-based estimator when $\overline{\xi}_{\tau_n} = \widehat{\xi}_{\tau_n}$. Finally, and for comparison purposes, the asymptotic variance in the i.i.d. case is γ^2, estimated by $(\widehat{\gamma}_n^H)^2$. In this case, an interval estimator is simply

$$\left[\overline{\xi}^{\star}_{\tau_n'} \left(\frac{1-\tau_n}{1-\tau_n'}\right)^{z_{\alpha/2}\widehat{\gamma}_n^H/\sqrt{n(1-\tau_n)}}, \; \overline{\xi}^{\star}_{\tau_n'} \left(\frac{1-\tau_n}{1-\tau_n'}\right)^{z_{1-\alpha/2}\widehat{\gamma}_n^H/\sqrt{n(1-\tau_n)}}\right] \tag{11}$$

and we call it the LAWS-IID-based estimator or QB-IID-based estimator depending on whether $\overline{\xi}_{\tau_n} = \widetilde{\xi}_{\tau_n}$ or $\overline{\xi}_{\tau_n} = \widehat{\xi}_{\tau_n}$.

A similar construction is of course feasible for XMES and is entirely similar, by substituting X_t for Y_t in both the Z_j and the Hill estimator, and by replacing $\overline{\xi}^{\star}_{\tau_n'}$ by either $\widetilde{\text{XMES}}^{\star}_{X,\tau_n'}$ or $\widehat{\text{XMES}}^{\star}_{X,\tau_n'}$, giving rise to respectively LAWS-D-based confidence intervals and QB-D-based confidence intervals. Their simpler i.i.d. counterparts are likewise called LAWS-IID-based confidence intervals and QB-IID-based confidence intervals.

6 Extreme expectile level selection

A crucial practical question in actuarial and financial risk management is the choice of the level of prudentiality of the risk measure under consideration. When working with quantile-based risk measures, one usually chooses extreme tail probabilities $\alpha_n \uparrow 1$ with $n(1-\alpha_n) \to c$, a finite constant, as $n \to \infty$, to allow for more prudent risk management. When expectiles are of interest, a reasonable idea is to, with the notation of Section 3, select τ_n' so that $\xi_{\tau_n'} \equiv q_{\alpha_n}$ for a given α_n. This was suggested by Bellini and Di Bernardino (2017) for normally distributed Y. For heavy-tailed and i.i.d. data, Daouia et al. (2018) instead suggested a nonparametric estimator of the level $\tau_n' = \tau_n'(\alpha_n)$ that satisfies $\xi_{\tau_n'} \equiv q_{\alpha_n}$. They find that this extreme expectile level satisfies $(1-\tau_n'(\alpha_n))/(1-\alpha_n) \to \gamma/(1-\gamma)$ as $n \to \infty$. If $\widehat{\gamma}_n$ is a consistent estimator of γ, one can then define a natural estimator of $\tau_n'(\alpha_n)$ as

$$\widehat{\tau}_n'(\alpha_n) = 1 - (1-\alpha_n)\,\frac{\widehat{\gamma}_n}{1-\widehat{\gamma}_n}.$$

By substituting this estimated value in place of τ_n' in the extrapolating LAWS estimator $\widetilde{\xi}^{\star}_{\tau_n'}$ and in the extrapolating QB estimator $\widehat{\xi}^{\star}_{\tau_n'}$, we obtain composite estimators of $\xi_{\tau_n'(\alpha_n)} \equiv q_{\alpha_n}$. It is interesting to note that if one uses the exact same estimator $\widehat{\gamma}_n$ in the extrapolation step and the calculation of $\widehat{\tau}_n'(\alpha_n)$, the composite extrapolating LAWS estimator is

$$\widetilde{\xi}^{\star}_{\widehat{\tau}_n'(\alpha_n)} = \left(\frac{1-\widehat{\tau}_n'(\alpha_n)}{1-\tau_n}\right)^{-\widehat{\gamma}_n} \widetilde{\xi}_{\tau_n} = (\widehat{\gamma}_n^{-1} - 1)^{\widehat{\gamma}_n}\,\widetilde{\xi}^{\star}_{\alpha_n}.$$

In other words, rewriting Equation (4) as $q_{\alpha_n} \equiv \xi_{\tau_n'(\alpha_n)} \approx (\gamma^{-1}-1)^{\gamma}\xi_{\alpha_n}$, the composite extrapolating LAWS estimator can be constructed by plugging in an estimator $\widehat{\gamma}_n$ and the extrapolating LAWS estimator at level α_n in the right-hand side of this approximation. This had not, to the best of our knowledge, been noted in the literature.

The available theory of the composite LAWS and QB estimators in Daouia et al. (2018) is limited to i.i.d. data. We give below a result showing their asymptotic normality in our dependent setting, when $\widehat{\gamma}_n$ is the Hill estimator.

Theorem 6.1. *Suppose the conditions of Theorem 3.5 hold with α_n in place of τ_n'. Then, if $\widehat{\gamma}_n = \widehat{\gamma}_n^H$ and $\overline{\xi}^{\star}$ is either $\widehat{\xi}^{\star}$ or $\widetilde{\xi}^{\star}$, we have*

$$\frac{\sqrt{n(1-\tau_n)}}{\log[(1-\tau_n)/(1-\alpha_n)]}\left(\frac{\overline{\xi}^{\star}_{\widehat{\tau}_n'(\alpha_n)}}{q_{\alpha_n}} - 1\right) \xrightarrow{d} \mathcal{N}\left(\frac{\lambda_1}{1-\rho}, \gamma^2\left[1 + 2\sum_{t=1}^{\infty} R_t(1,1)\right]\right).$$

An analogue construction is of course possible in the bivariate case and gives rise to composite estimators $\widetilde{\mathrm{XMES}}^{\star}_{X,\widehat{\tau}_n'(\alpha_n)}$ and $\widehat{\mathrm{XMES}}^{\star}_{X,\widehat{\tau}_n'(\alpha_n)}$ of $\mathrm{QMES}_{X,\alpha_n}$. The following result provides their asymptotic normality.

Theorem 6.2. *Suppose the conditions of Corollary 4.3 (resp. Corollary 4.4) hold with α_n in place of τ_n'. Then, if $\widehat{\gamma}_{X,n} = \widehat{\gamma}_{X,n}^H$ and $\overline{\mathrm{XMES}}^{\star} = \widetilde{\mathrm{XMES}}^{\star}$ (resp. $\overline{\mathrm{XMES}}^{\star} = \widehat{\mathrm{XMES}}^{\star}$), then*

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

7 Simulation experiments

7.1 Extreme expectile estimation

Here we investigate the finite-sample performance of the point and interval expectile estimators at extreme levels. We consider the following models:

(a) The AR(1) model $Y_{t+1} = 0.8\,Y_t + \varepsilon_{t+1}$, where the innovations ε_t are i.i.d. and have a common Student-t distribution with $\nu = 3$ degrees of freedom.

(b) The ARMA(1,1) model $Y_{t+1} = 0.95\,Y_t + \varepsilon_{t+1} + 0.9\,\varepsilon_t$, where the innovations ε_t are i.i.d. and have a common 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 (ε_t) is a sequence of i.i.d. standard Gaussian innovations.

(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 (ε_t) is a sequence of i.i.d. standard Gaussian innovations.

Strong linear dependence is present in models (a) and (b), while quadratic serial dependence is present in models (c) and (d). Recall that for standard linear time series with heavy-tailed innovations having balanced tails, which is the case for models (a) and (b), the tail index of the time series is the tail index of the innovations, so that the tail index of Y is 1/3 in models (a) and (b). In models (c) and (d), the marginal distribution is heavy-tailed, and the value of the tail index can be calculated numerically using *e.g.* Theorem 2.1 in Mikosch and Stărică (2000). It is found to be 0.262 and 0.239 in models (c) and (d) respectively.

For each model, we simulate 10^4 samples of size $n = 2500$ and consider the extreme level $\tau_n' = 0.9995 \approx 1 - 1/n$. On each simulated dataset, we calculate the extrapolating LAWS and QB estimators using the intermediate level $\tau_n = 1 - k/n$ for $k \in \{6, 8, \ldots, 700\}$. Then, we estimate an asymptotic confidence interval using the LAWS-D-based interval estimator in (10), with 95% nominal coverage probability. 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 value of the sample autocorrelation is small, *e.g.* smaller than 0.1. Checking if it contains the true expectile value allows us to compute a Monte Carlo approximation of the coverage probability. We do the same exercise for the QB-D-based, QB-IID-based and LAWS-IID-based procedures. Results are reported in Figure 1 (results at level $\tau_n' = 0.9999 > 1 - 1/n$ are similar).

It is readily seen that our proposed confidence intervals, derived from the theory of serial dependent data, behave far better in terms of coverage than the previously available i.i.d.-based intervals that are overall far too permissive. Even in the difficult ARCH and GARCH cases, our intervals represent significant improvement, by bringing down the non-coverage probability by about a half to around 15% in the ARCH cases, and by getting very close to the nominal level when k is in a neighbourhood of approximatively 60.

7.2 Extreme MES estimation

We consider the finite-sample performance of the point and interval MES estimators at extreme levels. We work with four models derived from models (a)-(d) in Section 7.1.

(e) $X_{t+1} = 0.8\, X_t + \varepsilon_{X,t+1}$ and $Y_{t+1} = 0.8\, Y_t + \varepsilon_{Y,t+1}$ where the pairs of innovations $(\varepsilon_{X,t}, \varepsilon_{Y,t})$ are i.i.d. For any t, the innovation $\varepsilon_{X,t}$ is distributed as $Z\mathbb{1}\{Z > 0\} - \sqrt{-Z}\mathbb{1}\{Z < 0\}$, where Z has a Student-t distribution with 3 degrees of freedom, and $\varepsilon_{Y,t}$ is itself Student-t distributed with 3 degrees of freedom. The dependence structure of the pair $(\varepsilon_{X,t}, \varepsilon_{Y,t})$ is given by a Student-t copula with correlation parameter $\rho = 0.8$ and 3 degrees of freedom.

(f) $X_{t+1} = 0.95\, X_t + \varepsilon_{X,t+1} + 0.9\, \varepsilon_{X,t}$, and $Y_{t+1} = 0.95\, Y_t + \varepsilon_{Y,t+1} + 0.9\, \varepsilon_{Y,t}$, where the pairs of innovations $(\varepsilon_{X,t}, \varepsilon_{Y,t})$ are i.i.d. For any t, the innovation $\varepsilon_{X,t}$ is distributed as $Z\mathbb{1}\{Z > 0\} - \sqrt{-Z}\mathbb{1}\{Z < 0\}$, where Z has a symmetric Pareto distribution with shape parameter $\zeta = 3$, and $\varepsilon_{Y,t}$ is itself symmetric Pareto distributed with shape parameter $\zeta = 3$. The dependence structure of the pair $(\varepsilon_{X,t}, \varepsilon_{Y,t})$ is given by a Gumbel copula with parameter $\theta = 2$.

(g) $X_{t+1} = \sigma_{X,t+1}\varepsilon_{X,t+1}$, where $\sigma^2_{X,t+1} = 0.4 + 0.6\, X_t^2$, and $Y_{t+1} = \sigma_{Y,t+1}\varepsilon_{Y,t+1}$, where $\sigma^2_{Y,t+1} = 0.4 + 0.6\, Y_t^2$, where the pairs of innovations $(\varepsilon_{X,t}, \varepsilon_{Y,t})$ are i.i.d. For any t, the innovation $\varepsilon_{X,t}$ has density $h(z) = 0.5\, \mathbb{1}\{-1 < z \le 0\} + 0.5\, e^{-z}\mathbb{1}\{z > 0\}$ and $\varepsilon_{Y,t}$ is standard Gaussian. The dependence structure of the pair $(\varepsilon_{X,t}, \varepsilon_{Y,t})$ is given by a Student-t copula with correlation parameter $\rho = 0.8$ and 3 degrees of freedom.

(h) $X_{t+1} = \sigma_{X,t+1}\varepsilon_{X,t+1}$, where $\sigma^2_{X,t+1} = 0.1 + 0.4\, X_t^2 + 0.4\, \sigma^2_{X,t}$, and $Y_{t+1} = \sigma_{Y,t+1}\varepsilon_{Y,t+1}$, where $\sigma^2_{Y,t+1} = 0.1 + 0.4\, Y_t^2 + 0.4\, \sigma^2_{Y,t}$, where the pairs of innovations $(\varepsilon_{X,t}, \varepsilon_{Y,t})$ are i.i.d. For any t, the innovation $\varepsilon_{X,t}$ has density $h(z) = 0.5\, \mathbb{1}\{-1 < z \le 0\} + 0.5\, e^{-z}\mathbb{1}\{z > 0\}$ and $\varepsilon_{Y,t}$ is standard Gaussian. The dependence structure of the pair $(\varepsilon_{X,t}, \varepsilon_{Y,t})$ 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 that the models considered here extend those of Section 7.1.

We simulate 10^4 samples of size $n = 2500$ for each model. Differently to Section 7.1, we consider the problem of estimating $\text{QMES}(\alpha_n)$ at levels α_n such that $\tau_n'(\alpha_n) = 0.9995$ and 0.9999, using our composite estimators $\widetilde{\text{XMES}}^{\star}_{X,\widehat{\tau}_n'(\alpha_n)}$ and $\widehat{\text{XMES}}^{\star}_{X,\widehat{\tau}_n'(\alpha_n)}$. In each model, the level α_n is first found theoretically using the asymptotic proportionality relationship $(1-\alpha_n) \approx (\gamma_Y^{-1} - 1)(1 - \tau_n'(\alpha_n))$ of Section 6. The true value of $\text{QMES}(\alpha_n) \equiv \text{XMES}(\tau_n'(\alpha_n))$ is determined by intensive Monte-Carlo simulation, with the true value of $\xi_{Y,\tau_n'(\alpha_n)} \equiv q_{Y,\alpha_n}$ taken from our earlier investigations in Section 7.1. Estimating a QMES at this level α_n allows us to get an idea of the performance of our composite QMES estimation method, in a problem comparable in difficulty to that of Section 7.1 (as far as the Y component is concerned). The construction of our confidence intervals follows the procedure described in Section 7.1, applied to the X_t component. In the QB composite estimator $\widehat{\text{XMES}}^{\star}_{X,\widehat{\tau}_n'(\alpha_n)}$, the Hill estimators of γ_X and γ_Y are used, at the same level k for the sake of simplicity.

Results are reported in Figure 1 at level $\tau_n'(\alpha_n) = 0.9995$ (results at level $\tau_n'(\alpha_n) = 0.9999$ are broadly similar). We see again with these results that our proposed confidence intervals are overall satisfactory, especially compared to the IID-based intervals, whose non-coverage probability is high in each case. Even in the substantially more difficult ARCH and GARCH cases, our intervals represent significant improvement compared to the i.i.d. method. We have also observed during our simulation experiments that our estimation procedure appears to be much more accurate than that of Daouia et al. (2018), whose bias and MSE were disturbingly high (see Figures 5 and 6 therein).

8 Real data analysis

Stock market index data. We consider here daily negative log-returns of the S&P 500 (GSPC) and Dow Jones Industrial Average (DJIA) indices from January 29, 1985 to December 12, 2019 (available from Yahoo! Finance). These samples of size $n = 8{,}785$ are plotted on the rightmost panels of Figure 2. They show evidence of the stylised facts such as heteroscedasticity and fat-tailedness found in financial time series (Embrechts et al., 1997). Throughout this section, we use the Hill estimator in our estimates and confidence intervals.

In the literature, the analysis of tail risk of loss returns is typically based on VaR at the 99.9% level (see *e.g.* Drees, 2003; de Haan et al., 2016) or on using a quantile at level $\alpha_n = 1 - p_n$ where p_n is not larger than $1/n$. Bellini and Di Bernardino (2017) showed that expectile-based forecasts provide capital requirements similar to those obtained with VaR as long as the extreme level τ_n' of the expectile is appropriately selected, as a larger value than the level of the pre-specified extreme VaR. For this selection, we use the method of Section 6 which, contrary to the method of Bellini and Di Bernardino (2017), does not depend on a specific parametric model. Precisely, here we fix $p_n = 1/n$ and $\alpha_n = 1 - p_n = 0.9998862$, and according to Section 6 we estimate first $\tau_n'(\alpha_n) = \tau_n'(0.9998862)$. Then we compute an estimate of the expectile at the extreme level $\widehat{\tau}_n'(\alpha_n)$ using the composite extrapolating LAWS estimator $\widetilde{\xi}^{\star}_{\tau_n'}$ and the composite extrapolating QB estimator $\widehat{\xi}^{\star}_{\tau_n'}$ for $\tau_n' = \widehat{\tau}_n'(\alpha_n)$. This accordingly produces estimators of $\xi_{\tau_n'(\alpha_n)}$, which is also the quantile-VaR $q_{\alpha_n} = q_{1-1/n}$. Let us reiterate here that our focus is the estimation of an extreme marginal expectile, rather than the completely different problem of dynamically predicting extreme expectiles.

The first panels of Figure 2 display $\widehat{\tau}_n'(\alpha_n)$ against k for $k \leq 700$, where as before $\tau_n = 1 - k/n$. After large fluctuations, the estimates stabilise around a (close for both series) common value, and then drift away due to bias from the centre of the distribution. The choice $k = 200$ seems to be a reasonable compromise. We further confirm this choice by calculating the composite extrapolating LAWS and composite extrapolating QB estimators, and the corresponding LAWS-D and QB-D confidence intervals of Section 5 at level $\tau_n' = \widehat{\tau}_n'(\alpha_n)$. The choice $k = 200$ once again seems sensible and is therefore adopted. It can be seen that our confidence intervals taking the dependence into account are indeed wider than the i.i.d.-based estimators, as a way to reflect better the uncertainty about the estimation. With $k = 200$, we find $\widehat{\tau}_n'(\alpha_n) \approx 0.9999423$ for the S&P 500 data, and 0.9999402 for the Dow Jones data. These levels are indeed larger than the original $\alpha_n = 1 - p_n = 0.9998862$.

We compare our extrapolating LAWS and QB methods with the traditional Weissman extreme quantile estimator at level α_n. This is

$$\widehat{q}^{\star}_{\alpha_n} = \left(\frac{1-\alpha_n}{1-\tau_n}\right)^{-\widehat{\gamma}^H_n} \widehat{q}_{\tau_n} = \left(\frac{1-\alpha_n}{1-\tau_n}\right)^{-\widehat{\gamma}^H_n} X_{n-\lfloor n(1-\tau_n)\rfloor,n}.$$

Confidence intervals can also be constructed for the extreme quantile q_{α_n} using this estimator: here we use a method developed by Drees (2003) which, contrary to ours, does not rely on a big-block/small-block argument, see Formula (33) therein. The estimates are reported in Table 1 and can be visualised in the rightmost panel of Figure 2. Quite reassuringly the methods give point estimates that are similar: note that the composite extrapolating QB point estimator is indeed nothing but $\widehat{q}^{\star}_{\alpha_n}$. However, on the third and fourth panels of Figure 2, it can be seen that the confidence intervals constructed on the basis of $\widehat{q}^{\star}_{\alpha_n}$ and the method of Drees (2003) are in general much more volatile than the LAWS-D-based interval; moreover, in a neighbourhood of our selected value of k, they are very close to the intervals based on i.i.d. theory. For our selected $k = 200$, according to the fourth panels of Figure 2, they do not contain the maximum observation in the sample, even though one is estimating $q_{\alpha_n} = q_{1-1/n}$, whereas the LAWS-D-based interval does contain this maximum value.

Financial returns of individual banks. We carry out an analysis of 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 July 3, 2000, to June 30, 2010, along with, for the same time period, daily loss returns (Y_t) of a value-weighted market index aggregating three markets: the New York Stock Exchange, American Express Stock Exchange and the National Association of Securities Dealers Automated Quotation system. These samples of data, of size $n = 2{,}513$, were already considered in Cai et al. (2015) and Daouia et al. (2018). Choosing $\alpha_n = 1 - 1/n = 0.9996021$, we display $\widehat{\tau}'_n(\alpha_n)$ against k, as well as the composite extrapolating LAWS and composite extrapolating QB estimators of $\mathrm{QMES}_{X,\alpha_n}$, and the corresponding LAWS-D and QB-D confidence intervals at level $\tau'_n = \widehat{\tau}'_n(\alpha_n)$, on the first three panels of Figure 3. The choice $k = 150$ seems reasonable here and we then find $\widehat{\tau}'_n(\alpha_n) \approx 0.9997239$ for the Goldman Sachs data, and 0.9996626 for the Morgan Stanley data. We compare our estimates to the Weissman-type extreme QMES estimator $\widehat{\mathrm{QMES}}^{*}_{X,\alpha_n}$. All estimates are reported in Table 2 and the LAWS estimates can be visualised in the rightmost panels of Figure 3. Like on our stock market index data, it can be seen that the confidence intervals constructed with the method of Drees (2003) are more volatile than ours. In addition, for our selected value of $k = n(1-\tau_n)$, they give similar results on the Goldman Sachs data, and somewhat shorter confidence intervals on the Morgan Stanley data.

Table 1: Estimates for the negative daily log-returns of the S&P 500 and Dow Jones indices, obtained with $k = 200$. Here $\alpha_n = 1 - 1/n = 0.9998862$.

Estimator	S&P 500	Dow Jones
$\widehat{\gamma}_n = \widehat{\gamma}_n^H$	0.3364 [0.2198, 0.4530]	0.3442 [0.2219, 0.4665]
$\widetilde{\xi}^{\star}_{\widehat{\tau}_n'(\alpha_n)}$	0.1358 [0.0676, 0.2727]	0.1360 [0.0657, 0.2813]
$\widehat{\xi}^{\star}_{\widehat{\tau}_n'(\alpha_n)}$	0.1398 [0.0696, 0.2807]	0.1394 [0.0674, 0.2884]
$\widehat{q}^{\star}_{\alpha_n}$	0.1398 [0.1124, 0.1739]	0.1394 [0.1025, 0.1896]

Table 2: Estimates for the loss returns of Goldman Sachs and Morgan Stanley, obtained with $k = 150$. Here $\alpha_n = 1 - 1/n = 0.9996021$.

Estimator	Goldman Sachs	Morgan Stanley
$\widehat{\gamma}_{X,n} = \widehat{\gamma}_{X,n}^H$	0.4096 [0.2815, 0.5377]	0.4589 [0.2966, 0.6212]
$\widetilde{\mathrm{XMES}}^{\star}_{X,\widehat{\tau}_n'(\alpha_n)}$	0.3419 [0.1705, 0.6853]	0.5901 [0.2445, 1.4238]
$\widehat{\mathrm{XMES}}^{\star}_{X,\widehat{\tau}_n'(\alpha_n)}$	0.3448 [0.1720, 0.6912]	0.6032 [0.2500, 1.4554]
$\widehat{\mathrm{QMES}}^{\star}_{X,\alpha_n}$	0.3448 [0.2158, 0.5509]	0.6032 [0.2944, 1.2359]

Acknowledgements

This research was in large part carried out when the authors were visiting each other at Bocconi University and the University of Nottingham where G. Stupfler was previously based. Support from the Nottingham PEF Fund is gratefully acknowledged. We would also like to thank Holger Drees for a detailed explanation of some of the technical details in Drees (2003) and Drees and Rootzén (2010). S.A. Padoan is supported by the Bocconi Institute for Data Science and Analytics (BIDSA), Italy. G. Stupfler is supported by the French National Research Agency under the grant ANR-19-CE40-0013-01/ExtremReg project.

References

Acharya, V.V., Pedersen, L.H., Philippon, T. and Richardson, M. (2017). Measuring systemic risk. *The Review of Financial Studies* **30**, 2–47.

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

Beirlant, J., Goegebeur, Y., Segers, J. and Teugels, J. (2004). *Statistics of Extremes: Theory and Applications*. John Wiley & Sons, Chichester.

Bellini, F. and Di Bernardino, E. (2017). Risk management with expectiles. *The European Journal of Finance* **23**, 487–506.

Bellini, F., Klar, B., Müller, A. and Gianin, E.R. (2014). Generalized quantiles as risk measures. *Insurance: Mathematics and Economics* **54**, 41–48.

Boussama, F., Fuchs, F. and Stelzer, R. (2011). Stationarity and geometric ergodicity of BEKK multivariate GARCH models. *Stochastic Processes and their Applications* **121**, 2331–2360.

Brownlees, C.T. and 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. and 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. and Guillou, A. (2018). Extreme quantile estimation for β-mixing time series and applications. *Insurance: Mathematics and Economics* **83**, 59–74.

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

Daouia, A., Girard, S. and Stupfler, G. (2019). Extreme M-quantiles as risk measures: From L^1 to L^p optimization. *Bernoulli* **25**, 264–309.

Daouia, A., Girard, S. and Stupfler, G. (2020). Tail expectile process and risk assessment. *Bernoulli* **26**, 531–556.

Doukhan, P. (1994). *Mixing: Properties and Examples.* Springer, New York.

Drees, H. (2000). Weighted approximations of tail processes for β-mixing random variables. *Annals of Applied Probability* **10**, 1274–1301.

Drees, H. (2002). Tail empirical processes under mixing conditions. *In Dehling, H.G., Mikosch, T. and Sørensen, M. (eds.), Empirical Process Techniques for Dependent Data.* Birkhäuser, Boston.

Drees, H. (2003). Extreme quantile estimation for dependent data, with applications to finance. *Bernoulli* **9**, 617–657.

Drees, H. and Rootzén, H. (2010). Limit theorems for empirical processes of cluster functionals. *Annals of Statistics* **38**, 2145–2186.

Ehm, W., Gneiting, T., Jordan, A. and Krüger, F. (2016). Of quantiles and expectiles: consistent scoring functions, Choquet representations and forecast rankings (with discussion). *Journal of the Royal Statistical Society: Series B* **78**, 505–562.

Embrechts, P., Klüppelberg, C. and Mikosch, T. (1997). *Modelling Extremal Events for Insurance and Finance*. Springer-Verlag, Berlin.

Engle, R.F., Jondeau, E. and Rockinger, M. (2015). Systemic risk in Europe. *Review of Finance* **19**, 145–190.

Francq, C. and 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.

de Haan, L. and Ferreira, A. (2006). *Extreme Value Theory: An Introduction.* Springer-Verlag, New York.

de Haan, L., Mercadier, C. and Zhou, C. (2016). Adapting extreme value statistics to financial time series: dealing with bias and serial dependence. *Finance and Stochastics* **20**, 321–354.

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

Holzmann, H. and Klar, B. (2016). Expectile asymptotics. *Electronic Journal of Statistics* **10**, 2355–2371.

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

Krätschmer, V. and Zähle, H. (2017). Statistical inference for expectile-based risk measures. *Scandinavian Journal of Statistics* **44**, 425–454.

Kuan, C.-M., Yeh, J.-H. and Hsu, Y.-C. (2009). Assessing value at risk with CARE, the Conditional Autoregressive Expectile models. *Journal of Econometrics* **150**, 261–270.

Mikosch, T. and 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. and Powell, J.L. (1987). Asymmetric least squares estimation and testing. *Econometrica* **55**, 819–847.

Rio, E. (2017). Asymptotic Theory of Weakly Dependent Random Processes. *In Probability Theory and Stochastic Modelling*, volume 80. Springer-Verlag, Berlin-Heidelberg.

Rootzén, H., Leadbetter, M.R. and de Haan, L. (1998). On the distribution of tail array sums for strongly mixing stationary sequences. *The Annals of Applied Probability* **8**, 868–885.

Sobotka, F. and Kneib, T. (2012). Geoadditive expectile regression. *Computational Statistics & Data Analysis* **56**, 755–767.

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

Ziegel, J.F. (2016). Coherence and elicitability. *Mathematical Finance* **26**, 901–918.

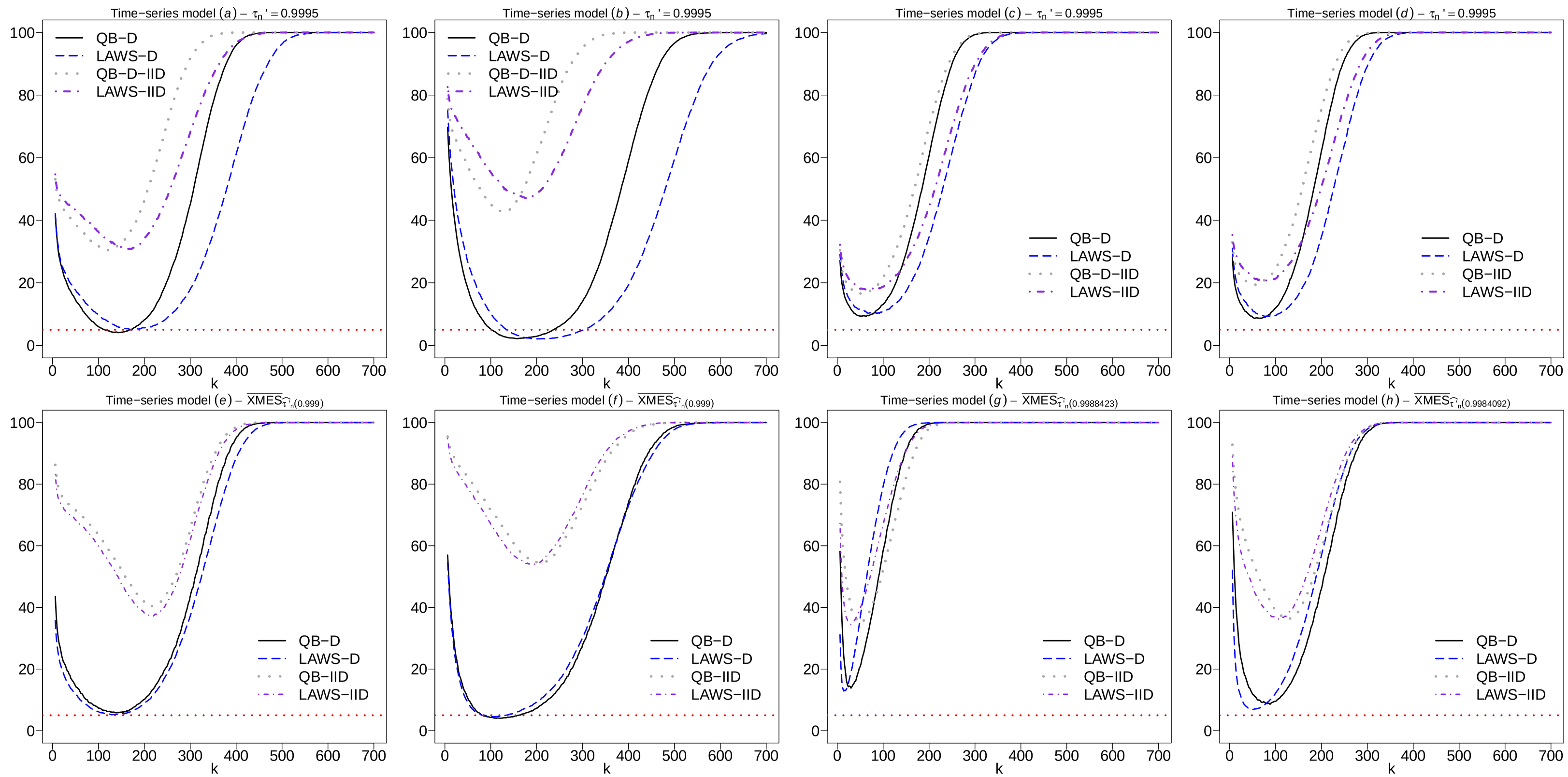

Figure 1: Actual non-coverage probabilities (in %) for the confidence intervals of the expectile $\xi_{\tau_n'}$ and the Marginal Expected Shortfall QMES(α_n), with 95% nominal level. From left to right, top row: models (a)-(d) and $\tau_n' = 0.9995$, bottom row: models (e)-(h) and level α_n such $\tau_n'(\alpha_n) = 0.9995$. In each panel, the horizontal dotted red line represents the 5% nominal non-coverage probability, and the solid black, blue dashed, dotted grey and dashed-dotted violet lines represent the actual non-coverage probabilities of the QB-D, LAWS-D, QB-IID and LAWS-IID confidence intervals.

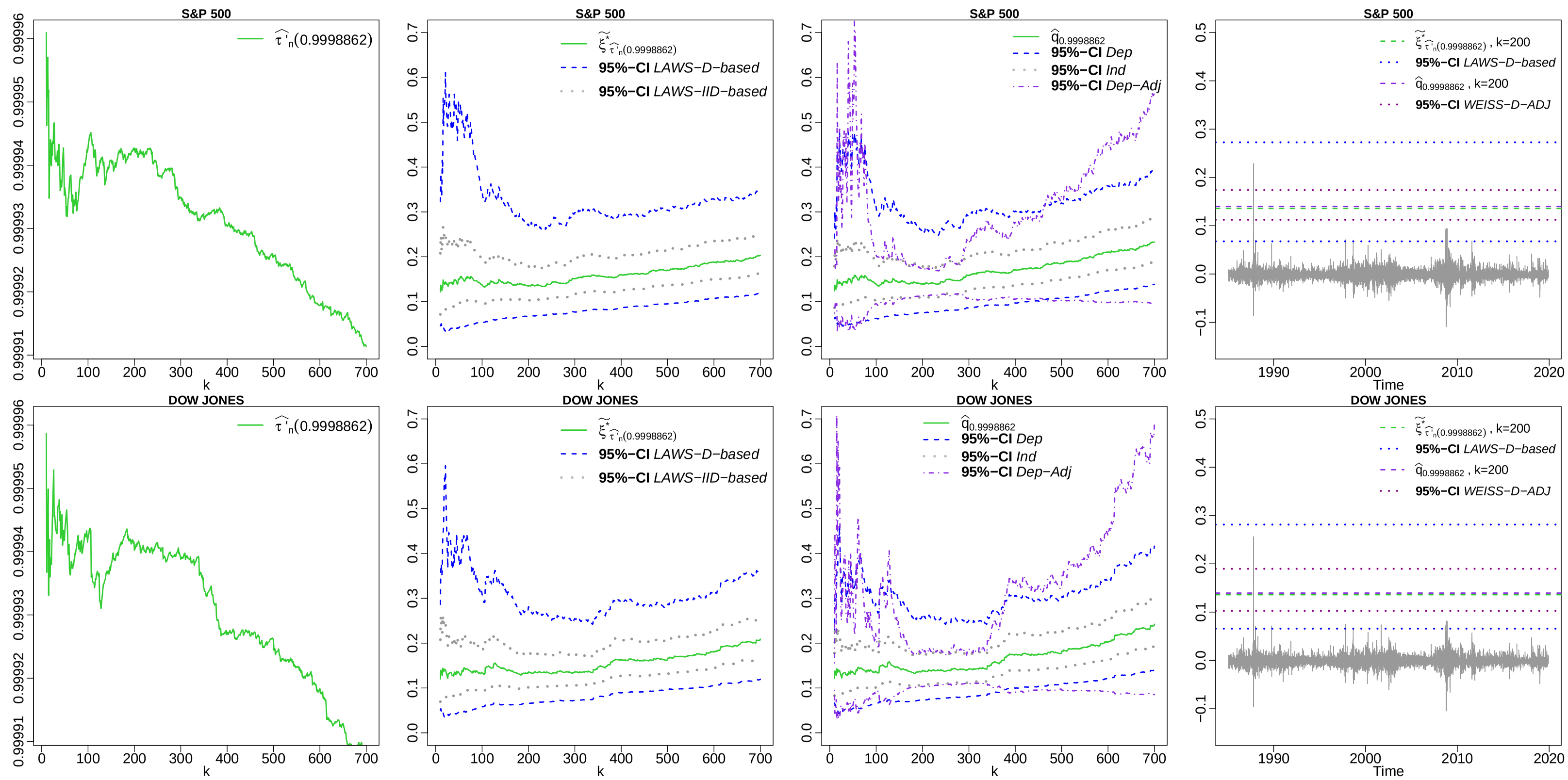

Figure 2: GSPC and DJIA negative daily log-returns data. From left to right, first panel: estimate $\widehat{\tau}'_n(0.9998862)$ (green line). Second panel: composite extrapolated LAWS estimate (green line) at level $\tau'_n = \widehat{\tau}'_n(0.9998862)$, with 95% confidence intervals based on i.i.d. theory (grey dotted line) and our LAWS-D procedure (blue dashed line). Third panel: Weissman quantile estimate (green line) at level $\alpha_n = 0.9998862$ (equivalently, composite extrapolated QB estimate at level $\tau'_n = \widehat{\tau}'_n(0.9998862)$), with 95% confidence intervals based on i.i.d. theory (grey dotted line), our LAWS-D procedure (blue dashed line) and the method of Drees (2003) (purple dashed-dotted line). Fourth panel: Negative daily log-returns data, with the composite extrapolated LAWS estimate (green dashed line) at level $\tau'_n = \widehat{\tau}'_n(0.9998862)$, compared to the standard Weissman quantile estimate $\widehat{q}^{\star}_{0.9998862}$ (purple dashed line), both with 95% confidence intervals (blue dotted line and purple dotted line respectively). Top row: GSPC data, bottom row: DJIA data.

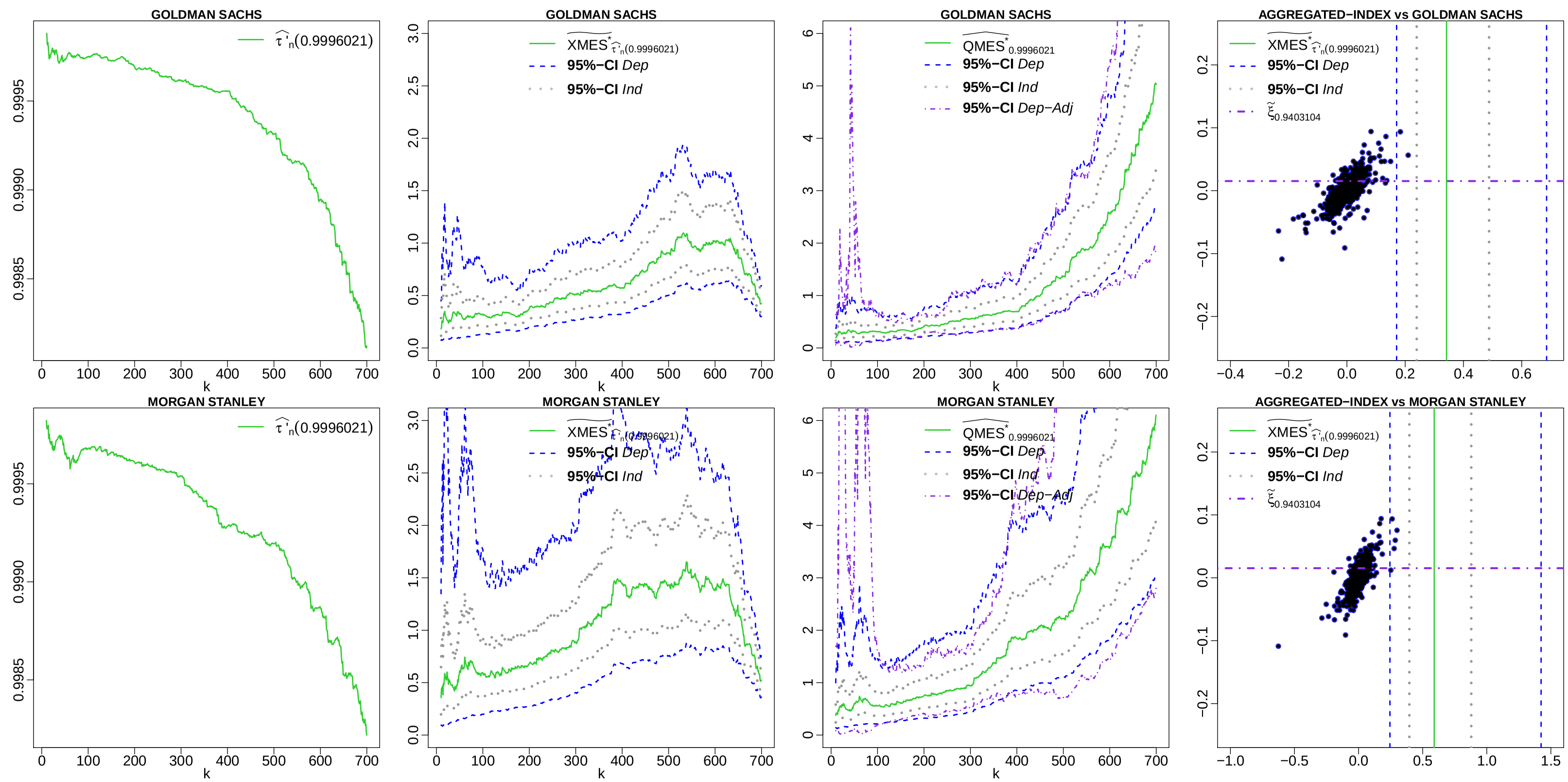

Figure 3: Goldman Sachs and Morgan Stanley loss returns data. From left to right, first panel: estimate $\widehat{\tau}_n'(0.9996021)$ (green line). Second panel: composite extrapolated LAWS-based estimate of the XMES (green line) at level $\tau_n' = \widehat{\tau}_n'(0.9996021)$, with 95% confidence intervals based on i.i.d. theory (grey dotted line) and our LAWS-D procedure (blue dashed line). Third panel: Weissman estimate of the QMES (green line) at level $\alpha_n = 0.9996021$ (equivalently, composite extrapolated QB estimate of the XMES at level $\tau_n' = \widehat{\tau}_n'(0.9996021)$), with 95% confidence intervals based on i.i.d. theory (grey dotted line), our LAWS-D procedure (blue dashed line) and the method of Drees (2003) (purple dashed-dotted line). Fourth panel: Scatterplot of the negative daily log-returns data, with the composite extrapolated LAWS estimate of XMES (green line) at level $\tau_n' = \widehat{\tau}_n'(0.9996021)$ with 95% confidence intervals (blue dashed line for the LAWS-D-based interval and grey dotted line for the LAWS-IID-based interval). The purple horizontal line represents the intermediate expectile level used in the anchor estimate $\widetilde{\text{XMES}}_{X,\tau_n}$. Top row: Goldman Sachs data, bottom row: Morgan Stanley data.