Tail Risk Inference via Expectiles in Heavy-Tailed Time Series

Anthony Davison, Simone A. Padoan, Gilles Stupfler

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

May 10, 2022

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 establish that extreme expectiles can be applied with the kind of dependent time series models relevant to finance. In this paper we provide a basis for inference on extreme expectiles and expectile-based marginal expected shortfall in a general β-mixing context that encompasses ARMA and GARCH models with heavy-tailed innovations. Our methods allow the estimation of marginal (pertaining to the stationary distribution) and dynamic (conditional on the past) extreme expectile-based risk measures. 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

Quantifying the risk associated to a profit-loss variable $Y \in \mathbb{R}$ is a key problem in financial econometrics. The best-known quantile-based risk measures have serious deficiencies: the Value-at-Risk is not coherent (Artzner et al. 1999) and only depends on the frequency of tail losses and not on their actual values, while the expected shortfall is not elicitable (Gneiting 2011) and thus does not benefit from the existence of a natural backtesting methodology. The expectile (Newey & Powell 1987), which extends the mean just as a quantile extends the median, addresses these issues. It is defined as

$$\xi_\tau = \arg\min_{\theta \in \mathbb{R}} \mathbb{E}\{\eta_\tau(Y-\theta) - \eta_\tau(Y)\}, \quad \text{with } \eta_\tau(u) = |\tau - \mathbb{1}\{u \leq 0\}|\, u^2, \tag{1}$$

in which $\mathbb{E}$ denotes expectation, $\mathbb{1}\{\cdot\}$ is the indicator function, and $\tau \in (0,1)$ is the level of the expectile. This risk measure is related to the popular gain-loss ratio performance measure (see Bellini & Di Bernardino 2017, and references therein), and it depends on both the realisations of Y and their probabilities. Moreover, expectiles with $\tau \geq 1/2$ induce the only risk measure that is law-invariant, coherent and elicitable. These results and others indicate that expectiles define sensible alternatives to the Value-at-Risk and expected shortfall, and inference for them has recently burgeoned.

Providing an accurate picture of financial and actuarial risk generally requires the estimation of risk measures at extreme levels, representing a very high level of prudentiality. For example, compliance with the European Union Solvency II directive essentially requires the estimation of the 99.5% quantile of financial losses over the next 12 months. A prerequisite for the wide adoption of expectiles as an instrument of risk protection is therefore a toolbox for their estimation at extreme levels, i.e., for $\tau \approx 1$, in models that accurately represent actuarial and financial data. The techniques available 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 away from 0 and 1. The only method for the estimation of marginal extreme expectiles with dependent data of which we are aware, due to Daouia et al. (2019), can only handle ϕ-mixing data, and as such does not allow the consideration of typical time series in finance, such as ARMA and GARCH models. Any inference based on such results may thus provide erroneous and overly-optimistic

results, unless one uses lower-frequency data in order to reduce dependence (as in Daouia et al. 2018, 2020), although this makes interpreting the results more difficult.

Here we develop new tools for inference for extreme expectile-based risk measures based on β-mixing heavy-tailed strictly stationary time series, which encompass many widely-used models. We first consider the prediction of marginal (pertaining to the stationary distribution of the series) tail expectiles, for which $\tau_n = 1 - p_n \to 1$ at any possible rate as the sample size $n \to \infty$; typically np_n is bounded, say $p_n \leq 1/n$, and then the expectile is expected to fall in a neighbourhood of or above the largest observations available. We then expand our statistical model to bivariate β-mixing time series $\{(X_t, Y_t), t = 1, 2, \ldots\}$ and establish estimation results for the marginal expected shortfall, which is important in assessing systemic risk (Acharya et al. 2017, Brownlees & Engle 2017). Our results apply to both the expectile- and quantile-based forms of the extreme marginal expected shortfall. We discuss the construction of confidence intervals for extreme expectile-based risk measures based on our theory and we show that they properly capture the amount of uncertainty in typical financial applications, unlike existing methods. Our last contribution is to discuss conditional or dynamic extreme expectile estimation in heteroskedastic regression models whose errors form a strictly stationary but dependent sequence. Our setup encompasses misspecified regression models, as well as popular financial time series models and in particular the celebrated GARCH-EVT framework of McNeil & Frey (2000). The asymptotic theory of marginal extreme expectile inference is shown to extend to this context when extreme conditional or dynamic expectiles are estimated based on the residuals of a regression model that are sufficiently close to the unobserved errors.

Our methods form part of the `R` package `ExtremeRisks` available on CRAN. The online Supplementary Material contains technical results, proofs and further numerical work. Section 2 of the paper explains our statistical context. Section 3 considers extreme expectile estimation. Section 4 introduces a 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. Section 6 focuses on conditional extreme expectile estimation in heteroskedastic time series models. The methods are applied to simulated data in Section 7

and to real financial data in Section 8. Section 9 draws conclusions.

2 Statistical model and time series framework

Let $(Y_t, t = 1, 2, \ldots)$ be a strictly stationary time series whose continuous marginal distribution F has survival function $\overline{F} = 1 - F$ and tail quantile function $U : s \mapsto \inf\{y \in \mathbb{R} : 1/\overline{F}(y) \geq s\}$. Throughout the paper Y should be seen as the negative of a generic financial position, so large positive values of Y represent extreme losses. We focus on heavy-tailed distributions and therefore assume that, for any $y > 0$, $\overline{F}(sy)/\overline{F}(s) \to y^{-1/\gamma}$ as $s \to \infty$, or equivalently $U(sy)/U(s) \to y^{\gamma}$. The tail index $\gamma > 0$ specifies the tail weight of $\overline{F}$. Together with the condition $\mathbb{E}|Y_-| < \infty$, where $Y_- = \min(Y, 0)$, the assumption $\gamma < 1$ ensures that $\mathbb{E}|Y| < \infty$, and then the expectile ξ_τ is well-defined by (1) for any τ.

We assume that the time series is β-mixing. 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 (Y_t). Then (Y_t) is β-mixing if the coefficients $\beta(l) = \sup_{m \geq 1} \beta_m(l)$ satisfy $\beta(l) \to 0$ as $l \to \infty$, where

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

Doukhan (1994, Section 2.4) shows that many Markov processes are geometrically β-mixing: under mild conditions there exists $a \in (0, 1)$ such that $\beta(l) \leq a^l$ for large enough l. This motivates the following basic assumption.

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

Our first goal is to estimate an unconditional marginal extreme expectile $\xi_{\tau'_n}$ of $Y = Y_1$, where $\tau'_n \to 1$ as $n \to \infty$, given data $Y_1, \ldots, Y_n$ from a time series satisfying Condition A.

3 Expectile estimation in time series

3.1 At intermediate levels

Let τ_n represent a sequence of probabilities such that $\tau_n \to 1$ and $n(1 - \tau_n) \to \infty$ as $n \to \infty$. A natural estimator of the expectile ξ_{τ_n} of the marginal distribution F is the em-

pirical counterpart of (1), i.e., the least asymmetrically weighted squares (LAWS) estimator $\widetilde{\xi}_{\tau_n} = \arg\min_{\theta\in\mathbb{R}} \sum_{t=1}^{n} \eta_{\tau_n}(Y_t - \theta)$. This is readily computed using iterative reweighted least squares. Another estimator exploits an asymptotic proportionality relationship between the high expectile ξ_τ and the corresponding quantile q_τ: following Bellini et al. (2014),

$$\lim_{\tau\to 1} \frac{\xi_\tau}{q_\tau} = (\gamma^{-1} - 1)^{-\gamma}, \text{ when } \gamma < 1. \tag{2}$$

Thus a quantile-based estimator of ξ_{τ_n} is $\widehat{\xi}_{\tau_n} = (\widehat{\gamma}_n^{-1} - 1)^{-\widehat{\gamma}_n} \widehat{q}_{\tau_n}$, where $\widehat{q}_{\tau_n} = Y_{n-\lfloor n(1-\tau_n)\rfloor,n}$ is the empirical counterpart of q_{τ_n}, $Y_{1,n} \le \cdots \le Y_{n,n}$ denote ascending order statistics, and $\widehat{\gamma}_n$ is a consistent estimator of γ. Here $\lfloor x \rfloor$ is the largest integer smaller than or equal to x. We focus on two simple estimators. The first, due to Hill (1975),

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

is the maximum likelihood estimator in a Pareto model. The second arises on 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, yielding the expectile-based (EB) estimator $\widehat{\gamma}_n^E = (1 + \widehat{\overline{F}}_n(\widetilde{\xi}_{\tau_n})/(1-\tau_n))^{-1}$.

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 $\overline{F}$ and a purely Pareto tail.

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

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

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

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

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

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

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

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

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

We start with a general result on the LAWS estimator $\widetilde{\xi}_{\tau_n}$.

Theorem 3.1. *Assume that Conditions A and B are satisfied and that there exists some* $\delta > 0$ *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*

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

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

The proof uses rather delicate arguments involving a tailored dependent central limit theory for tail array sums. The family of functions $R = (R_t, t = 1, 2, \ldots)$ specifies the extremal dependence within the time series between different time points; when $R_t \equiv 0$ for all $t \geq 1$ (e.g., when (Y_t) is i.i.d.), the asymptotic variance is $2\gamma^3/(1-2\gamma)$. The quantity $\sigma^2(\gamma, R)$ 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. 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. As mentioned in Section 2, geometrically strong β-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). More generally, assumptions such as Conditions A and B are typically satisfied by classical linear time series and heteroskedastic time series; see Section 4 in Drees (2000), Section 3 in Drees (2003) and Section 5 in de Haan et al. (2016), which also provide expressions for the R_t. The following corollary provides a rigorous statement in the geometrically mixing case.

Corollary 3.2. *Suppose that Conditions A and B(ii)-(iii) are satisfied, and that* $\beta(l) = \mathrm{O}(a^l)$ *for some* $a \in (0,1)$. *Suppose also 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, with the notation of Theorem 3.1,*

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

Theorem 3.1 and Corollary 3.2 represent a substantial step in the risk management literature, for they allow one to handle many important financial examples.

We now consider the estimators $\widehat{\xi}_{\tau_n}$ based on $\widehat{\gamma}_n^H$ and $\widehat{\gamma}_n^E$. In this case the crucial result, which is of interest in its own right, provides a joint Gaussian approximation of the tail empirical quantile process and its logarithm, $s \mapsto \widehat{q}_{1-(1-\tau_n)s} = Y_{n-\lfloor n(1-\tau_n)s\rfloor,n}$ and $s \mapsto \log \widehat{q}_{1-(1-\tau_n)s}$, in our mixing framework. See Theorem 2.4.8 in de Haan & Ferreira (2006) for a result restricted to the i.i.d. 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. *Suppose that Conditions A, B and C are satisfied, and 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 also that* $n(1-\tau_n)$ *is a sequence of integers and pick* $s_0 > 0$. *Then there exist 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 as* $n \to \infty$, *uniformly in* $s \in (0, s_0]$,

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

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

Condition B is used 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 $(u, v) = (u', v')$. A consequence of Theorem 3.3 is that 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*

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

where $m(\gamma) = (1-\gamma)^{-1} - \log(\gamma^{-1} - 1)$,

$$\lambda = \left(\frac{(\gamma^{-1}-1)^{-\rho}}{1-\gamma-\rho} + \frac{(\gamma^{-1}-1)^{-\rho} - 1}{\rho}\right)\lambda_1 + \gamma(\gamma^{-1}-1)^{\gamma}\mathbb{E}(Y)\lambda_2,$$

and

$$v^H(\gamma, R) = (1 + [m(\gamma)]^2)\left(1 + 2\sum_{t=1}^{\infty} R_t(1,1)\right) + 2m(\gamma)\int_0^1 \sum_{t=1}^{\infty}\left[\frac{R_t(s,1) + R_t(1,s)}{s} - 2R_t(1,1)\right] \mathrm{d}s.$$

To the best of our knowledge, Corollary 3.4 is the first result on the estimator $\widehat{\xi}_{\tau_n}$ at intermediate levels under weak dependence assumptions (i.e., in the β-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; owing to asymptotic variance considerations, $\widehat{\gamma}_n^E$ tends 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 that require an estimate of γ, as we illustrate in Sections B.1 and B.2 of the Supplementary Material when constructing such intervals for intermediate expectiles.

3.2 At extreme levels

We now consider estimating 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 $1/n$ (e.g., Cai et al. 2015). Our semiparametric approach is motivated by combining the heavy-tailed assumption and condition (2) and defines an extreme expectile estimator through a Weissman (1978)-type construction, whereby

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

for n large. This suggests 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 expectile ξ_{τ_n}. We call $\overline{\xi}^{\star}_{\tau_n'}$ the extrapolating LAWS estimator when $\overline{\xi}_{\tau_n} = \widetilde{\xi}_{\tau_n}$, then denoting it by $\widetilde{\xi}^{\star}_{\tau_n'}$, and call it the extrapolating quantile-based (QB) estimator when $\overline{\xi}_{\tau_n} = \widehat{\xi}_{\tau_n}$, then denoting it by $\widehat{\xi}^{\star}_{\tau_n'}$. 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 $n(1-\tau_n) \to \infty$, $n(1-\tau_n') \to c \in [0,\infty)$ 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 $n \to \infty$. Then, for $\widehat{\gamma}_n = \widehat{\gamma}_n^H$ and as $n \to \infty$, the weak convergence*

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

holds when

- $\overline{\xi}^{\star}_{\tau_n} = \widetilde{\xi}^{\star}_{\tau_n'}$, *provided there exists some $\delta > 0$ such that $\mathbb{E}|Y_-|^{2+\delta} < \infty$, $0 < \gamma < 1/(2+\delta)$, $\sum_{l\geq 1}[\beta(l)]^{\delta/(2+\delta)} < \infty$ and $r_n\{r_n/\sqrt{n(1-\tau_n)}\}^{\delta} \to 0$ as $n \to \infty$,*
- $\overline{\xi}^{\star}_{\tau_n} = \widehat{\xi}^{\star}_{\tau_n'}$, *without further assumptions.*

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 all $t \geq 1$, the asymptotic variance is γ^2. An analogous result holds for the estimator $\widehat{\gamma}_n^E$.

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

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

where $q_{Y,\tau}$ is the τ quantile of Y. A systemic crisis typically corresponds to a situation in which τ 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,

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

with $\xi_{Y,\tau}$ the τ expectile of Y. Estimation of $\mathrm{QMES}_{X,\tau}$ and $\mathrm{XMES}_{X,\tau}$ at extreme levels is considered by Cai et al. (2015) and Daouia et al. (2018), but only for i.i.d. 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 $\{(X_t, Y_t), t = 1, 2, \ldots\}$; 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 $\mathcal{F}_{1,m} = \sigma(X_1, Y_1, \ldots, X_m, Y_m)$ 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 (X_t, Y_t). Then the corresponding β-mixing coefficients can be defined as $b(l) = \sup_{m \geq 1} b_m(l)$, where

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

The sequence $\{(X_t, Y_t), t = 1, 2, \ldots\}$ is then said to be β-mixing if $b(l) \to 0$ as $l \to \infty$. If it is β-mixing, then $(Y_t, t = 1, 2, \ldots)$ is also β-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 $\{(X_t, Y_t), t = 1, 2, \ldots\}$ is strictly stationary, β-mixing and the one-dimensional marginal distribution functions F_X and F_Y of $(X_t, t = 1, 2, \ldots)$ and $(Y_t, t = 1, 2, \ldots)$ are continuous and heavy-tailed with respective tail indices γ_X and γ_Y;

(ii) there exist sequences of integers (l_n) and (r_n) such that $l_n \to \infty$, $r_n \to \infty$, $l_n/r_n \to 0$, $r_n/n \to 0$ and $n\, b(l_n)/r_n \to 0$ as $n \to \infty$;

(iii) for any $t = 1, 2, \ldots$ there exists a function r_t on $\mathcal{E}_4 = [0, \infty]^4 \setminus \{(\infty, \infty, \infty, \infty)\}$ such that

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

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

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

If $\{(X_t, Y_t), t = 1, 2, \ldots\}$ satisfies Condition D, then the univariate time series $(Y_t, t = 1, 2, \ldots)$ 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 $\mathrm{MES}_{X,\tau} = \mathbb{E}(X \mid Y > z_{Y,\tau})$ for $\tau \in (0, 1)$, where $z_{Y,\tau}$ is a risk measure on Y such that $\overline{F}(z_{Y,\tau})/(1 - \tau) \to z = z(\gamma_Y) \in (0, \infty)$ as $\tau \uparrow 1$. If 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 an extrapolation relationship similar to (3) suggests estimating $\mathrm{MES}_{X,\tau_n'} = \mathbb{E}(X \mid Y > z_{Y,\tau_n'})$ by

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

Replacing $\overline{z}_{Y,\tau_n}$ by $\widehat{q}_{Y,\tau_n} \equiv \widehat{q}_{\tau_n}$, $\widetilde{\xi}_{Y,\tau_n} \equiv \widetilde{\xi}_{\tau_n}$ or $\widehat{\xi}_{Y,\tau_n} = (\widehat{\gamma}_{Y,n}^{-1} - 1)^{-\widehat{\gamma}_{Y,n}} \widehat{q}_{Y,\tau_n} \equiv \widehat{\xi}_{\tau_n}$ yields estimators $\widehat{\mathrm{QMES}}^{\star}_{X,\tau_n'}$, $\widetilde{\mathrm{XMES}}^{\star}_{X,\tau_n'}$ or $\widehat{\mathrm{XMES}}^{\star}_{X,\tau_n'}$ of (4) or (5). The following second-order condition is needed to quantify the bias incurred in constructing these estimators.

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

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

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

The following theorem gives 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 $\rho_X < 0$, that there exists $\delta > 0$ such that $0 < \gamma_X < 1/(2+\delta)$, and that*

(i) $\tau_n,\ \tau_n' \uparrow 1$, with $n(1-\tau_n) \to \infty$, $n(1-\tau_n') \to c < \infty$ 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 exists $\varepsilon > 0$ 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) $\mathbb{E}(|X_-|^{1/\gamma_X}) < \infty$ 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, as $n \to \infty$:

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

(vi) the weak convergence $\sqrt{n(1-\tau_n)}(\widehat{\gamma}_{X,n} - \gamma_X) \xrightarrow{d} \Gamma$ holds, where Γ is a nondegenerate random variable, and

$$\sqrt{n(1-\tau_n)}\left(\frac{\overline{z}_{Y,\tau_n}}{z_{Y,\tau_n}} - 1\right) = \mathrm{O}_{\mathbb{P}}(1).$$

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

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

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 rates. In this condition, Γ is typically a normal distribution, as is for instance the case when $\widehat{\gamma}_{X,n}$ is a Hill estimator. Theorem 4.1 has the following important corollary.

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

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

holds for

- $\overline{\mathrm{MES}}^{\star}_{X,\tau'_n} = \widehat{\mathrm{QMES}}^{\star}_{X,\tau'_n}$, *without further assumptions,*
- $\overline{\mathrm{MES}}^{\star}_{X,\tau'_n} = \widetilde{\mathrm{XMES}}^{\star}_{X,\tau'_n}$, *assuming 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* $\sqrt{n(1-\tau_n)}q^{-1}_{Y,\tau_n} \to 0$ *as* $n \to \infty$,
- $\overline{\mathrm{MES}}^{\star}_{X,\tau'_n} = \widehat{\mathrm{XMES}}^{\star}_{X,\tau'_n}$, *assuming with the notation of Theorem 4.1 that* $\mathbb{E}|Y_-| < \infty$, $0 < \gamma_Y < 1$, *and* $\sqrt{n(1-\tau_n)}q^{-1}_{Y,\tau_n} \to 0$ *and* $\sqrt{n(1-\tau_n)}(\widehat{\gamma}_{Y,n} - \gamma_Y) = \mathrm{O}_{\mathbb{P}}(1)$ *as* $n \to \infty$.

5 Finite-sample inference and expectile level selection

The theory 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 $w(\gamma, R) = \gamma^2\{1 + 2\sum_{t=1}^{\infty} R_t(1,1)\}$ in Theorem 3.5. As F is continuous, $\mathbb{1}\{F(Y_1) > \tau_n\} = \mathbb{1}\{F(Y_1) \geq \tau_n\}$ with probability 1, and we can apply Proposition 2.1 in Drees (2003) to obtain that, when $n(1-\tau_n) \to \infty$, $r_n \to \infty$ and $r_n(1-\tau_n) \to 0$,

$$\lim_{n\to\infty} \frac{1}{r_n(1-\tau_n)} \mathrm{Var}\left[\sum_{i=1}^{r_n} \mathbb{1}\{F(Y_i) \geq \tau_n\}\right] = 1 + 2\sum_{t=1}^{\infty} R_t(1,1).$$

We split the data into big blocks of length r_n separated by small blocks of length l_n and define $\widehat{Z}_j = \sum_{t=1+j\ell_n}^{r_n+j\ell_n} \mathbb{1}\{\widehat{F}_n(Y_t) \geq \tau_n\}$ for $j = 0, 1, \ldots, m_n - 1$, where $m_n = \lfloor n/\ell_n \rfloor$, $\ell_n = r_n + l_n$, and $\widehat{F}_n$ is the empirical distribution function of all the data. We then compute the sample variance Σ_n of $\widehat{Z}_0, \ldots, \widehat{Z}_{m_n-1}$ and obtain an estimator

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

of $w(\gamma, R)$ (for consistency of $\widehat{w}_n(\gamma, R)$, see Section D of the Supplementary Material). As $n \to \infty$, Theorem 3.5 yields the weak convergence

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

Thus a $(1-\alpha)$-equi-tailed confidence interval for $\xi_{\tau_n'}$, called LAWS-D-ADJ in the sequel, has limits

$$\overline{\xi}_{\tau_n'}^{\star} \left(\frac{1-\tau_n}{1-\tau_n'}\right)^{-\widehat{b}_n \pm z_{1-\alpha/2}\sqrt{\widehat{w}_n(\gamma,R)/\{n(1-\tau_n)\}}}, \tag{6}$$

where $\overline{\xi}_{\tau_n'}^{\star}$ is the LAWS or quantile-based extrapolating estimator, the bias $\lambda_1/(1-\rho)$ is estimated by $\widehat{b}_n = \widehat{\gamma}_n^H \widehat{\beta}_n (1-\tau_n)^{-\widehat{\rho}_n}/(1-\widehat{\rho}_n)$, with $\widehat{\beta}_n$ and $\widehat{\rho}_n$ computed using the `R` package `evt0` (Manjunath & Caeiro 2013), and z_p is the p-quantile of the standard normal distribution. A simpler interval, which we call LAWS-D, sets $\widehat{b}_n = 0$ and thus ignores the bias. The asymptotic variance in the i.i.d. case, γ^2, is estimated by $(\widehat{\gamma}_n^H)^2$, resulting in so-called LAWS-IID intervals. The analogous intervals using the quantile-based estimator in place of the LAWS estimator are called QB-D-ADJ, QB-D and QB-IID below. Similar confidence intervals can be constructed for (5) by replacing Y_t by X_t in the $\widehat{Z}_j$ and the Hill estimators, and by replacing $\overline{\xi}_{\tau_n'}^{\star}$ either by $\widetilde{\mathrm{XMES}}_{X,\tau_n'}^{\star}$ or by $\widehat{\mathrm{XMES}}_{X,\tau_n'}^{\star}$.

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

$$\widehat{\tau}_n'(\alpha_n) = 1 - (1-\alpha_n)\frac{\widehat{\gamma}_n^H}{1-\widehat{\gamma}_n^H}. \tag{7}$$

Using this estimator in place of τ_n' in the extrapolating LAWS and quantile-based estimators yields composite estimators of $\xi_{\tau_n'(\alpha_n)} \equiv q_{\alpha_n}$. If one uses the same τ_n in the extrapolation

step and in the calculation of $\widehat{\tau}'_n(\alpha_n)$, the composite extrapolating LAWS estimator is

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

put another way, $\widetilde{\xi}^{\star}_{\widehat{\tau}'_n(\alpha_n)}$ can be constructed by inserting the estimator $\widehat{\gamma}^H_n$ and the extrapolating LAWS estimator at level α_n into the right-hand side of the approximation $q_{\alpha_n} \equiv \xi_{\tau'_n(\alpha_n)} \approx (\gamma^{-1}-1)^{\gamma}\xi_{\alpha_n}$ obtained from (2). So far as we know, this has not previously been noticed. Combined with our marginal expected shortfall estimators, this gives two estimators of $\mathrm{QMES}_{X,\alpha_n}$ whose asymptotic properties we now establish.

Theorem 5.1. *Suppose the conditions of Theorem 3.5 hold with α_n in place of τ'_n. Then, if $\widehat{\gamma}_n = \widehat{\gamma}^H_n$ and $\overline{\xi}^{\star}$ is either $\widehat{\xi}^{\star}$ or $\widetilde{\xi}^{\star}$, we have, as $n \to \infty$,*

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

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

Theorem 5.2. *Suppose the conditions of Corollary 4.2 hold with α_n in place of τ'_n. Then, if $\widehat{\gamma}_{X,n} = \widehat{\gamma}^H_{X,n}$ and $\overline{\mathrm{XMES}}^{\star} = \widetilde{\mathrm{XMES}}^{\star}$ (resp. $\overline{\mathrm{XMES}}^{\star} = \widehat{\mathrm{XMES}}^{\star}$), then, as $n \to \infty$,*

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

6 Conditional or dynamic extreme expectile estimation in heteroskedastic time series

More informative and/or reactive risk measure estimates for many applications are obtained when auxiliary information provided by covariates and/or the past of the time series is incorporated, but it is then necessary to estimate conditional or dynamic extreme expectiles. We tackle this within heteroskedastic regression models of the form $Y_t = g(\boldsymbol{X}_t) + \sigma(\boldsymbol{X}_t)\varepsilon_t$, where g and $\sigma > 0$ are measurable and unknown and the series $(\boldsymbol{X}_t)$, possibly partially or even totally unobserved, is independent of the strictly stationary series of heavy-tailed

innovations (ε_t). This covers classical regression models, possibly misspecified due to temporal dependence within the series (ε_t), as well as standard time series such as ARMA and GARCH models. In this context, the location equivariance and positive homogeneity of expectiles implies that

$$\xi_\tau(Y_t \mid \boldsymbol{X}_t) = g(\boldsymbol{X}_t) + \sigma(\boldsymbol{X}_t)\xi_\tau(\varepsilon), \text{ with } \varepsilon \stackrel{d}{=} \varepsilon_1.$$

If the errors ε_t were available, we could estimate $\xi_{\tau_n}(\varepsilon)$ directly using the LAWS estimator. We replace the unobserved ε_t by residuals $\widehat{\varepsilon}_t^{(n)}$, resulting in the residual-based LAWS estimator $\widetilde{\xi}_{\tau_n}(\varepsilon) = \arg\min_{\theta\in\mathbb{R}} \sum_{t=1}^n \eta_{\tau_n}(\widehat{\varepsilon}_t^{(n)} - \theta)$. This should be a $\sqrt{n(1-\tau_n)}$-(relatively) asymptotically normal estimator of $\xi_{\tau_n}(\varepsilon)$ for intermediate τ_n when the residuals and unobservable errors are similar, and we ensure this through the following condition.

Condition F. As $n \to \infty$, the array of random variables $(\widehat{\varepsilon}_t^{(n)}, 1 \le t \le n)$, satisfies

$$\sqrt{n(1-\tau_n)} \max_{1\le t\le n} \frac{|\widehat{\varepsilon}_t^{(n)} - \varepsilon_t|}{1+|\varepsilon_t|} \xrightarrow{\mathbb{P}} 0.$$

Under this condition the following result shows that the estimator $\widetilde{\xi}_{\tau_n}(\varepsilon)$ has the same asymptotic properties as its unobtainable counterpart resulting from asymmetric least squares estimation based on the innovations ε_t.

Theorem 6.1. *Suppose that the sequence of innovations $(\varepsilon_t, t = 1, 2, \ldots)$, rather than the time series $(Y_t, t = 1, 2, \ldots)$, satisfies Conditions A and B, and that Condition F holds. Suppose further that there exists $\delta > 0$ such that $\mathbb{E}|\varepsilon_-|^{2+\delta} < \infty$, $0 < \gamma < 1/(2+\delta)$ and $\sum_{l\ge1}[\beta(l)]^{\delta/(2+\delta)} < \infty$. 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,*

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

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

An alternative, quantile-based estimator of $\xi_{\tau_n}(\varepsilon)$ is $\widehat{\xi}_{\tau_n}(\varepsilon) = (\widehat{\gamma}_n^{-1}(\varepsilon) - 1)^{-\widehat{\gamma}_n(\varepsilon)}\widehat{q}_{\tau_n}(\varepsilon)$, where $\widehat{q}_{\tau_n}(\varepsilon) = \widehat{\varepsilon}^{(n)}_{n-\lfloor n(1-\tau_n)\rfloor,n}$ and $\widehat{\gamma}_n(\varepsilon)$ is any residual-based estimator of the tail index of ε. For the asymptotic analysis of such an estimator, an analogue of Theorem 3.3 is required;

again, under Condition F, the residual-based tail empirical process has the same behaviour as its counterpart based on the unobservable innovations. A rigorous result follows.

Theorem 6.2. *Suppose that the sequence of innovations* $(\varepsilon_t, t = 1, 2, \ldots)$*, rather than the time series* $(Y_t, t = 1, 2, \ldots)$*, satisfies Conditions A, B and C, and that Condition F holds. Suppose also 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 versions of the process* $s \mapsto \widehat{\varepsilon}_{n-\lfloor n(1-\tau_n)s \rfloor, n}$ *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* $\delta > 0$ *sufficiently small, we have as* $n \to \infty$*, uniformly in* $s \in (0, s_0]$*,*

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

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

We can then prove the following corollary on the asymptotic behaviour of the quantile-based estimator $\xi_{\tau_n}(\varepsilon)$ when

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

is the residual-based version of the Hill estimator.

Corollary 6.3. *Suppose that the conditions of Theorem 6.2 hold with* $\mathbb{E}|\varepsilon_-| < \infty$*,* $0 < \gamma < 1$ *and* $\mathbb{E}(\varepsilon) = 0$*, as well as* $\sqrt{n(1-\tau_n)}q_{\tau_n}^{-1}(\varepsilon) = \mathrm{O}(1)$ *and* $\sqrt{n(1-\tau_n)}A((1-\tau_n)^{-1}) \to \lambda \in \mathbb{R}$ *as* $n \to \infty$*. If* $\widehat{\gamma}_n(\varepsilon) = \widehat{\gamma}_n^H(\varepsilon)$ *then, as* $n \to \infty$*,*

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

$$\sqrt{n(1-\tau_n)}(\widehat{\gamma}_n(\varepsilon) - \gamma) \xrightarrow{d} \mathcal{N}\left(\frac{\lambda}{1-\rho}, \gamma^2\left\{1 + 2\sum_{t=1}^{\infty} R_t(1,1)\right\}\right),$$

with $b(\gamma,\rho) = (\gamma^{-1}-1)^{-\rho}/(1-\gamma-\rho) + \{(\gamma^{-1}-1)^{-\rho}-1\}/\rho$ *and*

$$v^H(\gamma,R) = (1+[m(\gamma)]^2)\left(1+2\sum_{t=1}^{\infty} R_t(1,1)\right) + 2m(\gamma)\int_0^1 \sum_{t=1}^{\infty}\left[\frac{R_t(s,1)+R_t(1,s)}{s} - 2R_t(1,1)\right]\mathrm{d}s.$$

As in the unconditional setting, we can now define classes of residual-based, plug-in extrapolating estimators of $\xi_{\tau_n'}(\varepsilon)$ at a level $\tau_n' \to 1$ such that $n(1-\tau_n') \to c < \infty$:

$$\overline{\xi}^{\star}_{\tau_n'}(\varepsilon) = \left(\frac{1-\tau_n'}{1-\tau_n}\right)^{-\widehat{\gamma}_n(\varepsilon)} \overline{\xi}_{\tau_n}(\varepsilon), \text{ where } \overline{\xi}_{\tau_n}(\varepsilon) = \widetilde{\xi}_{\tau_n}(\varepsilon) \text{ or } \widehat{\xi}_{\tau_n}(\varepsilon).$$

Given estimators $\overline{g}(\boldsymbol{x})$ and $\overline{\sigma}(\boldsymbol{x})$ of g and σ at a value $\boldsymbol{x}$ of the covariate, this readily yields estimators of extreme conditional or dynamic expectiles as $\overline{\xi}^{\star}_{\tau_n'}(Y_1 \mid \boldsymbol{X}_1 = \boldsymbol{x}) = \overline{g}(\boldsymbol{x}) + \overline{\sigma}(\boldsymbol{x})\overline{\xi}^{\star}_{\tau_n'}(\varepsilon)$. The following asymptotic normality result then holds.

Theorem 6.4. *Suppose that the conditions of Corollary 6.3 hold with* $\rho < 0$*, and that* $\tau_n' \uparrow 1$ *satisfies* $n(1-\tau_n') \to c \in [0,\infty)$ *and* $\sqrt{n(1-\tau_n)}/\log\{(1-\tau_n)/(1-\tau_n')\} \to \infty$*. If* $\widehat{\gamma}_n(\varepsilon) = \widehat{\gamma}_n^H(\varepsilon)$ *then the weak convergence*

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

holds as $n \to \infty$ *for*

- $\overline{\xi}_{\tau_n}(\varepsilon) = \widetilde{\xi}_{\tau_n}(\varepsilon)$*, assuming that there exists* $\delta > 0$ *such that* $\mathbb{E}|\varepsilon_-|^{2+\delta} < \infty$*,* $0 < \gamma < 1/(2+\delta)$*,* $\sum_{l\geq 1}[\beta(l)]^{\delta/(2+\delta)} < \infty$ *and* $r_n(r_n/\sqrt{n(1-\tau_n)})^{\delta} \to 0$ *as* $n \to \infty$*,*
- $\overline{\xi}_{\tau_n}(\varepsilon) = \widehat{\xi}_{\tau_n}(\varepsilon)$*, without further assumptions.*

In the heteroskedastic regression model $Y_t = g(\boldsymbol{X}_t) + \sigma(\boldsymbol{X}_t)\varepsilon_t$ *(with* $(\boldsymbol{X}_t)$ *independent of* (ε_t)*), if, at a given point* $\boldsymbol{x}$*, the estimators* $\overline{g}(\boldsymbol{x})$ *and* $\overline{\sigma}(\boldsymbol{x})$ *satisfy* $\overline{g}(\boldsymbol{x}) - g(\boldsymbol{x}) = \mathrm{O}_{\mathbb{P}}(1)$ *and* $\sqrt{n(1-\tau_n)}(\overline{\sigma}(\boldsymbol{x}) - \sigma(\boldsymbol{x})) = \mathrm{O}_{\mathbb{P}}(1)$*, then the estimator* $\overline{\xi}^{\star}_{\tau_n'}(Y_1 \mid \boldsymbol{X}_1 = \boldsymbol{x}) = \overline{g}(\boldsymbol{x}) + \overline{\sigma}(\boldsymbol{x})\overline{\xi}^{\star}_{\tau_n'}(\varepsilon)$ *is such that*

$$\frac{\sqrt{n(1-\tau_n)}}{\log\{(1-\tau_n)/(1-\tau_n')\}}\left(\frac{\overline{\xi}^{\star}_{\tau_n'}(Y_1 \mid \boldsymbol{X}_1 = \boldsymbol{x})}{\xi_{\tau_n'}(Y_1 \mid \boldsymbol{X}_1 = \boldsymbol{x})} - 1\right) \xrightarrow{d} \mathcal{N}\left(\frac{\lambda}{1-\rho}, \gamma^2\left\{1+2\sum_{t=1}^{\infty} R_t(1,1)\right\}\right)$$

as $n \to \infty$*, for the same choices of* $\overline{\xi}_{\tau_n}(\varepsilon)$ *under the same assumptions.*

We may finally provide a residual-based version of the composite extrapolating LAWS estimator, targeted at the (conditional) extreme quantile of fixed level α_n, that is,

$$\overline{\xi}^{\star}_{\widehat{\tau}'_n(\alpha_n)}(Y_1 \mid \boldsymbol{X}_1 = \boldsymbol{x}) = \overline{g}(\boldsymbol{x}) + \overline{\sigma}(\boldsymbol{x})\overline{\xi}^{\star}_{\tau'_n}(\varepsilon) \text{ with } \widehat{\tau}'_n(\alpha_n) = 1 - (1-\alpha_n)\,\frac{\widehat{\gamma}^H_n(\varepsilon)}{1-\widehat{\gamma}^H_n(\varepsilon)}.$$

Theorem 6.5. *Suppose that the conditions of Theorem 6.4 hold with α_n in place of τ'_n. Then, if $\widehat{\gamma}_n(\varepsilon) = \widehat{\gamma}^H_n(\varepsilon)$ and $\overline{\xi}^{\star}(\varepsilon)$ is either $\widehat{\xi}^{\star}(\varepsilon)$ or $\widetilde{\xi}^{\star}(\varepsilon)$, we have, as $n \to \infty$,*

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

Our marginal extreme expectile estimation and inference techniques therefore extend to the estimation of conditional and/or dynamic extreme expectiles of the form $\overline{\xi}^{\star}_{\tau'_n}(Y_t \mid \boldsymbol{X}_t = \boldsymbol{x})$ in heteroskedastic location-scale models, provided $\sqrt{n}$-consistent estimators of the location and scale functions g and σ are available. This is generally the case in linear heteroskedastic models, single-index models and additive models with dependent errors, and ARMA/GARCH models. In particular, our theoretical results encompass the GARCH-EVT framework of McNeil & Frey (2000), which has become very popular due to its capacity to estimate dynamic extreme risk measures without strong parametric specification of the innovation distribution. A related investigation, limited to i.i.d. ε_t and therefore unable to handle misspecified regression models, is performed by Girard et al. (2021*a*).

7 Simulation experiments

7.1 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 of a similar study of 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 $Y_{t+1} = 0.8\,Y_t + \varepsilon_{t+1}$, where the innovations ε_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 ε_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 $\sigma^2_{t+1} = 0.4 + 0.6\,Y_t^2$, and the ε_t are independent standard Gaussian innovations; and

(d) the GARCH(1,1) model $Y_{t+1} = \sigma_{t+1}\varepsilon_{t+1}$, where $\sigma^2_{t+1} = 0.1 + 0.4\,Y_t^2 + 0.4\,\sigma_t^2$, and the ε_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. Models (a) and (b) are standard linear time series, so the marginal distribution, like the innovations, has tail index $\gamma_Y = 1/3$. The marginal distribution in (c) and (d) is heavy-tailed, and the tail indices, calculated using Theorem 2.1 in Mikosch & Stărică (2000), equal $\gamma_Y = 0.262$ and $\gamma_Y = 0.239$ respectively.

We simulate 10^4 samples of size $n = 2500$ from 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 LAWS and quantile-based estimators using the intermediate level $\tau_n = 1 - k/n$ for $k \in \{6, 8, \ldots, 700\}$ for each simulated dataset and obtain 95% confidence intervals based on (6). 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 so that l_n is greater than or equal to a lag after which sample autocorrelations of the data from both (Y_t) and (Y_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 i.i.d., 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. Although the procedure depends somewhat on the tuning parameters, our confidence intervals perform very well overall for such sample sizes when using a rule of thumb to select a k lying in $[50, 200]$. Similar results at level $\tau_n' = 0.9999 > 1 - 1/n$ are shown in Figure B.2 of the Supplementary Material.

7.2 Extreme marginal expected shortfall estimation

To assess the finite-sample performance of the extreme marginal expected shortfall estimators we simulate data from four models in which the bivariate innovations are i.i.d.:

(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 $\sigma^2_{X,t+1} = 0.4 + 0.6\,X_t^2$ and $\sigma^2_{Y,t+1} = 0.4 + 0.6\,Y_t^2$. 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 (e); and

(h) $X_{t+1} = \sigma_{X,t+1}\varepsilon_{X,t+1}$, with $\sigma^2_{X,t+1} = 0.1 + 0.4\,X_t^2 + 0.4\,\sigma^2_{X,t}$, and $Y_{t+1} = \sigma_{Y,t+1}\varepsilon_{Y,t+1}$, with $\sigma^2_{Y,t+1} = 0.1 + 0.4\,Y_t^2 + 0.4\,\sigma^2_{Y,t}$. 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 7.1.

We simulate 10^4 samples of size $n = 2500$ for each model, and consider estimating $\mathrm{QMES}_{X,\alpha_n}$ at levels α_n such that $\tau'_n(\alpha_n) = 0.9995$ and 0.9999, by comparing our composite estimator $\widetilde{\mathrm{XMES}}^{\star}_{X,\widehat{\tau}'_n(\alpha_n)}$ with $\widehat{\mathrm{QMES}}^{\star}_{X,\alpha_n}$; recall that $(1-\alpha_n) \approx (\gamma_Y^{-1} - 1)\{1 - \tau'_n(\alpha_n)\}$, which allows to determine accurate approximations of α_n using the tail index values γ_Y calculated in Section 7.1, and the construction of $\widehat{\tau}'_n(\alpha_n)$ in Section 5. The true values of $\mathrm{QMES}_{X,\alpha_n} \approx \mathrm{XMES}_{X,\tau'_n(\alpha_n)}$, found by simulation, are in the bottom part of Table B.5 in the Supplementary Material. Estimation at this level of α_n gives us an idea of the performance of our composite estimation method in a problem comparable in difficulty

to that of Section 7.1, at least so far as the Y component is concerned. The confidence intervals are constructed as in Section 7.1, applied to the X_t component. For graphical clarity, we do not report results obtained through the quantile-based composite estimator $\widehat{\mathrm{XMES}}^{\star}_{X,\widehat{\tau}'_n(\alpha_n)}$, which were in line with those shown here.

The bottom row of Figure 1 gives results at level $\tau'_n(\alpha_n) = 0.9995$. The intervals for the linear time series have broadly the same properties as in the upper panels. The coverages of our proposed confidence intervals are very good 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 estimators providing reasonable results for a narrower range of values of k. Nevertheless, our intervals represent a significant improvement over those relying on the theory for i.i.d. observations and, with the aforementioned rule of thumb (selecting $k \in [50, 200]$), they 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.

8 Financial data analyses

8.1 Dow Jones and S&P 500 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. We provide marginal estimates and dynamic predictions of the future given past observations.

Estimation of marginal risk measures 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 τ_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 $\widehat{\tau}_n'(\alpha_n)$. Then we estimate the expectile at the extreme level $\widehat{\tau}_n'(\alpha_n)$ using the composite extrapolating LAWS estimator $\widetilde{\xi}^{\star}_{\tau_n'}$ and the corresponding quantile-based estimator $\widehat{\xi}^{\star}_{\tau_n'}$ for $\tau_n' = \widehat{\tau}_n'(\alpha_n)$. 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 in Figure 2 plot $\widehat{\tau}_n'(\alpha_n)$ 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 representing the composite extrapolating LAWS and quantile-based estimators and the confidence intervals of Section 5 at level $\tau_n' = \widehat{\tau}_n'(\alpha_n)$ in the middle two panels of Figure 2. They appear to be fairly stable when k is not too small, and $k = 200$ again seems sensible. As expected, the confidence intervals that allow for dependence are wider than those that do not. 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, both rather higher than the original $\alpha_n = 1 - p_n = 0.9998862$.

Figure 2 also compares our extrapolating LAWS and quantile-based estimators with the Weissman extreme quantile estimator at level α_n, i.e.,

$$\widehat{q}^{\star}_{\alpha_n} = \left(\frac{1-\alpha_n}{1-\tau_n}\right)^{-\widehat{\gamma}_n^H} \widehat{q}_{\tau_n} = \left(\frac{1-\alpha_n}{1-\tau_n}\right)^{-\widehat{\gamma}_n^H} Y_{n-\lfloor n(1-\tau_n)\rfloor,n}. \tag{8}$$

Confidence intervals for the extreme quantile q_{α_n} can also be constructed using this estimator; here we use a method of Drees (2003) for the estimation of the variance component $w(\gamma, R)$ in Section 5, and label this approach WEISS-D-ADJ. 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 Figure 2. The point estimates are reassuringly similar, but the third and fourth panels suggest that the confidence intervals based on $\widehat{q}^{\star}_{\alpha_n}$ are generally much more volatile than the LAWS interval; moreover, in a neighbourhood of $k \approx 200$, they are very close to the intervals based on i.i.d. 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 LAWS intervals contain the sample maximum.

Dynamic prediction of extreme risk Here we focus on the Dow Jones data. Starting on 29 January 1985, we fitted GARCH$(1,1)$ models using a Student$-t$ likelihood on rolling windows $(Y_j, \ldots, Y_{j+T-1})$ of size $T = 1{,}000$, corresponding to approximately four years of trading data. The choice of model was based on the AIC after fitting several classical heteroskedastic models to the last T data points in the series (from 15 December 2015 to 12 December 2019). The model was checked using the correlograms of the residuals and their squares, and the weighted Ljung–Box and Li–Mak tests of Fisher & Gallagher (2012) (calculations were performed using the R package `rugarch`, see Galanos & Kley 2022). All tests suggested that the residuals may be considered to be independent. This is supported by the calculation (on the last T days of data) of asymptotic confidence intervals for extreme expectiles and quantiles based on the dependence assumption, which tend to be very close to those based on i.i.d. theory; see the top row of Figure 4. The inference method of Drees (2003) provides unreasonably wide confidence intervals. The data $(Y_j, \ldots, Y_{j+T-1})$ are then filtered using the estimated GARCH$(1,1)$ structure, resulting in residuals $(\widehat{\varepsilon}_1^{(j)}, \ldots, \widehat{\varepsilon}_T^{(j)})$ that can be used to perform extreme-value predictions for Y_{j+T} given its past, following the ideas of Section 6. We took $k = 100$ based on a graphical analysis of the residual-based Hill estimator. These predictions, based on extreme expectile estimates and quantile estimates at the level $\tau_n' = 0.999$, are shown in the top row of Figure 4 from 16 January 1989 to 12 December 2019.

The predicted expectiles, which seem broadly correct and mirror the volatility bursts in the data, are substantially lower than the predicted quantiles. We checked estimation accuracy by counting the number of exceedances above our predictions: 10 (resp. 5) exceedances are observed above the LAWS (resp. quantile-based) expectile prediction, while only one exceedance is observed above the predicted quantile, although $(n - 1{,}000) \times (1 - 0.999) \approx 8$ exceedances above the 0.999-quantile are expected. For expectiles, the benchmark is harder to establish; the residual-based Hill estimator over the last T days of data was approximately 0.4, meaning that, by inverting (7), a rough estimate of the quantile level corresponding to the 0.999-expectile level is $1-(1-0.999)\times(1-0.4)/0.4 \approx 0.9985$. This suggests

that we should expect approximately $(n - 1{,}000) \times (1 - 0.9985) \approx 12$ exceedances above the dynamic 0.999-expectile, with our LAWS estimate achieving a number of exceedances close to this target. In this application the LAWS method seems to give better results than those for quantile-based estimation, perhaps because it tends to be more accurate than the latter when applied to zero-mean distributions, cf. Daouia et al. (2018).

8.2 Systemic risk measure estimation using financial returns

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 Figure 3 display values of $\widehat{\tau}'_n(\alpha_n)$ against k with again $\alpha_n = 1 - 1/n = 0.9996021$ and $\tau_n = 1 - k/n$.

These estimates initially fluctuate slightly, then stabilise, 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 LAWS estimator and the extrapolating QMES estimator of $\mathrm{QMES}_{X,\alpha_n}$, and the various confidence intervals at level $\tau'_n = \widehat{\tau}'_n(\alpha_n)$ are reported in the middle two 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 $\widehat{\tau}'_n(\alpha_n) \approx 0.9997239$ for Goldman Sachs and 0.9996626 for Morgan Stanley. All estimates are reported in Table 2 and the LAWS estimates are shown in the rightmost panels of Figure 3; they lead to conclusions similar to those in Section 8.1. For $k = 150$ they give similar results for Goldman Sachs and somewhat shorter confidence intervals for Morgan Stanley. As with the stock market index data, the confidence intervals constructed for $\mathrm{QMES}_{X,\alpha_n}$ with the

method of Drees (2003) are more volatile than ours.

8.3 Individual stock prediction given a market index

We finally consider a data set comprising the excess daily log-returns on General Electric and on the S&P 500 index, from 1 November 1993 to 31 March 2003; see the data frame `capm` in the R package `HRW` (Harezlak et al. 2021). Excess daily log-returns are obtained by subtracting the risk-free interest rate from raw daily log-returns (Harezlak et al. 2018, p.194). We denote by X_t (resp. Y_t) the negative excess daily log-return of the S&P 500 index (resp. General Electric). We focus on estimating the extreme conditional risk on General Electric given a value x of the (negative excess) return on the total market for the same day. A scatterplot of the data can be found in the bottom row of Figure 4.

An approximately linear relationship appears to link Y_t to X_t, though the presence or not of heteroskedasticity is less clear. Following the general procedure described in Section 6, we first estimate a linear regression model linking Y_t to X_t by ordinary least squares, leading to the estimated regression line $\widehat{g}(x) = \widehat{a} + \widehat{b}x = -0.000297 + 1.244\,x$. We then estimate the conditional variance of Y_t given $X_t = x$ by a residual-based local polynomial estimator, that is, $\widehat{\sigma}^2(x) = \widehat{c}_0 = \widehat{c}_0(x)$ where

$$(\widehat{c}_0, \widehat{c}_1) = \underset{(c_0,c_1)\in\mathbb{R}^2}{\arg\min} \sum_{t=1}^{n} (\widehat{\eta}_t^2 - c_0 - c_1(X_t - x))^2 K\left(\frac{X_t - x}{h}\right).$$

Here $\widehat{\eta}_t = Y_t - (\widehat{a} + \widehat{b}X_t)$ are the raw residuals from the regression model, K denotes the Gaussian kernel and $h \approx 0.00420$ is a bandwidth parameter, selected using the `dpill` function from the R package `KernSmooth` (Wand et al. 2021) with default settings. This estimate, shown in the first panel of the bottom row of Figure 4, leads to standardized residuals $\widehat{\varepsilon}_t = \widehat{\eta}_t/\widehat{\sigma}(X_t)$ shown in the the second panel and which appear to remain somewhat temporally dependent. Since the conditional variance estimator is nonparametric, it will tend to be inaccurate where data are sparse, so our extreme value analysis uses only those $\widehat{\varepsilon}_t$ whose corresponding X_t lie in $[-0.02, 0.02]$. We finally estimate an extreme expectile of level $\tau_n' = 1 - 1/n \approx 0.9995768$ with $k = 150$, using the residual-based LAWS estimator $\widetilde{\xi}_{\tau_n'}(\varepsilon)$, resulting in an estimate $\widetilde{\xi}_{\tau_n'}(Y \mid X = x) = -0.000297 + 1.244\,x + \widehat{\sigma}(x)\widetilde{\xi}_{\tau_n'}(\varepsilon)$ of the target extreme conditional expectile. These estimates and the corresponding 95%

confidence intervals are shown in the third and fourth panels of the bottom row of Figure 4.

The intervals based on our theory for dependent data are appreciably wider than those that presuppose independence, reflecting the temporal dependence in the data that the linear regression model did not eliminate. Moreover, and as expected, larger losses in the market lead to potentially larger extreme losses, although the relationship does not appear to be monotonic or linear due to the nonlinear profile of the conditional variance in the heteroskedastic model. An analogous analysis in which the regression of Y on X was estimated using local linear regression, rather than ordinary least squares, produced the same results. This suggests that only a broadly correct estimate of the regression function is required, with an accurate estimation of the variance component and extreme behaviour of the residuals being more important.

9 Conclusion

We discussed extreme risk inference via expectiles in heavy-tailed time series, from the marginal and conditional or dynamic prediction perspectives. The proposed inferential framework relies on β-mixing assumptions and large-sample theory and provides confidence intervals whose coverage is reasonably close to nominal, and which greatly improves on those based on independence.

We defer to future work the consideration of yet more general dependence frameworks, such as strong mixing. Another important topic is bias correction: Girard et al. (2021*b*) showed that bespoke bias reduction methods for extreme expectile estimation can greatly improve results when tail heaviness is moderate. It will be interesting to investigate the performance of bias correction methods with time series data.

Acknowledgements

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

Supplementary material

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

References

Acharya, V. V., Pedersen, L. H., 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.

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 β-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. (2019), 'Extreme M-quantiles as risk measures: From L^1 to L^p optimization', *Bernoulli* **25**, 264–309.

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.

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.

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

Drees, H. (2000), 'Weighted approximations of tail processes for β-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.

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

Fisher, T. J. & Gallagher, C. M. (2012), 'New weighted portmanteau statistics for time series goodness of fit testing', *Journal of the American Statistical Association* **107**, 777–787.

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.

Galanos, A. & Kley, T. (2022), *rugarch: Univariate GARCH Models*. R package version 1.4–6.

Girard, S., Stupfler, G. & Usseglio-Carleve, A. (2021*a*), 'Extreme conditional expectile estimation in heavy-tailed heteroscedastic regression models', *Annals of Statistics* **49**, 3358–3382.

Girard, S., Stupfler, G. & Usseglio-Carleve, A. (2021*b*), On automatic bias reduction for extreme expectile estimation. hal-03086048.

Gneiting, T. (2011), 'Making and evaluating point forecasts', *Journal of the American Statistical Association* **106**, 746–762.

Harezlak, J., Ruppert, D. & Wand, M. P. (2018), *Semiparametric Regression with R*, Springer Science+Business Media.

Harezlak, J., Ruppert, D. & Wand, M. P. (2021), *HRW: Datasets, Functions and Scripts for Semiparametric Regression Supporting Harezlak, Ruppert & Wand (2018)*. R package version 1.0–5.

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

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.

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. G. & Caeiro, F. (2013), *evt0: Mean of Order p, Peaks Over Random Threshold Hill and High Quantile Estimates*. R package version 1.1–3.

McNeil, A. J. & Frey, R. (2000), 'Estimation of tail-related risk measures for heteroscedastic financial time series: an extreme value approach', *Journal of Empirical Finance* **7**, 271–300.

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.

Wand, M. P., Moler, C. & Ripley, B. (2021), *KernSmooth: Functions for Kernel Smoothing Supporting Wand & Jones (1995)*. R package version 2.23–20.

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.

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
$\widehat{\gamma}_n^H$	0.336 [0.220, 0.453]	0.344 [0.222, 0.467]
$\widetilde{\xi}^{\star}_{\widehat{\tau}'_n(\alpha_n)}$	0.136 [0.064, 0.259]	0.136 [0.061, 0.263]
$\widehat{\xi}^{\star}_{\widehat{\tau}'_n(\alpha_n)}$	0.140 [0.064, 0.258]	0.139 [0.061, 0.260]
$\widehat{q}^{\star}_{\alpha_n}$	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
$\widehat{\gamma}_{X,n}^H$	0.410 [0.172, 0.648]	0.459 [0.279, 0.639]
$\widetilde{\mathrm{XMES}}^{\star}_{X,\widehat{\tau}'_n(\alpha_n)}$	0.342 [0.079, 1.046]	0.590 [0.185, 1.303]
$\widehat{\mathrm{QMES}}^{\star}_{X,\alpha_n}$ (CI via Section 5)	0.345 [0.080, 1.055]	0.603 [0.189, 1.332]
$\widehat{\mathrm{QMES}}^{\star}_{X,\alpha_n}$ (CI via Drees (2003))	0.345 [0.216, 0.551]	0.603 [0.294, 1.236]

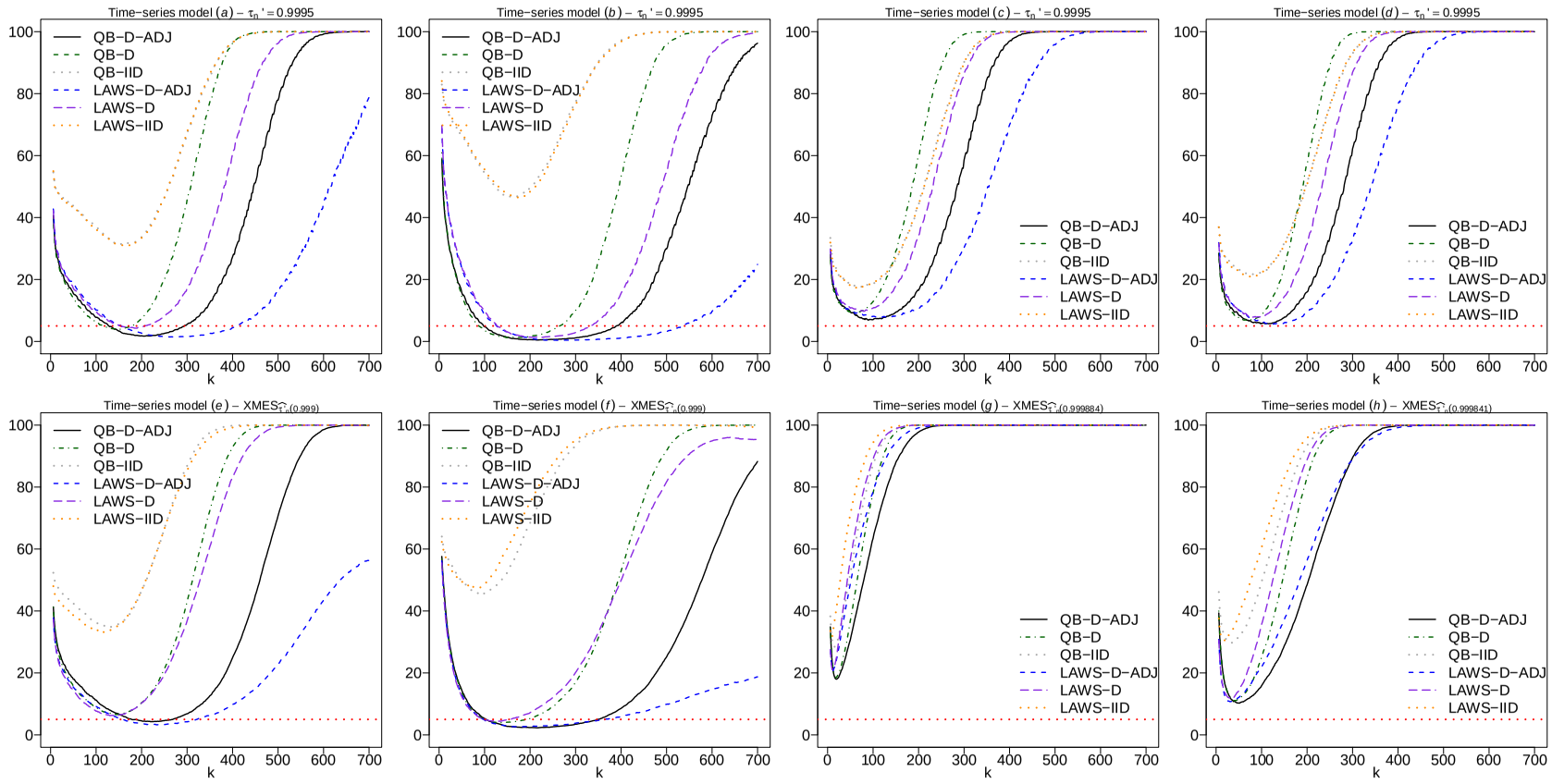

Figure 1: Empirical error rate (%) for nominal 95% confidence intervals for the expectile $\xi_{\tau_n'}$ (top row, with $\tau_n' = 0.9995$) and the marginal expected shortfall $\mathrm{QMES}_{X,\alpha_n}$ (bottom row, with level α_n such that $\tau_n'(\alpha_n) = 0.9995$). The true model is indicated 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 $\widehat{b} = 0$), QB-IID (ignoring the dependence), and likewise for the LAWS estimators. In the bottom panels the QB-D-ADJ, QB-D and QB-IID confidence intervals are constructed upon the estimator $\widehat{\mathrm{QMES}}^{\star}_{X,\alpha_n}$.

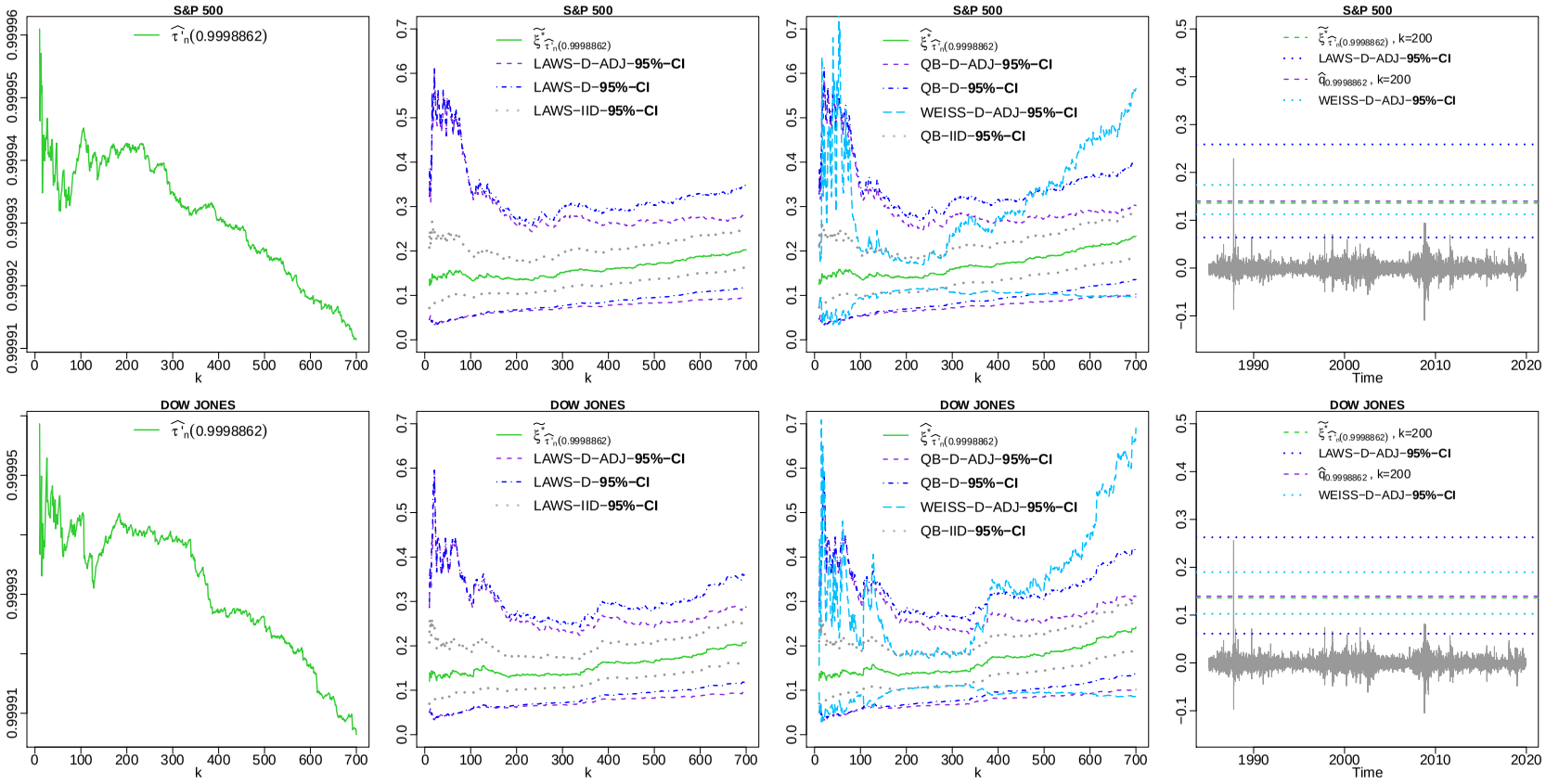

Figure 2: S&P 500 and Dow Jones Industrial Average data. From left to right: estimate $\widehat{\tau}'_n(\alpha_n)$, with $\alpha_n = 0.9998862$; marginal composite extrapolating LAWS estimate at level $\widehat{\tau}'_n(\alpha_n)$, with 95% LAWS-IID, LAWS-D and LAWS-D-ADJ confidence intervals; marginal composite extrapolating quantile-based estimate at level α_n, with 95% QB-IID, QB-D, QB-D-ADJ and WEISS-D-ADJ confidence intervals; the data, with marginal composite extrapolating LAWS estimate at level $\widehat{\tau}'_n(\alpha_n)$ and Weissman quantile estimate $\widehat{q}^{\star}_{\alpha_n}$, both with 95% confidence intervals.

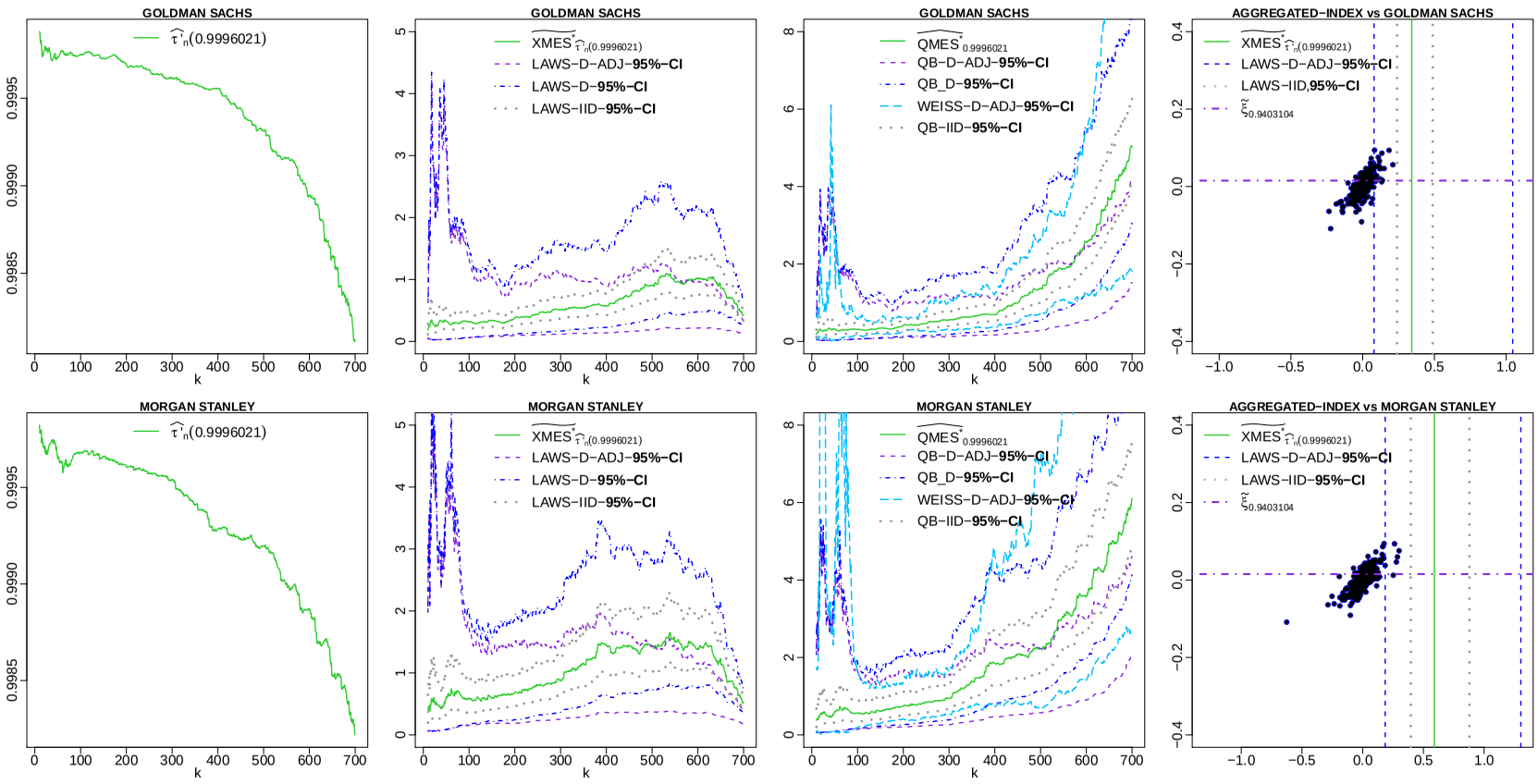

Figure 3: Goldman Sachs and Morgan Stanley data. From left to right: estimate $\widehat{\tau}'_n(\alpha_n)$, with $\alpha_n = 0.9996021$; composite extrapolating LAWS estimate of (5) at level $\widehat{\tau}'_n(\alpha_n)$, with 95% LAWS-IID, LAWS-D and LAWS-D-ADJ confidence intervals; extrapolating estimate of (4) at level α_n, with 95% QB-IID, QB-D, QB-D-ADJ and WEISS-D-ADJ confidence intervals; scatterplot of the data, with the composite extrapolating LAWS estimate of (5) at level $\widehat{\tau}'_n(\alpha_n)$ with 95% confidence intervals. The purple horizontal line is the intermediate level τ_n. The WEISS-D-ADJ confidence intervals are based on (8); see the caption to Figure 1 for the other abbreviations.

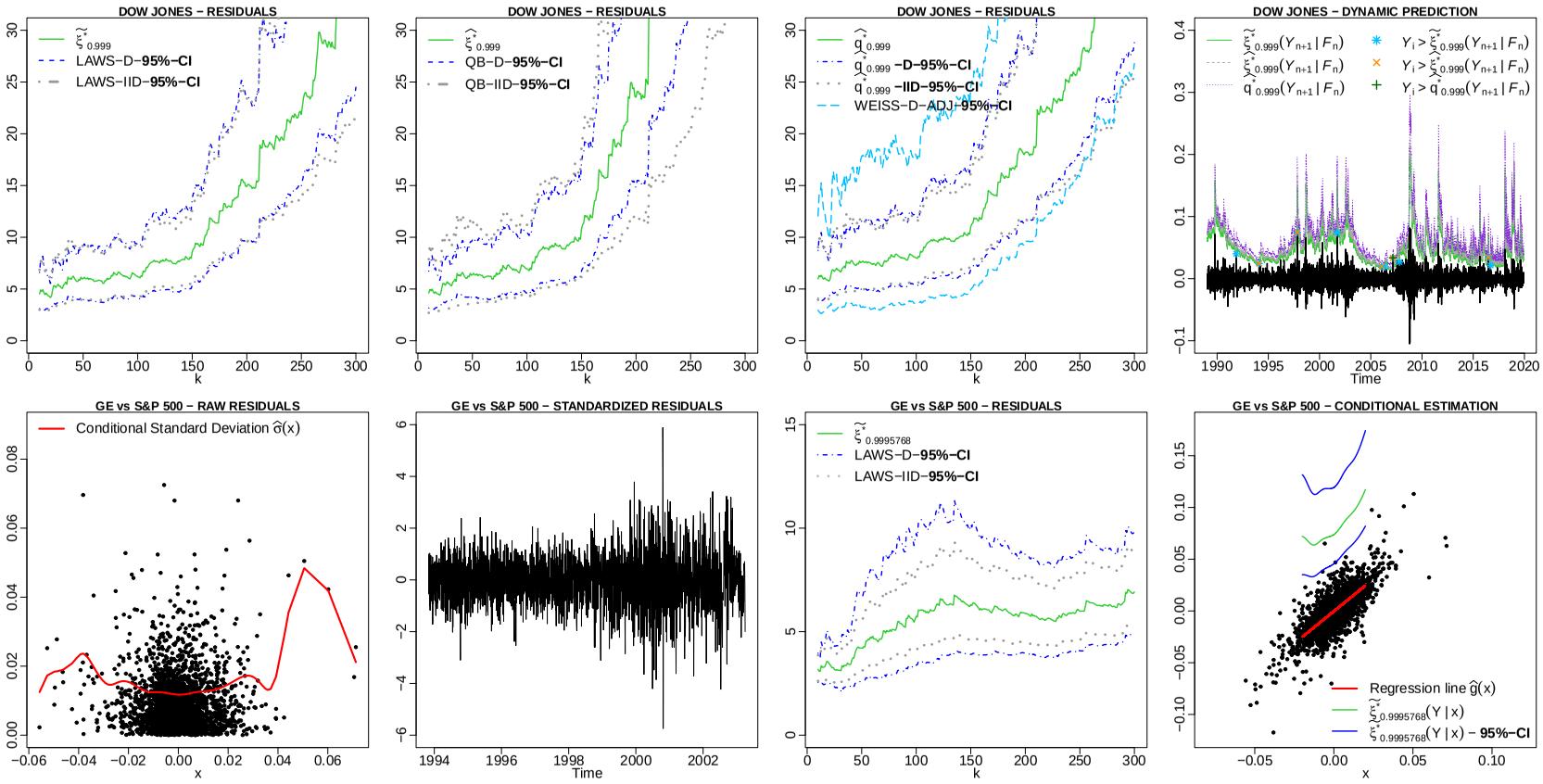

Figure 4: Top: Dow Jones Industrial Average data. From left to right: extrapolating LAWS estimate $\widetilde{\xi}^{\star}_{\tau'_n}(\varepsilon)$ at level $\tau'_n = 0.999$, with 95% LAWS-IID and LAWS-D confidence intervals; extrapolating quantile-based estimate $\widehat{\xi}^{\star}_{\tau'_n}(\varepsilon)$ at level τ'_n, with 95% QB-IID and QB-D confidence intervals; Weissman quantile estimate $\widehat{q}^{\star}_{\tau'_n}(\varepsilon)$ at level τ'_n, with 95% confidence intervals derived from i.i.d. theory, our dependence framework, and the WEISS-D-ADJ method; resulting dynamic predictions of extreme expectiles and quantiles at level $\tau'_n = 0.999$ superimposed to the data. Crosses denote exceedances above the predictions. Bottom: General Electric and S&P 500 data. From left to right: absolute raw residuals and conditional standard deviation estimate $\widehat{\sigma}$; standardized residuals; extrapolating residual-based LAWS estimate $\widetilde{\xi}^{\star}_{\tau'_n}(\varepsilon)$ at level $\tau'_n = 0.9995768$, with 95% LAWS-IID and LAWS-D confidence intervals; scatterplot of the data, with regression line and conditional prediction of the extreme expectile at level $\tau'_n = 0.9995768$ of the General Electric stock given the value of the S&P 500 index, with LAWS-D confidence intervals.