Limit Theory for Forecasts of Extreme Distortion Risk Measures and Expectiles

Yannick Hoga*

July 5, 2018

Abstract

In general location-scale models for asset returns, we develop central limit theory for tail risk forecasts. We do so for a wide range of risk measures, viz. distortion risk measures and expectiles. Two popular members of the class of distortion risk measures are the Value-at-Risk and the Expected Shortfall. The estimators we consider are motivated by a Pareto-type tail assumption and allow for extrapolation beyond the range of available observations. Simulations reveal that the finite-sample distributions are adequately approximated by the asymptotic distributions. An empirical application demonstrates that in sufficiently large samples our estimators outperform non-parametric alternatives when forecasting extreme risk.

Keywords: Central limit theory, Distortion risk measures, Expectiles, Extreme value theory, Location-scale model

JEL classification: C14 (Semiparametric and Nonparametric Methods), C22 (Time-Series Models), C58 (Financial Econometrics)

1 Motivation

Forecasting risk is one of the main tasks in financial risk management. Risk forecasts are deeply ingrained in the regulatory framework of the insurance industry (Solvency II) and the banking industry (Basel II and III). They are also used heavily within financial institutions to, e.g., set trading limits for traders (Jorion, 2006). As with other dynamic forecasts, there is estimation uncertainty in risk forecasts. A number of authors have quantified this estimation uncertainty; e.g., Gao and Song (2008), Linton and Xiao (2013), Francq and Zakoïan (2015) and Wang and Zhao (2016). However, these authors exclusively focus on Value-at-Risk (VaR), i.e., the value that is only exceeded with some small probability δ, and/or Expected Shortfall (ES), i.e., the average loss given a VaR exceedance. Also, their estimators cannot be adapted to the estimation of extreme risks, that may lie outside the range of observations.

*Faculty of Economics and Business Administration, University of Duisburg-Essen, Universitätsstraße 12, D-45117 Essen, Germany, yannick.hoga@vwl.uni-due.de. Support of DFG through grant HA 6766/2-2 is gratefully acknowledged.

However, adopting estimators—motivated by extreme value theory (EVT)—that can do so has been shown to be beneficial in many comparative studies of forecasting performance (e.g., McNeil and Frey, 2000; Gençay and Selçuk, 2004; Chan and Gray, 2006; Kuester *et al.*, 2006). To the best of our knowledge, only Chan *et al.* (2007), Martins-Filho *et al.* (2018) and Hoga (2018+a) study the estimation risk of estimators based on extreme value theory in a time series context.[1] Chan *et al.* (2007) and Hoga (2018+a) only derive their results for (ARMA–)GARCH models. A drawback of Martins-Filho *et al.* (2018) is that they assume finite fourth moments of the innovation distribution. This limits applicability in case of rather heavy-tails, where extreme value methods work particularly well vis-à-vis non-parametric alternatives (Hoga, 2018+c, Sec. 3). Again, they all consider estimation of only VaR and ES.

While VaR and ES are perhaps the most widely-used risk measures, they each have their distinct drawbacks. From a theoretical and a practical viewpoint, two desirable properties of risk measures are *coherence* (Artzner *et al.*, 1999) and *elicitability* (Gneiting, 2011). Coherence refers to a set of four elementary axioms for risk measures—translation invariance, subadditivity, positive homogeneity, and monotonicity. Elicitability is crucial for sensible comparisons of forecasting performance, as it entails the existence of a suitable—not necessarily unique—loss (or also: score) function with which to compare different sets of forecasts. VaR is not coherent but elicitable, whereas ES is coherent but not elicitable (Artzner *et al.*, 1999; Gneiting, 2011).

The good performance of EVT-based estimators and the drawbacks of VaR and ES motivate the study of EVT-based estimators of more general risk measures in this paper. Specifically, in general location-scale models, we quantify the uncertainty in forecasts of a wide range of extreme risk measures under some high-level conditions. Thus, the theory we develop is very general, both in the model assumed to be driving the returns and the choice of the risk measure. As regards the allowable risk measures, we consider extreme versions of Wang's (1996) distortion risk measures (DRMs) introduced by El Methni and Stupfler (2017). The class of Wang (1996) DRMs is very wide and includes—among many others—VaR, ES and the Wang (2000) transform; see El Methni and Stupfler (2017, Table 1) for a more complete overview.

Despite the generality of DRMs, the mean and VaR are the only elicitable DRMs (Wang and Ziegel, 2015; Kou and Peng, 2016). Thus, we shall also consider expectiles (Newey and Powell, 1987), which have recently become popular in risk management applications. For an extensive overview of the theoretical and practical appeal of expectiles in risk management, we refer to Daouia *et al.* (2018). One advantage of expectiles is that they are elicitable. In fact, Ziegel (2016) shows that expectiles are the only elicitable law-invariant[2] coherent risk measures. Nonetheless, DRMs may be *jointly* elicitable. For instance, VaR and ES are jointly elicitable, which allows for sensible forecast comparisons of the

[1]The literature that proposes extreme value methods to forecast risk without providing a measure of statistical uncertainty is much wider. Some references include McNeil and Frey (2000), Mancini and Trojani (2011), Chavez-Demoulin *et al.* (2014) and Bee *et al.* (2016).

[2]A risk measure is called *law-invariant* if it assigns the same (scalar) risk to X and Y whenever $X \overset{\mathcal{D}}{=} Y$.

pair (VaR, ES) (Fissler and Ziegel, 2016).

Expectiles also have drawbacks. First, they are not as intuitive to grasp as VaR and ES; see Subsection 2.1. Second, and perhaps more importantly, they depend on the complete probability distribution of the underlying random variable, whereas extreme DRMs—including VaR and ES in particular—specifically focus on the tail of interest. Third, by Proposition 2.2 of Emmer *et al.* (2015), expectiles are not comonotonically additive[3], i.e., they may attribute diversification benefits to comonotonic risks. This is particularly undesirable in portfolio construction (Emmer *et al.*, 2015, Sec. 4.2). For a more complete, succinct overview of the relative merits of VaR, ES and expectiles we refer to Table 1 of Emmer *et al.* (2015); see also Dowd and Blake (2006) for some in-depths discussion of VaR and ES. Summarizing, it is unlikely that there exists a universally preferred risk measure in any given situation based on both theoretical and practical considerations. This echoes Dowd and Blake (2006, p. 220), who emphasize that 'the most appropriate risk measure depends on the assumptions we make [...] and would appear also to be sometimes context-dependent. Any search for a single "best" risk measure—one that is best in all conceivable circumstances—would appear to be futile.' Thus, a sufficiently general estimation theory covering a wide class of risk measures, as presented in this paper, is desirable.

Sometimes the information conveyed by a point forecast of risk together with a confidence interval is not sufficient. For instance, Wang and Zhao (2016, p. 90) argue that it is important to allow for simultaneous inference of conditional risk measures at different risk levels, as this allows to dynamically manage the overall portfolio risk at different levels. Francq and Zakoïan (2016) argue in a similar vein. Thus, we also derive some uniform limit theory for our conditional risk measures that allows to construct confidence corridors for tail risk.

The paper is structured as follows. Section 2 presents the main results. It proceeds by first introducing extreme DRMs and expectiles (Subsection 2.1) along with suitable estimators (Subsection 2.2). Subsection 2.3 transfers these risk measures and estimators to a conditional setting in a general location-scale model. Finally, Subsection 2.4 presents pointwise and uniform asymptotic results for the estimators of conditional tail risk. The quality of the asymptotic approximations is assessed in a Monte Carlo study in Section 3. Section 4 presents an empirical application to asset returns, where EVT-based and non-parametric forecasting approaches are compared for a wide range of asset returns. The final section concludes.

[3]Let L_1 and L_2 be comonotonic r.v.s, i.e., there exist a r.v. X, and non-decreasing functions f_1 and f_2, s.t. $L_1 = f_1(X)$ and $L_2 = f_2(X)$. A risk measure ρ is then said to be *comonotonically additive*, if $\rho(L_1 + L_2) = \rho(L_1) + \rho(L_2)$.

2 Main results

2.1 Risk measures

In this section, we first define extreme DRMs and then expectiles. Let X denote a random variable with distribution function (d.f.) $F(\cdot)$ and quantile function $q(\cdot) = F^{\leftarrow}(\cdot)$. We assume X to be positive for the moment. Wang (1996) introduces the general family of DRMs, which are based on a *distortion function* $g : [0,1] \to [0,1]$, i.e., a non-decreasing and right-continuous function with $g(0) = 0$ and $g(1) = 1$. Then, the DRM of X with distortion function $g(\cdot)$ is

$$\mathrm{DRM}(X) := \int_0^\infty g(1 - F(x))\,\mathrm{d}x. \tag{1}$$

This representation nests several well-known risk measures, such as the Value-at-Risk at level δ, $q_\delta = F^{\leftarrow}(\delta)$ (for $g(x) = I_{\{x \geq 1-\delta\}}$), and the Expected Shortfall at level δ, $\mathrm{ES}_\delta = E[X \mid X > q_\delta]$ (for $g(x) = \min\left\{x/(1-\delta), 1\right\}$). Due to their generality, Wang (1996) DRMs have been studied intensively (Wirch and Hardy, 2001; Dowd and Blake, 2006). Wirch and Hardy (2001) show that a DRM is coherent if and only if $g(\cdot)$ is concave.

To investigate extreme risks, where the risk level $\delta \uparrow 1$ gets more extreme, El Methni and Stupfler (2017) introduce an extreme version of DRMs with distortion function not depending on δ. Define $F_\delta(x) = \max\left\{0, \frac{F(x)-\delta}{1-\delta}\right\}$. If the quantile function $q(\cdot)$ is continuous and strictly increasing in a neighbourhood of δ, then

$$F_\delta(x) = \max\left\{0, \frac{F(x) - F(q(\delta))}{1-\delta}\right\} = P\left\{X \leq x \mid X > q(\delta)\right\}.$$

Thus, the *extreme DRM* of El Methni and Stupfler (2017), defined as

$$\mathrm{DRM}_\delta := \mathrm{DRM}_\delta(X) := \int_0^\infty g(1 - F_\delta(x))\,\mathrm{d}x,$$

can be interpreted as the DRM of X given $X > q_\delta$. Letting $\delta \uparrow 1$ then leads to an extreme risk measure. As before, VaR_δ and ES_δ can be obtained as special cases for $g(x) = I_{\{x=1\}}$ and $g(x) = x$, respectively. If the quantile function $q(\cdot)$ is continuous and strictly increasing in a neighbourhood of infinity, then

$$\mathrm{DRM}_\delta = \int_0^1 F^{\leftarrow}(1 - (1-\delta)\alpha)\,\mathrm{d}g(\alpha) \tag{2}$$

for δ sufficiently close to 1. In this notation, DRM_δ can also be used for real-valued X. So at every appearance, X refers to a real random variable from now on.

Next, we introduce expectiles by analogy with quantiles, with which they share some conceptual similarities. Quantiles can be obtained as

$$q_\delta := q_\delta(X) := \operatorname*{arg\,min}_{q \in \mathbb{R}} \mathrm{E}[\eta_{\delta,1}(X - q)], \tag{3}$$

where $\eta_{\delta,m}(x) = |\delta - I_{\{x\le 0\}}| \cdot |x|^m$. Expectiles are defined via

$$\xi_\delta := \xi_\delta(X) := \arg\min_{q\in\mathbb{R}} \mathrm{E}[\eta_{\delta,2}(X-q)]. \tag{4}$$

The minimisation problem in (4) is well-defined if $\mathrm{E}\,|X| < \infty$.

Incidentally, the properties (3) and (4) naturally lead to elicitability of both quantiles and expectiles (Gneiting, 2011). For fixed $\delta > 0$, the loss (scoring) function $s_m : \mathbb{R}\times\mathbb{R} \to [0,\infty)$ can simply be chosen as $s_m(q,x) := \eta_{\delta,m}(x-q)$ for quantiles ($m = 1$) and expectiles ($m = 2$). Given a set of K competing forecasts $y_1^{(k)}, \ldots, y_n^{(k)}$ ($k = 1,\ldots,K$) and realised observations $x_1,\ldots,x_n$, a forecast ranking can be obtained by comparing the average scores

$$\overline{s}_m^{(k)} = \frac{1}{n}\sum_{i=1}^{n} s_m(y_i^{(k)}, x_i). \tag{5}$$

The set of forecasts $y_1^{(k)}, \ldots, y_n^{(k)}$ with the smallest score is preferred.

For $\delta = 1/2$, q_δ equals the median and ξ_δ equals the mean. For general $\delta \in (0,1)$, we can derive from a first-order condition for the minimisation problem in (3),

$$\delta = \frac{\int_{-\infty}^{q_\delta} \mathrm{d}F(x)}{\int_{-\infty}^{q_\delta} \mathrm{d}F(x) + \int_{q_\delta}^{\infty} \mathrm{d}F(x)},$$

and for the minimisation problem in (4) that

$$\delta = \frac{\int_{-\infty}^{\xi_\delta} |x-\xi_\delta|\,\mathrm{d}F(x)}{\int_{-\infty}^{\xi_\delta} |x-\xi_\delta|\,\mathrm{d}F(x) + \int_{\xi_\delta}^{\infty} |x-\xi_\delta|\,\mathrm{d}F(x)} = \frac{\mathrm{E}\left[|X-\xi_\delta| I_{\{X\le\xi_\delta\}}\right]}{\mathrm{E}\,|X-\xi_\delta|}.$$

Thus, q_δ specifies the value where the ratio of the average *number of observations* of the data below q_δ to the average *number of observations* of the data above and below q_δ is $100\delta\%$. In contrast, ξ_δ specifies the value where the ratio of the *average distance* of the data below ξ_δ to the *average distance* of the data above and below ξ_δ is $100\delta\%$. This interpretation demonstrates that expectiles not only depend on one tail of the distribution, like VaR and ES, but rather on the whole distribution. This may or may not be desirable from a risk management perspective; cf. Kuan *et al.* (2009, Sec. 2.2).

2.2 Estimation

We first introduce an estimator of extreme VaR, because estimators of extreme DRMs and expectiles are directly based on it. Let X denote a random variable with d.f. $F(\cdot)$. As mentioned in the Motivation, we consider random variables with Pareto-type tail. To make this more precise, let $U(t) = F^{\leftarrow}(1-1/t)$. Then, we say that X has Pareto-type tail if $U(\cdot)$ is regularly varying with *extreme value index* γ, i.e.,

$$\lim_{t\to\infty} \frac{U(tx)}{U(t)} = x^\gamma \qquad \text{for all } x > 0. \tag{6}$$

This condition is satisfied for the tails of a wide range of data, e.g., stock returns, trading volume, and city sizes (Gabaix, 2009).

Relation (6) leads to the following approximate relation between extreme quantiles $q_\beta = U(1/[1-\beta])$ and $q_\delta = U(1/[1-\delta])$ with $\beta < \delta$, both close to 1:

$$q_\delta \approx \left(\frac{1-\beta}{1-\delta}\right)^{\gamma} q_\beta. \tag{7}$$

This suggests that very extreme quantiles q_δ can be estimated by estimating a less extreme 'anchor' quantile q_β and then use the Pareto shape of the tail to extrapolate to q_δ via $([1-\beta]/[1-\delta])^{\gamma}$.

Let $X_1, \dots, X_n$ denote a sample of X. We mimic the asymptotics in (6) by imposing that β and δ converge to 1 as the sample size n gets larger.

Assumption 1. *The sequences β_n and δ_n both tend to* 1 *with $n(1-\beta_n) \to \infty$, $n(1-\delta_n) \to c > 0$, as $n \to \infty$.*

The sequences β_n and δ_n appearing subsequently are precisely those of Assumption 1.

Motivated by (7), we introduce the Weissman (1978) estimator

$$\widehat{q}_{\delta_n} := \widehat{q}_{\delta_n}(X) = \left(\frac{1-\beta_n}{1-\delta_n}\right)^{\widehat{\gamma}} X_{\lceil n\beta_n \rceil, n}, \tag{8}$$

where $X_{1,n} \leq \ldots \leq X_{n,n}$ denote the order statistics and $\widehat{\gamma} = \widehat{\gamma}(X)$ denotes a generic estimator of γ based on $X_1 \dots, X_n$. For instance, one could use the Hill (1975) estimator

$$\widehat{\gamma} = \widehat{\gamma}(X) = H_n(\lceil n(1-\beta_n)\rceil) \qquad \text{with} \quad H_n(k) = \frac{1}{k}\sum_{i=1}^{k} \log\left(X_{n-i+1,n}/X_{n-k,n}\right). \tag{9}$$

Since under the Pareto-type tail assumption (6) there is a close connection in the tail between quantiles and extreme DRMs/expectiles, one can use $\widehat{q}_{\delta_n}$ to estimate extreme DRMs and expectiles. For extreme DRMs the key result is Lemma 3 in El Methni and Stupfler (2017), which shows that under (6)

$$\frac{\mathrm{DRM}_{\delta_n}}{q_{\delta_n}} \underset{(n\to\infty)}{\longrightarrow} \int_0^1 s^{-\gamma}\,\mathrm{d}g(s), \tag{10}$$

provided $\int_0^1 s^{-\gamma-\iota}\,\mathrm{d}g(s) < \infty$ for some $\iota > 0$. Bellini and Di Bernardino (2017, Prop. 2.3) provide the corresponding result for expectiles, which states that under (6) with $\gamma \in (0,1)$,

$$\frac{\xi_{\delta_n}}{q_{\delta_n}} \underset{(n\to\infty)}{\longrightarrow} (\gamma^{-1}-1)^{-\gamma}. \tag{11}$$

Relations (10) and (11) provide the motivation for the plug-in estimators

$$\begin{aligned} \widehat{\mathrm{DRM}}_{\delta_n} &:= \widehat{\mathrm{DRM}}_{\delta_n}(X) := \int_0^1 s^{-\widehat{\gamma}}\,\mathrm{d}g(s)\widehat{q}_{\delta_n} \qquad \text{and} \\ \widehat{\xi}_{\delta_n} &:= \widehat{\xi}_{\delta_n}(X) := \left(\widehat{\gamma}^{-1}-1\right)^{-\widehat{\gamma}} \widehat{q}_{\delta_n}, \end{aligned} \tag{12}$$

where $\widehat{\gamma} = \widehat{\gamma}(X)$ and $\widehat{q}_{\delta_n} = \widehat{q}_{\delta_n}(X)$. El Methni and Stupfler (2017) and Daouia *et al.* (2018) study these estimators for independent, identically distributed (i.i.d.) data.

To derive limit theory for $\widehat{\gamma}$ and $\widehat{q}_{\delta_n}$, we impose a common second-order strengthening of (6) to control bias terms in the asymptotic approximations. El Methni and Stupfler (2017) and Daouia *et al.* (2018) also require such a second-order condition, termed $\mathcal{C}_2(\gamma, \rho, A)$, to study the asymptotics of $\widehat{\mathrm{DRM}}_{\delta_n}$ and $\widehat{\xi}_{\delta_n}$. Condition $\mathcal{C}_2(\gamma, \rho, A)$ is said to hold if there exist $\gamma > 0$, $\rho < 0$ and a function $A(\cdot)$ with $\lim_{t\to\infty} A(t) = 0$ and constant sign, such that for all $x > 0$,

$$\lim_{t\to\infty} \frac{\frac{U(tx)}{U(t)} - x^{\gamma}}{A(t)} = x^{\gamma} \frac{x^{\rho} - 1}{\rho}.$$

This is a standard assumption in EVT. Heavy-tailed distributions satisfying it are abundant (de Haan and Ferreira, 2006, p. 76). The smaller γ, the heavier the (right) tail of the distribution. The function $|A(\cdot)|$ is necessarily regularly varying with index ρ (de Haan and Ferreira, 2006, Thm. 2.3.3). Thus, the closer ρ is to $-\infty$, the faster $A(t)$ converges to 0 and—as a consequence—the more accurate the Pareto approximation $U(tx)/U(t) \approx x^{\gamma}$ in the tail.

2.3 Estimation of conditional risk measures

To estimate extreme conditional DRMs and expectiles based on a sequence of returns Y_t, we consider a standard location-scale model. As usual, let $\|\cdot\|$ denote the Euclidean norm.

Assumption 2. *The process* $\{Y_t\}_{t\in\mathbb{Z}}$ *is generated by*

$$Y_t = \mu_t(\boldsymbol{\theta}^{\circ}) + \sigma_t(\boldsymbol{\theta}^{\circ})\varepsilon_t, \qquad \varepsilon_t \overset{i.i.d.}{\sim} (0, 1), \tag{13}$$

where $\sigma_t(\boldsymbol{\theta}^{\circ}) > 0$ *almost surely and* $\boldsymbol{\theta}^{\circ} \in \mathbb{R}^p$ *a parameter vector. The innovations* ε_t *are independent of* $\mathcal{F}_{t-1} = \sigma(Y_{t-1}, Y_{t-2}, \ldots; \boldsymbol{X}_{t-1}, \boldsymbol{X}_{t-2}, \ldots)$*, the* σ*-field generated by* $Y_{t-1}, Y_{t-2}, \ldots$ *and possibly additional* $\mathbb{R}^d$*-valued covariates* $\boldsymbol{X}_{t-1}, \boldsymbol{X}_{t-2}, \ldots$*. Both* $\mu_t(\boldsymbol{\theta}^{\circ})$ *and* $\sigma_t(\boldsymbol{\theta}^{\circ})$ *are measurable with respect to* $\mathcal{F}_{t-1}$*. Furthermore,* $\mu_t(\boldsymbol{\theta}^{\circ}) \underset{(t\to\infty)}{=} \mathcal{O}_P(1)$ *and* $\sigma_t^{-1}(\boldsymbol{\theta}^{\circ}) \underset{(t\to\infty)}{=} \mathcal{O}_P(1)$*. We assume there exists a neighbourhood* $\boldsymbol{\Theta}_0$ *of* $\boldsymbol{\theta}^{\circ}$*, such that* $\mu_t(\boldsymbol{\theta})$ *and* $\sigma_t(\boldsymbol{\theta})$ *are differentiable on* $\boldsymbol{\Theta}_0$ *with derivatives that satisfy* $\mathrm{E}\left[\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}_0}\left\|\frac{\partial\mu_t}{\partial\boldsymbol{\theta}}(\boldsymbol{\theta})\right\|\right]^{\iota} < \infty$ *and* $\mathrm{E}\left[\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}_0}\left\|\frac{\partial\sigma_t}{\partial\boldsymbol{\theta}}(\boldsymbol{\theta})\right\|\right]^{\iota} < \infty$ *uniformly in* t *for some* $\iota > 0$*.*

Location scale models such as (13) can be justified from asset pricing theory (Campbell *et al.*, 1997). The location-scale model of Assumption 2 is very general, nesting (under suitable parameter restrictions) GARCH models, ARMA–GARCH models and many of their their myriad variations. Assumption 2 imposes mild smoothness conditions on the conditional mean and the conditional variance dynamics. The requirement that $\sigma_t^{-1}(\boldsymbol{\theta}^{\circ}) \underset{(t\to\infty)}{=} \mathcal{O}_P(1)$ is often trivially satisfied by a positivity constraint $\sigma_t^2 \geq \omega > 0$ almost surely.

Let us denote by $F_{t-1}(\cdot) = P\left\{Y_t \leq \cdot \mid \mathcal{F}_{t-1}\right\}$ the conditional distribution of Y_t given past information $\mathcal{F}_{t-1}$. In view of (2), we define a *conditional* extreme DRM via

$$\mathrm{DRM}_{\delta,t-1} := \mathrm{DRM}_{\delta,t-1}(Y_t) := \int_0^1 F_{t-1}^{\leftarrow}(1-(1-\delta)\alpha)\,\mathrm{d}g(\alpha), \tag{14}$$

where $g(\cdot)$ again denotes a distortion function. Following Newey and Powell (1987, p. 824), we define the *conditional* expectile as

$$\xi_{\delta,t-1} := \xi_{\delta,t-1}(Y_t) := \arg\min_q \mathrm{E}[\eta_{\delta,2}(Y_t-q) \mid \mathcal{F}_{t-1}].$$

Both $\mathrm{DRM}_{\delta,t-1}$ and $\xi_{\delta,t-1}$ are random variables and no longer scalars.

Under Assumption 2, the expressions for $\mathrm{DRM}_{\delta,t-1}$ and $\xi_{\delta,t-1}$ simplify. By $\mathcal{F}_{t-1}$-measurability of $\mu_t(\boldsymbol{\theta}^\circ)$ and $\sigma_t(\boldsymbol{\theta}^\circ)$ and the independence of ε_t from $\mathcal{F}_{t-1}$, it follows that

$$F_{t-1}^{\leftarrow}(\delta) = \mu_t(\boldsymbol{\theta}^\circ) + \sigma_t(\boldsymbol{\theta}^\circ)F^{\leftarrow}(\delta),$$

where $F(\cdot)$ denotes the d.f. of ε_t. Plugging this into (14) and recalling that $g(0)=0$, $g(1)=1$, gives

$$\begin{aligned}\mathrm{DRM}_{\delta,t-1} &= \mu_t(\boldsymbol{\theta}^\circ)\int_0^1 \mathrm{d}g(\alpha) + \sigma_t(\boldsymbol{\theta}^\circ)\int_0^1 F^{\leftarrow}(1-(1-\delta)\alpha)\,\mathrm{d}g(\alpha)\\ &= \mu_t(\boldsymbol{\theta}^\circ) + \sigma_t(\boldsymbol{\theta}^\circ)\,\mathrm{DRM}_\delta(\varepsilon). \end{aligned}\tag{15}$$

The fact that

$$\xi_{\delta,t-1} = \mu_t(\boldsymbol{\theta}^\circ) + \sigma_t(\boldsymbol{\theta}^\circ)\xi_\delta(\varepsilon) \tag{16}$$

follows from the location and scale equivariance of expectiles (Newey and Powell, 1987, Thm. 1).

Given a sample $Y_1,\dots,Y_n$ from the location-scale model (13), we now turn to estimators of $\mathrm{DRM}_{\delta_n,n}$ and $\xi_{\delta_n,n}$, which are both measures of risk inherent in Y_{n+1} given the past. The conditional mean $\mu_{n+1}(\boldsymbol{\theta}^\circ)$ and the conditional variance $\sigma_{n+1}(\boldsymbol{\theta}^\circ)$ in (15) and (16) can easily be obtained from $\mu_{n+1}(\widehat{\boldsymbol{\theta}}_n)$ and $\sigma_{n+1}(\widehat{\boldsymbol{\theta}}_n)$, where $\widehat{\boldsymbol{\theta}}_n$ is an estimator of the model parameters $\boldsymbol{\theta}^\circ$.[4] For instance, in GARCH(p,q) models, $\widehat{\boldsymbol{\theta}}_n$ may be obtained from standard Gaussian quasi-maximum likelihood estimation (QMLE). Regarding the parameter estimator, we require

Assumption 3. *The estimator $\widehat{\boldsymbol{\theta}}_n$ of $\boldsymbol{\theta}^\circ$ is such, that*

$$\frac{\sqrt{n(1-\beta_n)}}{\log([1-\beta_n]/[1-\delta_n])}(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}^\circ) = o_P(1).$$

Assumption 3 is a mild assumption that ensures the conditional mean and conditional variance can be estimated with sufficient precision. It holds in particular for any $\sqrt{n}$-consistent estimator of $\boldsymbol{\theta}^\circ$.

[4]Strictly speaking, $\mu_{n+1}(\widehat{\boldsymbol{\theta}}_n)$ and $\sigma_{n+1}(\widehat{\boldsymbol{\theta}}_n)$ can be functions of the unobserved infinite past. Conditions under which a truncation to the observed values of the finite past does not matter asymptotically are standard; see, e.g., Assumption A5 in Francq and Zakoïan (2015) and Francq and Zakoïan (2016), or Assumption A.3 in Du and Escanciano (2017). For ease of exposition, we ignore this technical issue, and refer to the simulations for evidence that this issue may be neglected for the models considered there.

For (ARMA–)GARCH models, Francq and Zakoïan (2010) provide many $\sqrt{n}$-consistent estimators under suitable parameter and moment restrictions.

Since the ε_t are unobserved, we use the standardised residuals

$$\widehat{\varepsilon}_t := \frac{Y_t - \mu_t(\widehat{\boldsymbol{\theta}}_n)}{\sigma_t(\widehat{\boldsymbol{\theta}}_n)}, \qquad t = 1, \dots, n, \tag{17}$$

to estimate $\mathrm{DRM}_{\delta_n}(\varepsilon)$ and $\xi_{\delta_n}(\varepsilon)$ in (15) and (16). To be able to apply the estimation theory of Subsection 2.2 to $\mathrm{DRM}_{\delta_n}(\varepsilon)$ and $\xi_{\delta_n}(\varepsilon)$, we impose

Assumption 4. *The tail of the innovations ε_t in Assumption 2 satisfies $\mathcal{C}_2(\gamma, \rho, A)$, where the function $A(\cdot)$ satisfies $\sqrt{n(1-\beta_n)}A([1-\beta_n]^{-1}) \underset{(n\to\infty)}{\longrightarrow} 0$.*

Many popular innovation distributions satisfy Assumption 4, e.g., the Student's t-distribution, which was popularized in that context by Bollerslev (1987). Hoga (2018+a, Remark 3 (c)) discusses some methods to check the plausibility of Assumption 4 empirically. Essentially, all methods from extreme value theory to verify a Pareto-shaped tail can be used for the residuals $\widehat{\varepsilon}_1, \dots, \widehat{\varepsilon}_n$, e.g., Hill plots or Pareto quantile plots (Beirlant *et al.*, 2004).

Estimators of $\mathrm{DRM}_{\delta_n,n}$ and $\xi_{\delta_n,n}$ can now be obtained from (15) and (16) as

$$\begin{aligned}
\widehat{\mathrm{DRM}}_{\delta_n,n} &:= \mu_{n+1}(\widehat{\boldsymbol{\theta}}_n) + \sigma_{n+1}(\widehat{\boldsymbol{\theta}}_n)\widehat{\mathrm{DRM}}_{\delta_n}(\widehat{\varepsilon}),\\
\widehat{\xi}_{\delta_n,n} &:= \mu_{n+1}(\widehat{\boldsymbol{\theta}}_n) + \sigma_{n+1}(\widehat{\boldsymbol{\theta}}_n)\widehat{\xi}_{\delta_n}(\widehat{\varepsilon}).
\end{aligned}$$

2.4 Asymptotic results

Before we can state our main result, we need one final assumption.

Assumption 5. *For a generic estimator $\widehat{\gamma}(\widehat{\varepsilon})$ of γ, we have*

$$\frac{\sqrt{n(1-\beta_n)}}{\widehat{\sigma}}(\widehat{\gamma}(\widehat{\varepsilon}) - \gamma) \underset{(n\to\infty)}{\overset{\mathcal{D}}{\longrightarrow}} Z, \qquad Z \sim \mathcal{N}(0,1) \tag{18}$$

$$\sqrt{n(1-\beta_n)}\left(\frac{\widehat{\varepsilon}_{\lceil n\beta_n\rceil,n}}{q_{\beta_n}(\varepsilon)} - 1\right) \underset{(n\to\infty)}{=} \mathcal{O}_P(1). \tag{19}$$

This high-level condition ensures that the tail shape and high, but within-sample, quantiles of the innovations ε_t can be estimated with sufficient precision based on the filtered residuals $\widehat{\varepsilon}_1, \dots, \widehat{\varepsilon}_n$ from (17). Assumption 5 holds true for GARCH models and the Hill (1975) estimator (Chan *et al.*, 2007); for ARMA–GARCH models and the Hill (1975) estimator, the moments ratio estimator (Daníelsson *et al.*, 1996) and the Csörgő and Viharos (1998) estimator (Hoga, 2018+a). Hill (2015, Theorem 1 and Remark 9) verifies Assumption 5 under some low-level conditions. Kim and Lee (2016, Section 2.2) do so for the PTTGARCH models of Pan *et al.* (2008). For non-parametric location-scale models it is checked by Martins-Filho *et al.* (2018) for a Peaks-over-Threshold (POT) estimator of γ. Overall, Assumption 5 is satisfied for a wide range of processes that are typically used to model returns on speculative assets. The respective estimators $\widehat{\sigma}$ of the asymptotic variance of $\widehat{\gamma}(\widehat{\varepsilon})$ can be obtained from the above cited references.

Theorem 1. *Suppose Assumptions 1–5 hold. Let $Z \sim \mathcal{N}(0,1)$ and denote by $q(\cdot)$ the quantile function of ε_0. Then, (i)*

$$\frac{1}{\widehat{\sigma}} \frac{\sqrt{n(1-\beta_n)}}{\log([1-\beta_n]/[1-\delta_n])} \left(\frac{\widehat{\mathrm{DRM}}_{\delta_n,n}}{\mathrm{DRM}_{\delta_n,n}} - 1 \right) \underset{(n\to\infty)}{\overset{\mathcal{D}}{\longrightarrow}} Z, \tag{20}$$

if $\int_0^1 s^{-\gamma-1/2-\eta}\,\mathrm{d}g(s) < \infty$ for some $\eta > 0$ and $q(\cdot)$ is continuous and strictly increasing in a neighbourhood of infinity, and (ii)

$$\frac{1}{\widehat{\sigma}} \frac{\sqrt{n(1-\beta_n)}}{\log([1-\beta_n]/[1-\delta_n])} \left(\frac{\widehat{\xi}_{\delta_n,n}}{\xi_{\delta_n,n}} - 1 \right) \underset{(n\to\infty)}{\overset{\mathcal{D}}{\longrightarrow}} Z, \tag{21}$$

if $\gamma \in (0,1)$, $\mathrm{E}\,|\varepsilon_0| < \infty$ and $q(\cdot)$ is strictly increasing.

The equivalent of (20) for unconditional DRMs can be found in El Methni and Stupfler (2017, Thm. 3), who show that (20) holds with $\widehat{\mathrm{DRM}}_{\delta_n,n}/\,\mathrm{DRM}_{\delta_n,n}$ replaced by $\widehat{\mathrm{DRM}}_{\delta_n}(\varepsilon)/\,\mathrm{DRM}_{\delta_n}(\varepsilon)$. Thus, the filtering (via the estimate $\widehat{\boldsymbol{\theta}}_n$) required for conditional DRMs does not introduce additional estimation effects. These disappear due to Assumption 3, which ensures that $\widehat{\boldsymbol{\theta}}_n$ converges faster than DRM estimates. A similar result holds true for (21), for which the unconditional analogue is Corollary 3 of Daouia *et al.* (2018). If $\widehat{\boldsymbol{\theta}}_n$ converges at the same rate as the risk measure estimates based on the residuals, estimation effects would appear in the limiting distribution. Gao and Song (2008) and Wang and Zhao (2016) demonstrate this for $\sqrt{n}$-consistent parameter and (non-parametric) VaR/ES estimators.

For $z \in \{\mathrm{DRM}, \xi\}$, we obtain the following asymptotic $(1-\alpha)$-confidence interval for $z_{\delta_n,n}$ from Theorem 1:

$$I_{1-\alpha} = \widehat{z}_{\delta_n,n} \exp\left\{ \mp\Phi(1-\alpha/2)\widehat{\sigma} \frac{\log([1-\beta_n]/[1-\delta_n])}{\sqrt{n(1-\beta_n)}} \right\}. \tag{22}$$

Note that since $\log(x) \sim x-1$ as $x \to 1$, $\widehat{z}_{\delta_n,n}/z_{\delta_n,n} - 1$ can be replaced by $\log(\widehat{z}_{\delta_n,n}/z_{\delta_n,n})$ in Theorem 1. Drees (2003) and Gomes and Pestana (2007) show in simulations that $\log\left(\widehat{q}_{\delta_n}(\varepsilon)/q_{\delta_n}(\varepsilon)\right)$ is better approximated by the limiting normal distribution than $\widehat{q}_{\delta_n}(\varepsilon)/q_{\delta_n}(\varepsilon)-1$. Thus, since $\widehat{z}_{\delta_n}(\widehat{\varepsilon})$ is a simple function of $\widehat{q}_{\delta_n}(\widehat{\varepsilon})$, we use the log-formulation of Theorem 1 to construct confidence intervals in (22).

As argued in the Motivation, it is desirable to derive some uniform limit theory to obtain a more complete picture of tail risk. To do so, we write $\delta_n = 1-p_n$ for p_n chosen in the obvious way. Then, we define $\delta_n(t) = 1-p_n t$ for $t > 0$. Our aim is to derive limit theory for $\widehat{\mathrm{DRM}}_{\delta_n(t),n}$ and $\widehat{\xi}_{\delta_n(t),n}$ that is uniform in the tail, i.e., uniform in $t \in [\underline{t}, \overline{t}]$. Wang and Zhao (2016) also derive some uniform limit theory. Yet, their results are only valid for VaR and for $\delta_n = 1-p$ uniformly in $p \in [\varepsilon,\ 1-\varepsilon]$ $(\varepsilon > 0)$, so that the tail is explicitly excluded.

Theorem 2. *For $0 < \underline{t} < \overline{t} < \infty$, we have that under the conditions of Theorem 1 (i)*

$$\sup_{t\in[\underline{t},\overline{t}]}\left|\frac{1}{\widehat{\sigma}}\frac{\sqrt{n(1-\beta_n)}}{\log([1-\beta_n]/[1-\delta_n(t)])}\left(\frac{\widehat{\mathrm{DRM}}_{\delta_n(t),n}}{\mathrm{DRM}_{\delta_n(t),n}}-1\right)\right|\xrightarrow[(n\to\infty)]{\mathcal{D}}|Z|,$$

and under the conditions of Theorem 1 (ii)

$$\sup_{t\in[\underline{t},\overline{t}]}\left|\frac{1}{\widehat{\sigma}}\frac{\sqrt{n(1-\beta_n)}}{\log([1-\beta_n]/[1-\delta_n(t)])}\left(\frac{\widehat{\xi}_{\delta_n(t),n}}{\xi_{\delta_n(t),n}}-1\right)\right|\xrightarrow[(n\to\infty)]{\mathcal{D}}|Z|.$$

Again, we use the log-formulation of Theorem 2 to obtain the asymptotic $(1-\alpha)$-confidence corridor

$$\widehat{z}_{\delta_n(t),n}\exp\left\{-\Phi(1-\alpha/2)\widehat{\sigma}\frac{\log([1-\beta_n]/[1-\delta_n(t)])}{\sqrt{n(1-\beta_n)}}\right\}\leq z_{\delta_n(t),n}$$
$$\leq \widehat{z}_{\delta_n(t),n}\exp\left\{+\Phi(1-\alpha/2)\widehat{\sigma}\frac{\log([1-\beta_n]/[1-\delta_n(t)])}{\sqrt{n(1-\beta_n)}}\right\},\qquad t\in[\underline{t},\overline{t}],$$

for $z_{\delta_n(t),n}$ ($z\in\{\mathrm{DRM},\xi\}$). As in the case of (unconditional) VaR, dealt with by Hoga (2018+c, Sec. 2.3), the confidence corridor simply consists of the pointwise confidence intervals from Theorem 1. Thus the width of the confidence corridor does not depend on $\underline{t}$ and $\overline{t}$, as might have been expected.

3 Monte Carlo simulations

We investigate coverage of the asymptotic confidence intervals (22) for $\mathrm{DRM}_{\delta_n,n}$ and $\xi_{\delta_n,n}$ suggested by Theorem 1. For conciseness, we use $g(x)=I_{\{x=1\}}$ for $\mathrm{DRM}_{\delta_n,n}$, so that it corresponds to conditional VaR, denoted by $q_{\delta_n,n}$. Throughout, we use Hill (1975) estimates $\widehat{\gamma}(\widehat{\varepsilon})$ of γ.

We let $\{Y_t\}_{t=-v+1,\ldots,n}$ follow a popular GARCH$(1,1)$ model

$$Y_t=\sigma_t\varepsilon_t,\qquad \text{where}\quad \sigma_t^2=\omega^\circ+\alpha^\circ Y_{t-1}^2+\beta^\circ\sigma_{t-1}^2$$

with $(\omega^\circ,\alpha^\circ,\beta^\circ)=(0.00001,0.1,0.85)$. Estimates $(\widehat{\omega}^\circ,\widehat{\alpha}^\circ,\widehat{\beta}^\circ)$ of the parameters are obtained from standard Gaussian QMLE if $\mathrm{E}[\varepsilon_t^4]<\infty$, and Laplace QMLE if $\mathrm{E}[\varepsilon_t^4]=\infty$. The standardized residuals are calculated via $\widehat{\varepsilon}_t=Y_t/\widehat{\sigma}_t$, where $\widehat{\sigma}_t^2=\widehat{\omega}^\circ+\widehat{\alpha}^\circ Y_{t-1}^2+\widehat{\beta}^\circ\sigma_{t-1}^2$. Due to initialization effects in the variance equation, we discard the first $v=10$ residuals, giving us $\widehat{\varepsilon}_1,\ldots,\widehat{\varepsilon}_n$ to apply the theory of Section 2. We use trajectories of length $n\in\{500,\ 1000,\ 2000\}$ and let the risk level δ_n vary with n to be consistent with Assumption 1. Specifically, we choose $1-\delta_n\in\{10\%,\ 5\%,\ 1\%,\ 0.5\%\}$ for $n=500$, $1-\delta_n\in\{5\%,\ 1\%,\ 0.5\%,\ 0.1\%\}$ for $n=1000$ and $1-\delta_n\in\{1\%,\ 0.5\%,\ 0.1\%,\ 0.05\%\}$ for $n=2000$.

For the distribution of ε_t, consider a Burr(β,λ,η) distribution of type XII with d.f.

$$F(x)=1-\left(\frac{\beta}{\beta+x^\tau}\right)^\lambda,\qquad x>0,\quad \beta,\ \tau,\ \lambda>0.$$

This distribution has extreme value index $\gamma = 1/(\tau\lambda)$ and second-order parameter $\rho = -1/\lambda$ in $\mathcal{C}_2(\gamma, \rho, A)$. We use this distribution because it allows to vary γ separately from ρ. Recall from the discussion of condition $\mathcal{C}_2(\gamma, \rho, A)$ that a smaller ρ aligns with a better fit to a Pareto tail. To obtain zero-mean innovations with unit variance, we choose $\varepsilon_t = R_t B_t / \sqrt{\mathrm{E}[B_t^2]}$, where R_t are i.i.d. Rademacher random variables (i.e., equal ± 1 with probability 1/2), independent of the $B_t \overset{\text{i.i.d.}}{\sim} \mathrm{Burr}(\beta, \lambda, \eta)$. As parameters for the Burr distribution, we take $(\beta, \lambda, \eta) \in \{(1, 0.25, 12),\ (1, 1, 3)\}$ to obtain models with $\gamma = 1/3$, and $(\beta, \lambda, \eta) \in \{(1, 0.25, 20),\ (1, 1, 5)\}$ to obtain models with $\gamma = 1/5$. We abbreviate the corresponding four GARCH(1, 1) models by their respective values of (γ, ρ), i.e., $(1/3, -4)$, $(1/3, -1)$, $(1/5, -4)$, $(1/5, -1)$. For $\gamma = 1/3$ ($\gamma = 1/5$), the innovations have infinite (finite) fourth moments, so parameters are estimated via Laplace (Gaussian) QMLE.

Regarding the choice of β_n in $\widehat{q}_{\delta_n,n}$ and $\widehat{\xi}_{\delta_n,n}$, we opt for a proposal of Daníelsson *et al.* (2016). We follow Hoga (2018+a) in its implementation. Let $1 \leq k_{\min} < k_{\max}$ denote the minimal and maximal number of upper order statistics to use in extreme value index estimation of the standardized residuals $\widehat{\varepsilon}_t$. The $(1 - j/n)$-quantile of ε_0 can be estimated either non-parametrically via $\widehat{\varepsilon}_{n-j,n}$, or using the (semi-parametric) Pareto-based estimate in (8) (setting $\delta_n = 1 - j/n$), i.e.,

$$\widehat{q}_{1-j/n,\beta_n} := \widehat{q}_{1-j/n}(\widehat{\varepsilon}) = \left(\frac{1-\beta_n}{j/n}\right)^{\widehat{\gamma}(\widehat{\varepsilon})} \widehat{\varepsilon}_{\lceil n\beta_n \rceil, n}.$$

This approximation should be good in the sense that the absolute distance between the Pareto-motivated and the non-parametric estimate, $|\widehat{q}_{1-j/n,\beta_n} - \widehat{\varepsilon}_{n-j,n}|$, is small for all $j = 1, \ldots, k_{\max}$. Thus, we choose $\beta_n = \beta^*$ with

$$\beta^* := \underset{\beta_n \in \{1-k_{\min}/n, \ldots, 1-k_{\max}/n\}}{\arg\min} \left[\sup_{j=1,\ldots,k_{\max}} |\widehat{q}_{1-j/n,\beta_n} - \widehat{\varepsilon}_{n-j,n}| \right]. \tag{23}$$

Chan *et al.* (2007) use $k_n = \lfloor 1.5(\log n)^2 \rfloor$ order statistics for the Hill estimator. This value motivates the choice of $k_{\min} = \lfloor (\log n)^2 \rfloor$ and $k_{\max} = \lfloor 4(\log n)^2 \rfloor$ in our data-driven approach in (23). Thus, for (e.g.) $n = 1000$ we use between $k_{\min}/n \approx 5\%$ and $k_{\max}/n \approx 20\%$ of the largest residuals, s.t. $80\% \lessapprox \beta^* \lessapprox 95\%$. Thus, we estimate the $\delta_n =$ 95%-, 99%-, 99.5%-, and 99.9%-quantiles based on a non-parametric estimate of the β^*-quantile with $80\% \lessapprox \beta^* \lessapprox 95\%$ combined with some extrapolation to the desired extreme level via $[(1-\beta^*)/(1-\delta_n)]^{\widehat{\gamma}(\widehat{\varepsilon})}$. So any advantage in quantile estimation of the extreme value vis-à-vis the non-parametric approach comes from this extrapolation step.

Since $\beta_n = \beta^*$ is less extreme than δ_n, Assumption 1 is met. Assumptions 2 and 3 are also satisfied, since Laplace and Gaussian QMLE are $\sqrt{n}$-consistent (Francq and Zakoïan, 2010). As required by Assumption 4, the Burr distribution satisfies $\mathcal{C}_2(\gamma, \rho, A)$. Finally, Assumption 5 is satisfied for $\widehat{\sigma} = \widehat{\gamma}(\widehat{\varepsilon})$ by results of Chan *et al.* (2007), Hoga (2018+a), or Hill (2015, Example 3.3). Here, $\widehat{\sigma}$ estimates the asymptotic variance of the Hill estimator $\widehat{\gamma}(\varepsilon)$ for the i.i.d. innovations (de Haan and Ferreira, 2006, Thm. 3.2.5), but also the asymptotic variance of $\widehat{\gamma}(\widehat{\varepsilon})$ for the (non-i.i.d.) filtered residuals $\widehat{\varepsilon}_t$. To explicitly account for possible remaining serial dependence in the $\widehat{\varepsilon}_1, \ldots, \widehat{\varepsilon}_n$, instead of $\widehat{\sigma} = \widehat{\gamma}(\widehat{\varepsilon})$ we

(γ,ρ)	$1-\beta^*$	Risk measure	$1-\delta_n$	Bias	RMSE	Coverage	Int. length
$(1/3,-4)$	15%	CVaR	10%	0.24	0.57	94.4	2.12
			5%	0.23	0.65	93.9	2.72
			1%	0.13	1.51	84.9	4.69
			0.5%	0.02	2.48	78.0	5.84
		CExp	10%	0.44	0.60	80.2	1.67
			5%	0.32	0.65	88.5	2.15
			1%	0.14	1.48	79.2	3.73
			0.5%	0.04	2.40	71.2	4.67
$(1/3,-1)$	11%	CVaR	10%	0.30	0.71	97.4	2.95
			5%	0.21	0.88	97.5	3.84
			1%	0.43	2.07	92.4	6.90
			0.5%	0.79	3.99	88.0	8.99
		CExp	10%	0.58	0.84	89.3	2.40
			5%	0.45	0.99	93.2	3.13
			1%	0.69	2.34	86.1	5.64
			0.5%	1.09	4.25	81.6	7.38
$(1/5,-4)$	15%	CVaR	10%	0.33	0.46	86.6	1.49
			5%	0.34	0.58	87.5	1.77
			1%	0.32	0.84	82.7	2.47
			0.5%	0.31	1.12	77.5	2.79
		CExp	10%	0.48	0.54	52.8	1.13
			5%	0.38	0.53	79.1	1.35
			1%	0.30	0.65	83.1	1.88
			0.5%	0.29	0.84	79.4	2.13
$(1/5,-1)$	11%	CVaR	10%	0.37	0.61	94.2	2.15
			5%	0.33	0.66	95.3	2.57
			1%	0.48	1.11	90.3	3.66
			0.5%	0.65	1.56	86.8	4.28
		CExp	10%	0.68	0.78	62.7	1.63
			5%	0.47	0.64	89.8	1.95
			1%	0.43	0.88	89.6	2.78
			0.5%	0.54	1.22	86.7	3.25

Table 1: Values β^*, bias, RMSE, coverage probabilities and interval lengths of asymptotic 95%-confidence intervals averaged over 10,000 repetitions for $n=500$.

use Hill's (2010) estimator

$$\widehat{\sigma}_H^2 := \frac{1}{\lceil n(1-\beta_n)\rceil}\sum_{i,j=1}^{n} w\left(\frac{s-t}{\gamma_n}\right)\left[\log_+\left(\frac{\widehat{\varepsilon}_i}{\widehat{\varepsilon}_{\lceil n\beta_n\rceil,n}}\right)-(1-\beta_n)\widehat{\gamma}\right]\left[\log_+\left(\frac{\widehat{\varepsilon}_j}{\widehat{\varepsilon}_{\lceil n\beta_n\rceil,n}}\right)-(1-\beta_n)\widehat{\gamma}\right],$$

where $\log_+(x) := \log(\max\{x,1\})$, $w(\cdot)$ is the Bartlett kernel, and $\gamma_n\to\infty$ the bandwidth with $\gamma_n=o(n)$, and $\sqrt{n}(1-\beta_n)\to\infty$. The estimator $\widehat{\sigma}_H^2$ is consistent for the asymptotic variance of the

(γ,ρ)	$1-\beta^*$	Risk measure	$1-\delta_n$	Bias	RMSE	Coverage	Int. length
$(1/3,-4)$	9.4%	CVaR	5%	0.27	0.70	97.4	3.09
			1%	0.19	1.26	94.5	5.26
			0.5%	0.13	1.94	90.3	6.59
			0.01%	−0.12	5.13	77.2	11.3
		CExp	5%	0.35	0.66	94.1	2.45
			1%	0.17	1.22	90.8	4.18
			0.5%	0.11	1.90	85.8	5.24
			0.1%	−0.08	4.74	70.7	9.08
$(1/3,-1)$	7.6%	CVaR	5%	0.32	0.90	97.9	3.93
			1%	0.40	2.10	95.6	6.97
			0.5%	0.61	2.75	92.0	8.85
			0.1%	1.60	6.74	81.4	15.6
		CExp	5%	0.48	0.89	94.1	3.16
			1%	0.52	2.02	91.1	5.62
			0.5%	0.75	2.84	86.8	7.16
			0.1%	1.77	6.79	76.0	12.7
$(1/5,-4)$	9.4%	CVaR	5%	0.35	0.52	93.3	2.00
			1%	0.37	0.72	90.8	2.76
			0.5%	0.37	0.96	88.7	3.18
			0.1%	0.36	1.79	78.1	4.36
		CExp	5%	0.39	0.49	86.1	1.52
			1%	0.33	0.56	90.4	2.10
			0.5%	0.33	0.73	89.2	2.42
			0.1%	0.31	1.33	79.9	3.31
$(1/5,-1)$	7.5%	CVaR	5%	0.40	0.62	96.1	2.60
			1%	0.44	0.94	93.2	3.67
			0.5%	0.53	1.24	89.9	4.22
			0.1%	0.85	2.28	81.7	5.85
		CExp	5%	0.52	0.64	89.9	1.98
			1%	0.41	0.75	92.4	2.79
			0.5%	0.45	0.95	89.8	3.20
			0.1%	0.67	1.73	82.4	4.44

Table 2: Same as Table 1 for $n = 1000$.

Hill (1975) estimator under very weak conditions on the temporal dependence and is also applicable for filtered residuals (Hill, 2015, Rem. 8). Following Hill (2010), we use the bandwidth $\gamma_n = [n(1-\beta^*)]^{0.25}$.

Tables 1–3 display the simulation results for both conditional VaR (CVaR) and conditional Expectiles (CExp). We draw the following conclusions:

1. The better the Pareto-approximation (as indicated by a smaller value of ρ), the more observations are used for estimation. For both values of $\gamma \in \{1/3, 1/5\}$ and (e.g.) $n = 1000$, 9.4% of the largest residuals are used for estimation when $\rho = -4$, whereas for $\rho = -1$ only 7.5% are

(γ, ρ)	$1-\beta^*$	Risk measure	$1-\delta_n$	Bias	RMSE	Coverage	Int. length
$(1/3, -4)$	5.7%	CVaR	1%	0.30	1.09	97.9	5.92
			0.5%	0.26	1.52	96.8	7.34
			0.1%	0.04	4.02	89.2	12.5
			0.05%	−0.06	6.02	84.0	15.9
		CExp	1%	0.26	1.04	96.6	4.70
			0.5%	0.20	1.48	94.1	5.83
			0.1%	0.02	3.82	84.3	10.0
			0.05%	−0.05	5.58	78.5	12.7
$(1/3, -1)$	5.0%	CVaR	1%	0.44	1.51	97.7	7.17
			0.5%	0.53	2.59	96.2	9.16
			0.1%	1.07	5.12	89.3	15.7
			0.05%	1.47	7.55	85.6	20.0
		CExp	1%	0.48	1.44	94.9	5.75
			0.5%	0.57	2.64	92.8	7.36
			0.1%	1.12	5.05	84.4	12.6
			0.05%	1.50	7.14	80.3	16.1
$(1/5, -4)$	5.7%	CVaR	1%	0.41	0.64	95.3	3.03
			0.5%	0.41	0.80	94.6	3.50
			0.1%	0.42	1.45	88.5	4.81
			0.05%	0.43	1.90	84.5	5.51
		CExp	1%	0.36	0.51	94.7	2.31
			0.5%	0.35	0.62	94.5	2.66
			0.1%	0.36	1.08	89.5	3.66
			0.05%	0.37	1.42	85.8	4.19
$(1/5, -1)$	5.0%	CVaR	1%	0.46	0.80	95.8	3.74
			0.5%	0.50	1.01	93.7	4.28
			0.1%	0.71	1.89	86.9	5.92
			0.05%	0.86	2.43	83.6	6.85
		CExp	1%	0.42	0.64	95.0	2.84
			0.5%	0.43	0.78	93.4	3.25
			0.1%	0.56	1.42	87.4	4.49
			0.05%	0.68	1.83	84.3	5.20

Table 3: Same as Table 1 for $n = 2000$.

used. Thus, the data-driven method of choosing β^* picks up the different qualities of the Pareto approximations.

2. Bias and, even more so, root mean square error (RMSE) tend to decrease the less extreme the probability level $1-\delta_n$, and also the smaller the value of ρ (i.e., the better the approximation to the Pareto tail). Although bias is comparable for CVaR and CExp estimation, CExp estimates tend to be more precise in terms of RMSE. This is as expected, because for all $\gamma \in (0, 1/2)$ the limit in (11) is smaller than 1 and thus VaR is more extreme than the expectile at the same

level. Put differently, VaR at level δ_n is equal to an expectile at a more extreme level $\delta_n' > \delta_n$ (Kuan *et al.*, 2009, p. 263).

3. The more extrapolation is required (i.e., the smaller $1 - \delta_n$), the worse coverage tends to be. This may be explained as follows. For all values of $1 - \delta_n$, the limiting distribution in Theorem 1 is the same. Thus, the different degree of required extrapolation (as measured by $[(1 - \beta^*)/(1 - \delta_n)]^{\widehat{\gamma}(\widehat{\varepsilon})}$)—which is of course again subject to estimation uncertainty via $\widehat{\gamma}(\widehat{\varepsilon})$—may not be sufficiently reflected in the confidence interval (22) in finite samples. Hoga (2018+a) shows that these distortions can be alleviated using self-normalized confidence intervals. These, however, require functional central limit theory to hold in Assumption 5, which is more difficult to derive.

4. The fact that RMSE is lower for CExp than for CVaR forecasts is also reflected in the narrower confidence intervals, suggesting less estimation uncertainty. Again, the more accurate the Pareto approximation, the lower the estimation uncertainty for CVaR and CExp, as measured by the narrower confidence intervals.

5. As n gets larger, we observe the following: For a fixed value of $1-\delta_n$, say 1%, bias is roughly constant as the sample size increases, while—as expected—the RMSE decreases. Not surprisingly, coverage also improves the larger the sample size, even as $1 - \delta_n$ decreases. What is surprising, however, is that interval lengths for a fixed $1-\delta_n$ increase the larger the sample. For instance, for $(\gamma, \rho) = (1/3, -4)$ the 95%-confidence interval for 1%-CExp has average length 3.73 for $n = 500$, while for $n = 2000$—where coverage is quite accurate—the average length is 4.70.

In general, Tables 1–3 reveal that the degree to which asymptotic distributions are accurate depends not only on n but also on $1 - \delta_n$. This is in line with the asymptotic results of Theorem 1.

4 Application

Numerous CVaR and CES forecast comparisons can be found in the literature (see Rocco, 2014, and references therein). So we compare the CVaR and CExp forecasting performance in this application. We include CVaR since CExp estimators closely rely on CVaR estimates, so it is of interest to compare the two. Specifically, we forecast one-day-ahead CVaR and CExp for returns $\{Y_t\}_{t=1,\dots,N}$ on a wide range of speculative assets—stocks (Siemens), stock indices (CAC40), CBOE Volatility Index (VIX), commodities (WTI oil spot price) and foreign exchange (USD/EUR exchange rate). We consider 20 years of daily log-returns from 1998 to 2017 except for the exchange rate data, where data availability is restricted to the period from 1999 onwards, since the Euro was introduced in 1999. For the Euro this gives us $N = 4772$ returns and roughly $N = 5000$ for the other assets. The oil price and the exchange rate data is taken from *fred.stlouisfed.org* (series DCOILWTICO and DEXUSEU) and the other three series are downloaded from *finance.yahoo.com* (ticker symbols SIE.DE, FCHI and VIX).

As a location-scale model we choose a benchmark GARCH$(1,1)$ specification, which we estimate via Gaussian QMLE.[5] We proceed by estimating the returns based on a rolling window $\{Y_{j+t}\}_{t=-v+1,\ldots n}$ ($j = v,\ldots,N-n$) of length $n+v$. Similarly as in the Monte Carlo study, we only use the last n standardized residuals $\{\widehat{\varepsilon}_{j+t}\}_{t=1,\ldots,n}$ in $\widehat{\mathrm{DRM}}_{\delta_n}(\widehat{\varepsilon})$ and $\widehat{\xi}_{\delta_n}(\widehat{\varepsilon})$. Again, we choose $v = 10$, $n \in \{500, 1000, 2000\}$ and we let $1-\delta_n \in \{10\%,\ 5\%,\ 1\%,\ 0.5\%\}$ for $n = 500$, $1-\delta_n \in \{5\%,\ 1\%,\ 0.5\%,\ 0.1\%\}$ for $n = 1000$ and $1-\delta_n \in \{1\%,\ 0.5\%,\ 0.1\%\, 0.05\%\}$ for $n = 2000$.

Based on the residuals, we compare three different approaches to estimate $\mathrm{DRM}_{\delta_n}(\varepsilon) = q_{\delta_n}(\varepsilon)$ and $\xi_{\delta_n}(\varepsilon)$ in (15) and (16), respectively. First, we use $\widehat{q}_{\delta_n}(\widehat{\varepsilon})$ and $\widehat{\xi}_{\delta_n}(\widehat{\varepsilon})$ based on the Hill (1975) estimator, precisely as in the simulations.

As a second estimator, we apply POT methodology to the (almost i.i.d.) standardized residuals. Very briefly, POT fits—via maximum likelihood—a generalized Pareto distribution (GPD) with d.f. $G_{\gamma,\sigma}(x) = 1-(1+x/[\gamma\sigma])^{-\gamma}$ ($\sigma > 0, \gamma \in \mathbb{R}$) to the excesses above some high threshold $\widehat{\varepsilon}_{\lceil n\beta_n\rceil,n}$. The ML estimates $(\widehat{\gamma}^{\mathrm{ML}}, \widehat{\sigma}^{\mathrm{ML}})$ of (γ,σ) can then be used to estimate a more extreme VaR $q_{\delta_n}(\varepsilon)$ with $\delta_n > \beta_n$ via

$$\widehat{q}_{\delta_n}^{\mathrm{POT}} = \widehat{\varepsilon}_{\lceil n\beta_n\rceil,n} + \widehat{\sigma}^{\mathrm{ML}}\widehat{\gamma}^{\mathrm{ML}}\left[\left(\frac{1-\delta_n}{1-\beta_n}\right)^{-1/\widehat{\gamma}^{\mathrm{ML}}} - 1\right].$$

For more detail we refer to McNeil and Frey (2000, Sec. 2.2). Martins-Filho *et al.* (2018) provide asymptotic theory for POT-based VaR estimates in non-parametric location-scale models. As an expectile estimator we then consider $\widehat{\xi}_{\delta_n}^{\mathrm{POT}}(\widehat{\varepsilon}) = \left(\frac{1}{\widehat{\gamma}^{\mathrm{ML}}} - 1\right)^{-\widehat{\gamma}^{\mathrm{ML}}} \widehat{q}_{\delta_n}^{\mathrm{POT}}$, similarly as in (12). Following McNeil and Frey (2000), Mancini and Trojani (2011) and Chavez-Demoulin *et al.* (2014), we choose $\beta_n = 0.9$ so that the largest 10% of the residuals is used for estimation.

To assess the potential benefits of both semi-parametric extreme value approaches, we consider a third set of completely non-parametric estimators of $q_{\delta_n}(\varepsilon)$ and $\xi_{\delta_n}(\varepsilon)$. We define a nonparametric quantile estimator via the empirical counterpart of (3), i.e.,

$$\widehat{q}_{\delta_n}^{\mathrm{NP}} = \arg\min_{q\in\mathbb{R}} \frac{1}{n}\sum_{t=1}^{n} \eta_{\delta_n,1}(\widehat{\varepsilon}_{j+t} - q).$$

An expectile estimator can be defined similarly via the empirical counterpart of (4), i.e.,

$$\widehat{\xi}_{\delta_n}^{\mathrm{NP}} = \arg\min_{q\in\mathbb{R}} \frac{1}{n}\sum_{t=1}^{n} \eta_{\delta_n,2}(\widehat{\varepsilon}_{j+t} - q).$$

We evaluate the three different CVaR and CExp forecasts via their average scores $\overline{s}_m^{(k)}$, $k = 1,2,3$,

[5] We have also tried the GJR–GARCH(1,1) of Glosten *et al.* (1993), which was recommended in an extreme value context by Trapin (2018). While this model led to somewhat lower scores than those reported in Tables 4–6 for the GARCH$(1,1)$, the relative differences between forecasts were unaffected.

Data	Est.	CVaR				CExp			
		$1-\delta_n$				$1-\delta_n$			
		10%	5%	1%	0.5%	10%	5%	1%	0.5%
SIE.DE	Hill	0.999	0.999	0.976	0.943	1.006	1.004	0.985	0.785***
	POT	1.000	0.999	0.968	0.941	1.035***	1.058***	1.054	0.861***
VIX	Hill	1.002	0.999	0.989	0.968	1.019***	1.009**	1.048***	1.100***
	POT	1.000	1.000	0.990	0.973*	1.041***	1.076***	1.089***	1.068**
CAC40	Hill	1.002	1.000	0.970***	0.961	1.009	0.996	0.988	0.983
	POT	1.000	0.999	0.980**	0.965*	1.078***	1.101***	1.070*	1.044
USD/EUR	Hill	1.003	0.997	1.003	1.009	1.028***	1.007**	1.015	1.024
	POT	1.000	0.997*	1.004	0.997	1.113***	1.160***	1.182***	1.146***
WTI	Hill	1.000	1.000	0.993	0.995	1.018**	1.002	1.010	1.030
	POT	1.000	0.999	0.991	0.981	1.091***	1.122***	1.149***	1.125**

Table 4: Score ratios $\overline{s}_m^{(k)}/\overline{s}_m^{(3)}$ for $k=1,2$ competing forecasts based on $n=500$ observations. The score $\overline{s}_m^{(1)}$ is based on $\widehat{q}_{\delta_n}$ / $\widehat{\xi}_{\delta_n}$ with the Hill (1975) estimator (Hill), $\overline{s}_m^{(2)}$ is based on $\widehat{q}_{\delta_n}^{\mathrm{POT}}$ / $\widehat{\xi}_{\delta_n}^{\mathrm{POT}}$ (POT) and $\overline{s}_m^{(3)}$ on $\widehat{q}_{\delta_n}^{\mathrm{NP}}$ / $\widehat{\xi}_{\delta_n}^{\mathrm{NP}}$ (NP). Significantly different performance of Hill/POT vis-à-vis NP at the 10%/5%/1%-level is indicated by a */**/*** above the score ratio.

in (5) with $m=1$ ($m=2$) for the CVaR (CExp) forecasts. The different forecasts are abbreviated by Hill ($k=1$), POT ($k=2$) and NP ($k=3$). Tables 4–6 display the score ratios $\overline{s}_m^{(k)}/\overline{s}_m^{(3)}$ for $k=1$ (Hill) and $k=2$ (POT) for $n\in\{500,1000,2000\}$. Recall that lower average scores are preferable, so that score ratios below (above) 1 indicate EVT-based estimates are better (worse) than non-parametric ones. To test whether the average scores of Hill/POT are significantly different from NP, we use a standard Diebold and Mariano (1995) test. Significantly different performance at the 10%/5%/1%-level is indicated by a */**/*** above the score ratio.

We draw the following conclusions from Tables 4–6:

1. With the exception of Siemens shares, the score ratios for CVaR forecasts tend to be lower than those for CExp forecasts. Indeed, while for CVaR estimation a sample size of $n=500$ already suffices for extreme value methods to be preferable to non-parametric estimates (Table 4), for CExp forecasting the advantages only clearly emerge for $n=2000$ (Table 6).

   This may be explained as follows. Both semi-parametric expectile estimators are motivated by the approximation of expectiles as a constant multiple of VaR given in (11). Thus, this additional approximation may hurt the CExp forecasting precision relative to that of CVaR. Of course, if the approximation is very exact, it offers very precise estimates. This may be the case for Siemens returns.

2. The EVT-based estimates have a strong tendency to improve relative to NP the larger the sample size n; see, for instance, the USD/EUR returns. This may reflect the fact that one needs

Data	Est.	CVaR				CExp			
		$1-\delta_n$				$1-\delta_n$			
		5%	1%	0.5%	0.1%	5%	1%	0.5%	0.1%
SIE.DE	Hill	1.000	1.004	1.002	0.965	1.008*	0.986	0.964**	0.758***
	POT	1.000	1.010	1.009	1.012	1.034***	1.026**	0.978	0.676***
VIX	Hill	0.999	0.996	0.988	0.965	1.006	1.004	1.006	1.006
	POT	0.999	0.994	0.995	0.997	1.051***	1.063**	1.050	0.940
CAC40	Hill	0.999	0.996	0.971*	0.906	1.000	0.995	0.985	0.975
	POT	1.000	0.992	0.977	0.924	1.101***	1.071	1.059	1.002
USD/EUR	Hill	1.000	0.994	0.997	0.937	1.010	1.001	1.007	1.035
	POT	0.999	0.995	0.988	0.937	1.149***	1.181***	1.156**	1.049
WTI	Hill	0.999	1.001	1.010	0.893	1.003	1.001	1.011	0.992
	POT	1.001**	1.005	1.003	0.946	1.115***	1.178***	1.178***	1.038

Table 5: Same as Table 4 with forecasts based on $n = 1000$.

sufficiently precise γ-estimates calculated on a reasonably large sample size for extrapolation to work well and, hence, provide an advantage over non-parametric methods.

3. For CVaR forecasts at the least extreme levels (10% for $n = 500$, 5% for $n = 1000$ and 1% for $n = 2000$) there is little difference between the semi-parametric and the non-parametric estimates. Recall that any advantage in quantile estimation via EVT comes from extrapolation. Yet, for the least extreme levels there is little scope for extrapolation, so that performance is quite similar. For more extreme levels, large score differences in favour of EVT can emerge. Yet, few of these are significant, which may simply reflect that statistical inference for tail quantities requires more data. However, the consistency with which Hill and POT beat NP for extreme risk and large sample sizes suggests some merit in EVT methods even in the absence of statistical significance.

4. The performance of Hill and POT is quite similar for CVaR forecasts. However, POT-based estimates appear slightly worse for CExp forecasting, particularly for $n \in \{500, 1000\}$.

Overall, the Hill method appears slightly preferable over POT, since it performs better for CExp. The degree to which EVT methods are superior to non-parametric ones depends on the sample size and the scope for extrapolation based on the semi-parametric Pareto tail assumption. Another advantage of the Hill/POT approach vis-à-vis NP is that—using the theory developed in this paper—estimation risk can be quantified for a wide range of risk measure estimates. For instance, we are not aware of any attempts to quantify estimation uncertainty of the non-parametric expectile estimator $\widehat{\xi}_{\delta_n}^{\mathrm{NP}}$ in some location-scale model. Having a measure of uncertainty is of course crucial, because as Dowd and Blake (2006, p. 221) point out 'any estimated risk measure reported on its own is close to meaningless without some indicator of how precise the estimate might be.'

Data	Est.	CVaR				CExp			
		$1-\delta_n$				$1-\delta_n$			
		1%	0.5%	0.1%	0.05%	1%	0.5%	0.1%	0.05%
SIE.DE	Hill	0.987*	0.986	0.944	0.936	1.003	1.007	1.025	0.933
	POT	0.986**	0.986	0.932	1.021	1.025	1.021	0.941	0.795***
VIX	Hill	1.003	0.995	0.876	0.804	0.998	0.995	0.953	0.886
	POT	0.997	0.989	0.938	0.881	1.017	1.001	0.951	0.864
CAC40	Hill	1.005	0.987	0.912	0.771	0.995	0.995	0.948	0.897
	POT	0.999	0.983	0.959	0.916	1.074	1.060	0.958	0.893
USD/EUR	Hill	1.000	1.007	0.938	0.815	0.998	0.999	0.972	0.924
	POT	0.998	0.994	0.957	0.893	1.154**	1.114	0.992	0.923
WTI	Hill	1.003	0.990	0.931	0.814	0.997	0.999	0.951	0.905
	POT	1.006	0.991	0.973	0.951	1.090*	1.091	0.950	0.877

Table 6: Same as Table 4 with forecasts based on $n = 2000$.

5 Conclusion

We derive (uniform) central limit theory for forecasts of a wide range of risk measures in location-scale models. The estimators are motivated by a Pareto-type tail assumption for the innovations. This Pareto assumption allows all risk measure estimators to be expressed as simple functions of the Weissman (1978) quantile estimators, which is the reason for the generality of our approach regarding the risk measures. The finite-sample confidence intervals for the risk measures appear to work reasonably well, as demonstrated in Monte Carlo simulations, even though they may be improved using self-normalization or, potentially, other more refined methods. An empirical application demonstrates that the semi-parametric Pareto-type tail assumption can be used to obtain risk measure estimates that are preferable to completely non-parametric estimates in large samples.

A possible avenue for future research is to investigate bias-reduction techniques for the extreme DRM and expectile estimates based on the standardized residuals. This work could build on Gomes and Pestana (2007), who propose a bias-reduced VaR estimator for i.i.d. data. Another extension worth pursuing is to consider extreme conditional risk measure estimates in multivariate models for asset returns, similarly as was done in an unconditional context by Hoga (2018+b).

Appendix

For brevity, we put

$$\widehat{\mu}_{n+1} := \mu_{n+1}(\widehat{\boldsymbol{\theta}}_n), \quad \widehat{\sigma}_{n+1} := \sigma_{n+1}(\widehat{\boldsymbol{\theta}}_n) \qquad \text{and} \qquad \mu_{n+1} := \mu_{n+1}(\boldsymbol{\theta}^{\circ}), \quad \sigma_{n+1} := \sigma_{n+1}(\boldsymbol{\theta}^{\circ}).$$

Lemma 1. *Suppose Assumptions 2–3 hold. Then,*

$$\begin{aligned} \widehat{\mu}_{n+1} - \mu_{n+1} &= o_P\left(\frac{\log([1-\beta_n]/[1-\delta_n])}{\sqrt{n(1-\beta_n)}}\right), \qquad \textit{and} \\ \frac{\widehat{\sigma}_{n+1}}{\sigma_{n+1}} - 1 = &= o_P\left(\frac{\log([1-\beta_n]/[1-\delta_n])}{\sqrt{n(1-\beta_n)}}\right). \end{aligned}$$

Proof: It follows from Markov's inequality and Assumption 2 that, uniformly in n,

$$P\left\{\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}_0}\left\|\frac{\partial\mu_{n+1}}{\partial\boldsymbol{\theta}}(\boldsymbol{\theta})\right\| > K\right\} \leq K^{-\iota}\,\mathrm{E}\left[\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}_0}\left\|\frac{\partial\mu_{n+1}}{\partial\boldsymbol{\theta}}(\boldsymbol{\theta})\right\|\right]^{\iota} \underset{(K\to\infty)}{\longrightarrow} 0. \tag{A.1}$$

Let $\boldsymbol{\theta}, \widetilde{\boldsymbol{\theta}} \in \boldsymbol{\Theta}_0$. Then, we get by the mean value theorem that for some $\boldsymbol{\theta}_* \in \boldsymbol{\Theta}_0$ between $\boldsymbol{\theta}$ and $\widetilde{\boldsymbol{\theta}}$ that

$$\begin{aligned} \left|\mu_{n+1}(\boldsymbol{\theta}) - \mu_{n+1}(\widetilde{\boldsymbol{\theta}})\right| &= \left|\frac{\partial\mu_{n+1}}{\partial\boldsymbol{\theta}}(\boldsymbol{\theta}_*)^{\top}(\boldsymbol{\theta} - \widetilde{\boldsymbol{\theta}})\right| \\ &\leq \sup_{\boldsymbol{\theta}_*\in\boldsymbol{\Theta}_0}\left\|\frac{\partial\mu_{n+1}}{\partial\boldsymbol{\theta}}(\boldsymbol{\theta}_*)\right\| \cdot \left\|\boldsymbol{\theta} - \widetilde{\boldsymbol{\theta}}\right\|, \end{aligned} \tag{A.2}$$

where we used the Cauchy–Schwarz inequality in the second step. For brevity, put $c_n = \frac{\sqrt{n(1-\beta_n)}}{\log([1-\beta_n]/[1-\delta_n])}$. Now, for $\delta > 0$ chosen such that $\|\boldsymbol{\theta} - \boldsymbol{\theta}^{\circ}\| < \delta$ implies $\boldsymbol{\theta} \in \boldsymbol{\Theta}_0$, we have that

$$\begin{aligned} P\left\{c_n|\widehat{\mu}_{n+1} - \mu_{n+1}| > \varepsilon\right\} &\leq P\left\{c_n|\widehat{\mu}_{n+1} - \mu_{n+1}| > \varepsilon,\ \left\|\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}^{\circ}\right\| < \delta\right\} + P\left\{\left\|\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}^{\circ}\right\| \geq \delta\right\} \\ &= P\left\{c_n|\mu_{n+1}(\widehat{\boldsymbol{\theta}}_n) - \mu_{n+1}(\boldsymbol{\theta}^{\circ})| > \varepsilon,\ \left\|\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}^{\circ}\right\| < \delta\right\} + o(1) \\ &\overset{(A.2)}{\leq} P\left\{c_n \sup_{\boldsymbol{\theta}_*\in\boldsymbol{\Theta}_0}\left\|\frac{\partial\mu_{n+1}}{\partial\boldsymbol{\theta}}(\boldsymbol{\theta}_*)\right\| \cdot \left\|\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}^{\circ}\right\| > \varepsilon,\ \left\|\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}^{\circ}\right\| < \delta\right\} + o(1) \\ &= o(1), \end{aligned}$$

where we used Assumption 3 ($c_n\left\|\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}^{\circ}\right\| = o_P(1)$) and (A.1) ($\sup_{\boldsymbol{\theta}_*\in\boldsymbol{\Theta}_0}\left\|\frac{\partial\mu_{n+1}}{\partial\boldsymbol{\theta}}(\boldsymbol{\theta}_*)\right\| = \mathcal{O}_P(1)$) in the final step. The first claim follows.

It follows along similar lines that

$$\widehat{\sigma}_{n+1} - \sigma_{n+1} = o_P\left(\frac{\log([1-\beta_n]/[1-\delta_n])}{\sqrt{n(1-\beta_n)}}\right).$$

Dividing by $\sigma_{n+1} > 0$ does not change the rate, since $\sigma_{n+1}^{-1} = \mathcal{O}_P(1)$ by Assumption 2. $\square$

Proof of Theorem 1: We start with the proof of (20). Write

$$\frac{\widehat{\mathrm{DRM}}_{\delta_n,n}}{\mathrm{DRM}_{\delta_n,n}}-1=\frac{\frac{\widehat{\mu}_{n+1}-\mu_{n+1}}{\sigma_{n+1}\,\mathrm{DRM}_{\delta_n}(\varepsilon)}+\frac{\widehat{\sigma}_{n+1}}{\sigma_{n+1}}\left(\frac{\widehat{\mathrm{DRM}}_{\delta_n}(\widehat{\varepsilon})}{\mathrm{DRM}_{\delta_n}(\varepsilon)}-1\right)+\left(\frac{\widehat{\sigma}_{n+1}}{\sigma_{n+1}}-1\right)}{\frac{\mu_{n+1}}{\sigma_{n+1}\,\mathrm{DRM}_{\delta_n}(\varepsilon)}+1}$$
$$=\frac{(I)+(II)+(III)}{(IV)}.$$

We deal with the terms (I)–(IV) separately.

Consider (I). Due to (10) and the fact that $q_{\delta_n}(\varepsilon)\underset{(n\to\infty)}{\longrightarrow}\infty$ under the Pareto-type tail Assumption 4 (de Haan and Ferreira, 2006, Proposition B.1.9.1), we get $\mathrm{DRM}_{\delta_n}(\varepsilon)\underset{(n\to\infty)}{\longrightarrow}\infty$. Combine this with Lemma 1 and $\sigma_{n+1}^{-1}=\mathcal{O}_P(1)$ (from Assumption 2) to obtain

$$(I)=o_P\left(\frac{\log([1-\beta_n]/[1-\delta_n])}{\sqrt{n(1-\beta_n)}}\right).$$

Also from Lemma 1, we get

$$(III)=o_P\left(\frac{\log([1-\beta_n]/[1-\delta_n])}{\sqrt{n(1-\beta_n)}}\right).$$

Since $\mu_{n+1}=\mathcal{O}_P(1)$, $\sigma_{n+1}^{-1}=\mathcal{O}_P(1)$ (by Assumption 2) and $\mathrm{DRM}_{\delta_n}(\varepsilon)\underset{(n\to\infty)}{\longrightarrow}\infty$,

$$(IV)\overset{P}{\underset{(n\to\infty)}{\longrightarrow}}1.$$

Finally, for (II) we write

$$\frac{\widehat{\mathrm{DRM}}_{\delta_n}(\widehat{\varepsilon})}{\mathrm{DRM}_{\delta_n}(\varepsilon)}-1=\frac{\widehat{q}_{\delta_n}(\widehat{\varepsilon})}{q_{\delta_n}(\varepsilon)}\cdot\int_0^1 s^{-\widehat{\gamma}}\,\mathrm{d}g(s)\cdot\frac{q_{\delta_n}(\varepsilon)}{\mathrm{DRM}_{\delta_n}(\varepsilon)}-1$$
$$=\frac{\widehat{q}_{\delta_n}(\widehat{\varepsilon})}{q_{\delta_n}(\varepsilon)}\cdot\frac{\int_0^1 s^{-\widehat{\gamma}}\,\mathrm{d}g(s)}{\int_0^1 s^{-\gamma}\,\mathrm{d}g(s)}\cdot\left[1+o\left(\frac{1}{\sqrt{n(1-\beta_n)}}\right)\right]-1, \tag{A.3}$$

where the final step follows from Lemma 3 (ii) in El Methni and Stupfler (2017). Equation (9) in El Methni and Stupfler (2017) demonstrates that

$$\frac{\int_0^1 s^{-\widehat{\gamma}}\,\mathrm{d}g(s)}{\int_0^1 s^{-\gamma}\,\mathrm{d}g(s)}=1+\mathcal{O}_P\left(\frac{1}{\sqrt{n(1-\beta_n)}}\right) \tag{A.4}$$

under Assumption 5. Similarly as in the proof of Theorem 4.3.8 in de Haan and Ferreira (2006) (setting $k_n=n(1-\beta_n)$ and $p_n=1-\delta_n$ in their notation) it follows from Assumption 5 that

$$\frac{1}{\widehat{\sigma}}\frac{\sqrt{n(1-\beta_n)}}{\log([1-\beta_n]/[1-\delta_n])}\left(\frac{\widehat{q}_{\delta_n}(\widehat{\varepsilon})}{q_{\delta_n}(\varepsilon)}-1\right)\overset{\mathcal{D}}{\underset{(n\to\infty)}{\longrightarrow}}Z. \tag{A.5}$$

Combining (A.3) with (A.4) and (A.5) gives

$$\frac{1}{\widehat{\sigma}}\frac{\sqrt{n(1-\beta_n)}}{\log([1-\beta_n]/[1-\delta_n])}\left(\frac{\widehat{\mathrm{DRM}}_{\delta_n}(\widehat{\varepsilon})}{\mathrm{DRM}_{\delta_n}(\varepsilon)}-1\right)\xrightarrow[(n\to\infty)]{\mathcal{D}} Z.$$

Together with Lemma 1, this implies

$$(II)\xrightarrow[(n\to\infty)]{\mathcal{D}} Z.$$

Putting together the results for (I)–(IV), the conclusion follows.

It remains to prove the second statement (21). The proof is very similar, so we only highlight the main differences. Write

$$\begin{aligned}\frac{\widehat{\xi}_{\delta_n,n}}{\xi_{\delta_n,n}}-1&=\frac{\frac{\widehat{\mu}_{n+1}-\mu_{n+1}}{\sigma_{n+1}\xi_{\delta_n}(\varepsilon)}+\frac{\widehat{\sigma}_{n+1}}{\sigma_{n+1}}\left(\frac{\widehat{\xi}_{\delta_n}(\widehat{\varepsilon})}{\xi_{\delta_n}(\varepsilon)}-1\right)+\left(\frac{\widehat{\sigma}_{n+1}}{\sigma_{n+1}}-1\right)}{\frac{\mu_{n+1}}{\sigma_{n+1}\xi_{\delta_n}(\varepsilon)}+1}\\&=\frac{(I_*)+(II_*)+(III_*)}{(IV_*)}.\end{aligned}$$

By using (11) instead of (10) for (I_*), it follows as before that

$$(I_*)=o_P\left(\frac{\log([1-\beta_n]/[1-\delta_n])}{\sqrt{n(1-\beta_n)}}\right).$$

The terms (III_*) and (IV_*) converge at the same rate as (III) and (IV), respectively. Finally, consider (II_*) and write

$$\begin{aligned}\frac{\widehat{\xi}_{\delta_n}(\widehat{\varepsilon})}{\xi_{\delta_n}(\varepsilon)}-1&=\frac{\widehat{q}_{\delta_n}(\widehat{\varepsilon})}{q_{\delta_n}(\varepsilon)}\cdot(\widehat{\gamma}^{-1}-1)^{-\widehat{\gamma}}\cdot\frac{q_{\delta_n}(\varepsilon)}{\xi_{\delta_n}(\varepsilon)}-1\\&=\frac{\widehat{q}_{\delta_n}(\widehat{\varepsilon})}{q_{\delta_n}(\varepsilon)}\cdot\frac{(\widehat{\gamma}^{-1}-1)^{-\widehat{\gamma}}}{(\gamma^{-1}-1)^{-\gamma}}\cdot\left[1+o\left(\frac{1}{\sqrt{n(1-\beta_n)}}\right)\right]-1,\end{aligned}\tag{A.6}$$

where the second equality follows from Corollary 1 of Daouia *et al.* (2018); see also Equation (B.26) in the supplementary material to Daouia *et al.* (2018). Apply the Delta method to (18) in Assumption 5 to obtain that

$$\frac{1}{\widehat{\sigma}}\sqrt{n(1-\beta_n)}\left(\frac{(\widehat{\gamma}^{-1}-1)^{-\widehat{\gamma}}}{(\gamma^{-1}-1)^{-\gamma}}-1\right)\xrightarrow[(n\to\infty)]{\mathcal{D}}\Big[(1-\gamma)^{-1}-\log(\gamma^{-1}-1)\Big]Z.$$

Hence,

$$\frac{(\widehat{\gamma}-1)^{-\widehat{\gamma}}}{(\gamma-1)^{-\gamma}}=1+\mathcal{O}_P\left(\frac{1}{\sqrt{n(1-\beta_n)}}\right).\tag{A.7}$$

Combining (A.6) with (A.7) and (A.5) gives

$$\frac{1}{\widehat{\sigma}}\frac{\sqrt{n(1-\beta_n)}}{\log([1-\beta_n]/[1-\delta_n])}\left(\frac{\widehat{\xi}_{\delta_n}(\widehat{\varepsilon})}{\xi_{\delta_n}(\varepsilon)}-1\right)\xrightarrow[(n\to\infty)]{\mathcal{D}} Z.$$

Together with Lemma 1, this yields

$$(II_*) \underset{(n\to\infty)}{\overset{\mathcal{D}}{\longrightarrow}} Z.$$

Putting together the results for (I_*)–(IV_*), the conclusion follows similarly as before. $\square$

Proof of Theorem 2: We start with the result for $\widehat{\mathrm{DRM}}_{\delta_n(t),n}$. The proof is similar to that of Theorem 1 (i). Simply replace δ_n by $\delta_n(t)$ at every occurence in (I)–(IV) to obtain, say, (I^t)–(IV^t). Then, it is easy to check that (I^t), (III^t) and (IV^t) converge uniformly in $t \in [\underline{t}, \overline{t}]$ at the same rates as (I), (III) and (IV), respectively. Thus, it remains to consider

$$(II^t) = \frac{\widehat{\sigma}_{n+1}}{\sigma_{n+1}} \left(\frac{\widehat{\mathrm{DRM}}_{\delta_n(t)}(\widehat{\varepsilon})}{\mathrm{DRM}_{\delta_n(t)}(\varepsilon)} - 1 \right).$$

Write

$$\frac{\widehat{\mathrm{DRM}}_{\delta_n(t)}(\widehat{\varepsilon})}{\mathrm{DRM}_{\delta_n(t)}(\varepsilon)} - 1 = \frac{\widehat{q}_{\delta_n(t)}(\widehat{\varepsilon})}{q_{\delta_n(t)}(\varepsilon)} \cdot \frac{\int_0^1 s^{-\widehat{\gamma}}\,\mathrm{d}g(s)}{\int_0^1 s^{-\gamma}\,\mathrm{d}g(s)} \cdot \frac{q_{\delta_n(t)}(\varepsilon)}{\mathrm{DRM}_{\delta_n(t)}(\varepsilon)} \int_0^1 s^{-\gamma}\,\mathrm{d}g(s) - 1. \tag{A.8}$$

A close inspection of the proof of Lemma 3 (ii) of El Methni and Stupfler (2017) shows that

$$\frac{q_{\delta_n(t)}(\varepsilon)}{\mathrm{DRM}_{\delta_n(t)}(\varepsilon)} \int_0^1 s^{-\gamma}\,\mathrm{d}g(s) = 1 + o\left(\frac{1}{\sqrt{n(1-\beta_n)}} \right) \tag{A.9}$$

also holds uniformly in $t \in [\underline{t}, \overline{t}]$.

Exactly as in the proof of Theorem 2 in Hoga (2018+c) (with, in his notation, $k_n = n(1-\beta_n)$ and $p_n = 1-\delta_n$) it follows that

$$\frac{\sqrt{n(1-\beta_n)}}{\log([1-\beta_n]/[1-\delta_n(t)])} \left(\frac{\widehat{q}_{\delta_n(t)}(\widehat{\varepsilon})}{q_{\delta_n(t)}(\varepsilon)} - 1 \right) = \frac{\sqrt{n(1-\beta_n)}}{\log([1-\beta_n]/[1-\delta_n])} \left(\frac{\widehat{q}_{\delta_n}(\widehat{\varepsilon})}{q_{\delta_n}(\varepsilon)} - 1 \right) + o_P(1) \tag{A.10}$$

uniformly in $t \in [\underline{t}, \overline{t}]$.

Combining (A.4), (A.9), (A.10) with (A.8), we obtain

$$\frac{\sqrt{n(1-\beta_n)}}{\log([1-\beta_n]/[1-\delta_n(t)])} \left(\frac{\widehat{\mathrm{DRM}}_{\delta_n(t)}(\widehat{\varepsilon})}{\mathrm{DRM}_{\delta_n(t)}(\varepsilon)} - 1 \right) = \frac{\sqrt{n(1-\beta_n)}}{\log([1-\beta_n]/[1-\delta_n])} \left(\frac{\widehat{q}_{\delta_n}(\widehat{\varepsilon})}{q_{\delta_n}(\varepsilon)} - 1 \right) + o_P(1)$$

uniformly in $t \in [\underline{t}, \overline{t}]$. The conclusion follows with (A.5).

The proof for $\widehat{\xi}_{\delta_n(t),n}$ follows along similar lines using arguments in the proof of Theorem 1 (ii). We omit the details. The crucial difference is that

$$\frac{q_{\delta_n(t)}(\varepsilon)}{\xi_{\delta_n(t)}(\varepsilon)} = 1 + o\left(\frac{1}{\sqrt{n(1-\beta_n)}} \right)$$

holds uniformly in $t \in [\underline{t}, \overline{t}]$, as a careful reading of the proof of Corollary 1 in Daouia *et al.* (2018) shows; cf. (A.6). $\square$

References

Artzner P, Delbaen F, Eber JM, Heath D. 1999. Coherent measures of risk. *Mathematical Finance* **9**: 203–228.

Bee M, Dupuis DJ, Trapin L. 2016. Realizing the extremes: Estimation of tail-risk measures from a high-frequency perspective. *Journal of Empirical Finance* **36**: 86–99.

Beirlant J, Goegebeur Y, Segers J, Teugels J. 2004. *Statistics of Extremes*. Chichester: Wiley.

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

Bollerslev T. 1987. A conditionally heteroskedastic time series model for speculative prices and rates of return. *The Review of Economics and Statistics* **69**: 542–547.

Campbell JY, Lo AL, MacKinlay AC. 1997. *The Econometrics of Financial Markets*. Princeton: Princeton University Press.

Chan KF, Gray P. 2006. Using extreme value theory to measure value-at-risk for daily electricity spot prices. *International Journal of Forecasting* **22**: 283–300.

Chan NH, Deng SJ, Peng L, Xia Z. 2007. Interval estimation of value-at-risk based on GARCH models with heavy-tailed innovations. *Journal of Econometrics* **137**: 556–576.

Chavez-Demoulin V, Embrechts P, Sardy S. 2014. Extreme-quantile tracking for financial time series. *Journal of Econometrics* **181**: 44–52.

Csörgő S, Viharos L. 1998. Estimating the tail index. In Szyszkowicz B (ed.) *Asymptotic Methods in Probability and Statistics*, Amsterdam: Elsevier, pages 833–881.

Daníelsson J, de Haan L, Ergun LM, de Vries CG. 2016. Tail index estimation: Quantile driven threshold selection.

Daníelsson J, Jansen DW, de Vries CG. 1996. The method of moments ratio estimator for the tail shape parameter. *Communications in Statistics - Theory and Methods* **25**: 711–720.

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

de Haan L, Ferreira A. 2006. *Extreme Value Theory*. New York: Springer.

Diebold FX, Mariano RS. 1995. Comparing predictive accuracy. *Journal of Business & Economic Statistics* **13**: 253–263.

Dowd K, Blake D. 2006. After VaR: The theory, estimation, and insurance applications of quantile-based risk measures. *The Journal of Risk and Insurance* **73**: 193–229.

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

Du Z, Escanciano JC. 2017. Backtesting expected shortfall: Accounting for tail risk. *Management Science* **63**: 940–958.

El Methni J, Stupfler G. 2017. Extreme versions of Wang risk measures and their estimation for heavy-tailed distributions. *Statistica Sinica* **27**: 907–930.

Emmer S, Kratz M, Tasche D. 2015. What is the best risk measure in practice? A comparison of standard measures. *Journal of Risk* **18**: 31–60.

Fissler T, Ziegel JF. 2016. Higher order elicitability and Osband's principle. *The Annals of Statistics* **44**: 1680–1707.

Francq C, Zakoïan JM. 2010. *GARCH Models: Structure, Statistical Inference and Financial Applications*. Chichester: Wiley.

Francq C, Zakoïan JM. 2015. Risk-parameter estimation in volatility models. *Journal of Econometrics* **184**: 158–173.

Francq C, Zakoïan JM. 2016. Looking for efficient QML estimation of conditional VaRs at multiple risk levels. *Annals of Economics and Statistics* **123/124**: 9–28.

Gabaix X. 2009. Power laws in economics and finance. *Annual Review of Economics* **1**: 255–293.

Gao F, Song F. 2008. Estimation risk in GARCH VaR and ES estimates. *Econometric Theory* **24**: 1404–1424.

Gençay R, Selçuk F. 2004. Extreme value theory and value-at-risk: Relative performance in emerging markets. *International Journal of Forecasting* **20**: 287–303.

Glosten LR, Jagannathan R, Runkle DE. 1993. On the relation between the expected value and the volatility of the nominal excess return on stocks. *The Journal of Finance* **48**: 1779–1801.

Gneiting T. 2011. Making and evaluating point forecasts. *Journal of the American Statistical Association* **106**: 746–762.

Gomes MI, Pestana D. 2007. A sturdy reduced-bias extreme quantile (VaR) estimator. *Journal of the American Statistical Association* **102**: 280–292.

Hill B. 1975. A simple general approach to inference about the tail of a distribution. *The Annals of Statistics* **3**: 1163–1174.

Hill JB. 2010. On tail index estimation for dependent, heterogeneous data. *Econometric Theory* **26**: 1398–1436.

Hill JB. 2015. Tail index estimation for a filtered dependent time series. *Statistica Sinica* **25**: 609–629.

Hoga Y. 2018+a. Confidence intervals for conditional tail risk measures in ARMA–GARCH models. Forthcoming in *Journal of Business & Economic Statistics* (DOI: 10.1080/07350015.2017.1401543) .

Hoga Y. 2018+b. Detecting tail risk differences in multivariate time series. Forthcoming in *Journal of Time Series Analysis* (DOI: 10.1111/jtsa.12) .

Hoga Y. 2018+c. Extreme conditional tail moment estimation under serial dependence. Forthcoming in *Journal of Financial Econometrics* (DOI: 10.1093/jjfinec/nby016) .

Jorion P. 2006. *Value at Risk: the New Benchmark for Managing Financial Risk*. Chicago: McGraw–Hill, 3rd edn.

Kim L, Lee S. 2016. On the tail index inference for heavy-tailed GARCH-type innovations. *Annals of the Institute of Statistical Mathematics* **68**: 237–267.

Kou S, Peng X. 2016. On the measurement of economic tail risk. *Operations Research* **64**: 1056–1072.

Kuan CM, Yeh JH, Hsu YC. 2009. Assessing value at risk with CARE, the Conditional Autoregressive Expectile models. *Journal of Econometrics* **150**: 261–270.

Kuester K, Mittnik S, Paolella MS. 2006. Value-at-risk prediction: A comparison of alternative strategies. *Journal of Financial Econometrics* **4**: 53–89.

Linton O, Xiao Z. 2013. Estimation of and inference about the expected shortfall for time series with infinite variance. *Econometric Theory* **29**: 771–807.

Mancini L, Trojani F. 2011. Robust value at risk prediction. *Journal of Financial Econometrics* **9**: 281–313.

Martins-Filho C, Yao F, Torero M. 2018. Nonparametric estimation of conditional value-at-risk and expected shortfall based on extreme value theory. *Econometric Theory* **34**: 23–67.

McNeil AJ, 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.

Newey WK, Powell JL. 1987. Asymmetric least squares estimation and testing. *Econometrica* **55**: 819–847.

Pan J, Wang H, Tong H. 2008. Estimation and tests for power-transformed and threshold GARCH models. *Journal of Econometrics* **142**: 352–378.

Rocco M. 2014. Extreme value theory in finance: A survey. *Journal of Economic Surveys* **28**: 82–108.

Trapin L. 2018. Can volatility models explain extreme events? *Journal of Financial Econometrics* **16**: 297–315.

Wang CS, Zhao Z. 2016. Conditional value-at-risk: Semiparametric estimation and inference. *Journal of Econometrics* **195**: 86–103.

Wang R, Ziegel JF. 2015. Elicitable distortion risk measures: A concise proof. *Statistics & Probability Letters* **100**: 172–175.

Wang SS. 1996. Premium calculation by transforming the layer premium density. *ASTIN Bulletin* **26**: 71–92.

Wang SS. 2000. A class of distortion operators for pricing financial and insurance risks. *Journal of Risk and Insurance* **67**: 15–36.

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

Wirch JL, Hardy MR. 2001. Distortion risk measures: Coherence and stochastic dominance. In *International Congress on Insurance: Mathematics and Economics, Lisbon.*

Ziegel JF. 2016. Coherence and elicitability. *Mathematical Finance* **26**: 901–918.