Extreme Conditional Tail Moment Estimation under Dependence

Yannick Hoga*

March 13, 2017

*Faculty of Economics and Business Administration, University of Duisburg-Essen, Universitätsstraße 12, D-45117 Essen, Germany, tel. +49 201 1834365, yannick.hoga@vwl.uni-due.de. The author would like to thank Christoph Hanck for his detailed comments. Full responsibility is taken for all remaining errors. Support of DFG (HA 6766/2-2) is gratefully acknowledged.

Abstract

A wide range of risk measures can be written as functions of conditional tail moments and Value-at-Risk, for instance the Expected Shortfall. In this paper we derive joint central limit theory for semi-parametric estimates of conditional tail moments, including in particular Expected Shortfall, at arbitrarily small risk levels. We also derive confidence corridors for Value-at-Risk at different levels far out in the tails, which allows for simultaneous inference. We work under a semi-parametric Pareto-type assumption on the distributional tail of the observations and only require an extremal-near epoch dependence (E-NED) assumption. In simulations, our semi-parametric expected shortfall estimate is shown to be more accurate in terms of root mean square error than extant non-parametric estimates. An empirical application illustrates the proposed methods.

Keywords: Value-at-Risk, Expected Shortfall, E-NED, Pareto-type Tails, Confidence Corridor

JEL classification: C12 (Hypothesis Testing), C13 (Estimation), C14 (Semiparametric and Nonparametric Methods)

1 Motivation

The need to quantify risk, defined broadly, has lead to a burgeoning literature on risk measures. Two of the most popular risk measures in the financial industry are the Value-at-Risk at level $p \in (0,1)$ (VaR_p), defined as the upper p-quantile of the distribution of losses X, and the Expected Shortfall (ES) at level p, defined as the expected loss given an exceedance of VaR_p, $\mathrm{ES}_p = \mathrm{E}\left[X \mid X > \mathrm{VaR}_p\right]$. ES is defined if $\mathrm{E}\,|X| < \infty$ and is sometimes also called conditional tail expectation or tail-VaR. In contrast to ES, VaR is not a coherent risk measure in the sense of Artzner *et al.* (1999) and is uninformative as to the expected loss beyond the VaR. Yet, VaR is easy to estimate and to backtest (e.g., Daníelsson, 2011).

A unifying perspective on VaR, ES and a wide range of other popular risk measures was presented by El Methni *et al.* (2014). They introduced the conditional tail moment (CTM), i.e., the a-th moment ($a > 0$) of the loss given a VaR_p-exceedance, $\mathrm{CTM}_a(p) = \mathrm{E}\left[X^a \mid X > \mathrm{VaR}_p\right]$. For $a = 1$, the conditional tail moment reduces to the ES. For an appropriate choice of $a < 1$ the conditional tail moment may still be used for extremely heavy-tailed time series with $\mathrm{E}\,|X| = \infty$, when ES can no longer be used. For instance, there is evidence that economic losses in the aftermath of natural disasters have infinite means (Ibragimov *et al.*, 2009; Ibragimov and Walden, 2011). Then, El Methni *et al.* (2014) showed that many risk measures are functions of VaR and CTMs. Hence, by virtue of the continuous mapping theorem, weak limit theory for estimators of these risk measures can be grounded on joint asymptotics of VaR and CTM estimates.

Denote the ordered observations of a time series $X_1, \ldots, X_n$ by $X_{(1)} \geq \ldots \geq X_{(n)}$. While –

in the spirit of El Methni *et al.* (2014) – we develop limit theory for many risk measures, we shall frequently focus on our estimator of ES (or, equivalently, $\mathrm{CTM}_1(p)$). ES estimation for time series is a topic of recent interest, yet the literature almost exclusively focuses on the case where $\mathrm{E}\,|X_i|^2 < \infty$; see, e.g., Scaillet (2004); Chen (2008). However, evidence for infinite variance models is widespread. For instance, IGARCH models have a tail index equal to 2 and hairline infinite variance (Ling, 2007, Thm. 2.1 (iii)). We refer to Engle and Bollerslev (1986) and the references therein for evidence of the plausibility of IGARCH models for exchange rates and interest rates. Infinite variance phenomena can be found more generally in, e.g., insurance and internet traffic applications (Resnick, 2007, Examples 4.1 & 4.2), and emerging market stock returns and exchange rates (Hill, 2013, 2015a).

To the best of our knowledge, only Linton and Xiao (2013) and Hill (2015a) avoid a finite variance assumption for ES estimation of time series. Linton and Xiao (2013) essentially study a simple non-parametric estimate of ES,

$$\widehat{\mathrm{ES}}_p = \frac{1}{pn} \sum_{i=1}^{n} X_i I_{\left\{X_i \geq X_{(\lfloor pn \rfloor)}\right\}}, \tag{1}$$

where I_A denotes the indicator function for a set A, and $\lfloor \cdot \rfloor$ rounds to the nearest smallest integer. Linton and Xiao (2013) assume regularly varying tails

$$P\left\{|X_i| > x\right\} = x^{-1/\gamma} L(x), \qquad \text{where } L(\cdot) \text{ is slowly varying.} \tag{2}$$

In the case of the Pareto distribution $L(\cdot)$ is identically a constant, which is why distributions with (2) may be said to be of Pareto-type. Concretely, Linton and Xiao (2013) impose $\gamma \in (1/2, 1)$. Since moments of order greater than or equal to $1/\gamma$ do not exist but smaller ones do (de Haan and Ferreira, 2006, Ex. 1.16), this rules out infinite-mean models by $\gamma < 1$ (in which case ES does not exist anyway) and finite variance models by $\gamma > 1/2$. For geometrically strong-mixing $\{X_i\}$, they derive the stable limit of $n^{1-\gamma}(\widehat{\mathrm{ES}}_p - \mathrm{ES}_p)$, which however depends on the unknown γ. For feasible inference, they consider a subsampling procedure. Hill (2015a), who also works with geometrically strong-mixing random variables (r.v.s), uses a tail-trimmed estimate

$$\widehat{\mathrm{ES}}_p^{(*)} = \frac{1}{pn} \sum_{i=1}^{n} X_i I_{\left\{X_{(k_n)} \geq X_i \geq X_{(\lfloor pn \rfloor)}\right\}}, \tag{3}$$

where the integer trimming sequence $k_n < n$ tends to infinity with $k_n = o(n)$. This improves the convergence rate to $\sqrt{n}/g(n)$ for some slowly varying function $g(n) \to \infty$ if $\gamma \in [1/2, 1)$. His results also extend to $\gamma < 1/2$, where he obtains the standard $\sqrt{n}$-rate. In both cases, Hill (2015a) delivers standard Gaussian limit theory, although – in contrast to Linton and Xiao (2013) – he requires a second-order refinement of (2). To deal with possibly non-vanishing bias terms that may arise due to trimming, Hill (2015a) exploits regular variation and proposes an ES estimator $\widehat{\mathrm{ES}}_p^{(2)} = \widehat{\mathrm{ES}}_p^{(*)} + \widehat{\mathcal{R}}_n^{(2)}$ with optimal bias correction $\widehat{\mathcal{R}}_n^{(2)}$.

Despite working under a *semi*-parametric Pareto-tail assumption as in (2), Linton and Xiao (2013) and Hill (2015a) (essentially) only consider *non*-parametric estimators of ES, viz., $\widehat{\mathrm{ES}}_p$ and $\widehat{\mathrm{ES}}_p^{(2)}$. Only Hill (2015b) exploits assumption (2) for purposes of bias correction via $\widehat{\mathcal{R}}_n^{(2)}$ in the ES estimate $\widehat{\mathrm{ES}}_p^{(2)}$. In this paper we take a different tack and use (2) as a motivation for a truly *semi*-parametric of ES, and indeed more generally of CTMs. In a regression environment with covariates and independent, identically distributed (i.i.d.) observations, similar estimates have been studied by El Methni *et al.* (2014).

Our first main contribution is to derive the joint weak Gaussian limit of our VaR and CTM estimators under a general notion of dependence, covering and significantly extending the geometrically strong-mixing framework of Linton and Xiao (2013) and Hill (2015a). Thus, not only do we cover estimators of ES (as Linton and Xiao, 2013, and Hill, 2015a, do), but also – among others – those of VaR, conditional tail variance (Valdez, 2005) and conditional tail skewness (Hong and Elshahat, 2010); see El Methni *et al.* (2014). In our extreme value setting, we necessarily require that $p = p_n \to 0$ as $n \to \infty$, thus disadvantaging our estimator in a direct comparison of the rates obtained by Linton and Xiao (2013) and Hill (2015a) for $\widehat{\mathrm{ES}}_p$ and $\widehat{\mathrm{ES}}_p^{(2)}$; see also Remark 6 below. Nonetheless, we obtain a convergence rate that can improve the $n^{1-\gamma}$-rate for $\widehat{\mathrm{ES}}_p$. While the $\sqrt{n}/g(n)$-rate of $\widehat{\mathrm{ES}}_p^{(2)}$ cannot be beaten, we show in simulations that our estimator still has a lower root mean square error (RMSE). This is true for a wide range of values $p \in \{0.005, 0.01, 0.05, 0.1\}$, where – quite expectedly, as we focus on $p = p_n \to 0$ – the relative advantage becomes larger, the smaller p.

Our second main contribution is to derive confidence corridors for VaR at different levels. This is important because '[i]n financial risk management, the portfolio manager may be interested in different percentiles [...] of the potential loss and draw some simultaneous inference. This type of information provides the basis for dynamically managing the portfolio to control the overall risk at different levels' (Wang and Zhao, 2016, p. 90). Working with VaR (albeit conditioned on past returns) Wang and Zhao (2016) derive a functional central limit theorem for VaR estimates indexed by the level $p \in [\delta, 1-\delta]$ for some $\delta > 0$. While Wang and Zhao (2016, Rem. 2) conjecture that an extension to the interval $p \in (0, 1)$ may be possible, their current results exclude the tails of the distributions, which are of particular interest in risk management. We fill this gap in the present extreme value setting, where the tail is the natural focus.

The rest of the paper proceeds as follows. Section 2 states the main theoretical results. Subsection 2.1 derives joint central limit theory for CTMs and VaR. Subsection 2.2 derives confidence corridors for VaR at different levels, allowing for simultaneous inference. In the simulations in Section 3, the finite-sample performance is illustrated and compared with $\widehat{\mathrm{ES}}_p^{(2)}$. An application in Section 4 applies the results to the time series of VW log-returns during the attempted takeover by Porsche, that ultimately failed. The final Section 5 concludes. Proofs are relegated to the Appendix.

2 Main results

2.1 Limit theory for extreme conditional tail moments

Let $\{X_i\}$ be a strictly stationary sequence of non-negative r.v.s, whose right tail will be studied as is customary in extreme value theory. In practice, non-negativity may be achieved via a simple transformation, e.g., $X_i I_{\{X_i \geq 0\}}$ or $-X_i I_{\{-X_i \geq 0\}}$ if interest centers on the right- or left-tail, respectively. Define the survivor function $\overline{F}(\cdot) = 1 - F(\cdot)$, where F denotes the distribution function of X_1. We assume regularly varying tails $\overline{F}(\cdot) \in RV_{-1/\gamma}$, i.e.,

$$\lim_{x\to\infty} \frac{\overline{F}(\lambda x)}{\overline{F}(x)} = \lambda^{-1/\gamma} \qquad \forall\ \lambda > 0, \tag{4}$$

where $\gamma > 0$ is called the *extreme value index* and $\alpha = 1/\gamma$ the *tail index*. Note that (4) is equivalent to

$$\overline{F}(x) = x^{-1/\gamma} L(x), \qquad \text{where } L(\cdot) \text{ is slowly varying, i.e.,} \quad \lim_{x\to\infty} \frac{L(\lambda x)}{L(x)} = 1. \tag{5}$$

This in turn is equivalent to (de Haan and Ferreira, 2006, p. 25)

$$U(x) = x^{\gamma} L_U(x), \qquad \text{where } U(x) = F^{\leftarrow}(1 - 1/x) \quad \text{and} \quad L_U(\cdot) \text{ is slowly varying.} \tag{6}$$

Since (4) is an asymptotic relation, we require an *intermediate sequence* $k_n \to \infty$ with $k_n = o(n)$ and $k_n < n$ for statistical purposes. This sequence k_n is restricted by the following assumption.

Assumption 1. *There exists a function $A(\cdot)$ with $\lim_{x\to\infty} A(x) = 0$ such that for some $\rho < 0$*

$$\lim_{x\to\infty} \frac{\frac{\overline{F}(\lambda x)}{\overline{F}(x)} - \lambda^{-1/\gamma}}{A(x)} = \lambda^{-1/\gamma} \frac{\lambda^{\rho/\gamma} - 1}{\gamma\rho} \qquad \forall\ \lambda > 0. \tag{7}$$

Additionally, $\sqrt{k_n} A\big(U(n/k_n)\big) \longrightarrow 0$, as $n \to \infty$.

Remark 1. This assumption controls the speed of convergence in (4) and is consequently referred to as a second-order condition in extreme value theory (EVT). Equivalently, it may also be written in terms of the quantile function $U(\cdot)$ from (6) (see de Haan and Ferreira, 2006, Thm. 2.3.9). In this form, it is widely-used in tail index (e.g., Einmahl *et al.*, 2016; Hoga, 2017+a) and extreme quantile estimation (e.g., Chan *et al.*, 2007; Hoga, 2017+b). Examples of d.f.s satisfying Assumption 1 are abundant. For instance, d.f.s expanding as

$$\overline{F}(x) = c_1 x^{-1/\gamma} + c_2 x^{-1/\gamma+\rho/\gamma}\big(1+o(1)\big), \qquad x\to\infty, \qquad (c_1>0,\ c_2\neq 0,\ \gamma>0,\ \rho<0) \tag{8}$$

fulfill Assumption 1 with the indicated γ and ρ, and $k_n = o(n^{-2\rho/(1-2\rho)})$ (de Haan and Ferreira, 2006, pp. 76-77). The more negative ρ, the closer the tail is to actual Pareto decay ($\rho=-\infty$). In the Pareto case, $k_n = o(n)$ can be chosen quite large, which is desirable for reasons detailed in Remark 6. The expansion in (8) is satisfied by, e.g., the Student t_ν-distribution with $\gamma = 1/\nu$ and $\rho=-2$, where $\nu>0$ denotes the degrees of freedom.

Define $x_p = F^{\leftarrow}(1-p)$ as the $(1-p)$-quantile for short. Most of the literature, including Linton and Xiao (2013) and Hill (2015a), focuses on the case where $p\in(0,1)$ is fixed. EVT however allows for $p=p_n\to 0$ as $n\to\infty$. Approximations derived from EVT often provide better approximations when p is small – the case of particular interest in risk management –, as they take the semi-parametric tail (4) into account. The following two motivations show how the regular variation of the tail is taken into account.

First, we use the regular-variation assumption (4) to estimate x_{p_n} in $\mathrm{CTM}_a(p_n) = \mathrm{E}\left[X^a \mid X > x_{p_n}\right]$ as follows. Note that p_n can be very small, such that x_{p_n} may lie outside the range of observations $X_1,\ldots,X_n$. Then, the idea is to base estimation of x_{p_n} on a less extreme (in-sample) quantile $x_{k_n/n}$ and use (4) to extrapolate from that estimate. Concretely, set $x = x_{k_n/n}$, $\lambda = x_{p_n}/x_{k_n/n}$ and use (4) as an approximation to obtain

$$\left(\frac{x_{p_n}}{x_{k_n/n}}\right)^{-1/\gamma} \approx \frac{1-F(x_{p_n})}{1-F(x_{k_n/n})} \approx \frac{np_n}{k_n}. \tag{9}$$

Replacing population with empirical quantities, this approximation motivates the so-called Weissman (1978) estimator $\widehat{x}_{p_n} = d_n^{\widehat{\gamma}} X_{(k_n+1)}$, where $d_n = k_n/(np_n)$. It has been used in, e.g., Drees (2003), Chan *et al.* (2007), or Hoga and Wied (2017). Of course, there is a wide range of estimators $\widehat{\gamma}$. We will use the Hill (1975) estimator

$$\widehat{\gamma} = \frac{1}{k_n}\sum_{i=1}^{k_n} \log\left(X_{(i)}/X_{(k_n+1)}\right)$$

in the following, which is arguably the most popular one (see, e.g., Hsing, 1991; Hill, 2010, and the references therein).

For the second approximation we exploit (4) once again. Together with Pan *et al.* (2013, Thm. 4.1), which was obtained from Karamata's theorem, this assumption implies $\mathrm{CTM}_a(p_n) \sim \frac{x_{p_n}^a}{1-a\gamma}$ as $n\to\infty$. Asymptotic equivalence, $a_n \sim b_n$, is defined as $\lim_{n\to\infty} a_n/b_n = 1$. Thus, the following estimate suggests itself:

$$\widehat{\mathrm{CTM}}_a(p_n) := \frac{\widehat{x}^a_{p_n}}{1 - a\widehat{\gamma}}. \tag{10}$$

This estimator accounts for the regular variation both in estimating x_{p_n} (through (9)) as well as in calculating the expected loss above x_{p_n} (through $\mathrm{CTM}_a(p_n) \sim \frac{x^a_{p_n}}{1-a\gamma}$).

Next, we introduce a sufficiently general dependence concept. The asymptotic behavior of $\widehat{\mathrm{CTM}}_a(p_n)$ crucially relies on that of $\widehat{\gamma}$ (see the proof of Theorem 1). To the best of our knowledge, the most general conditions under which extreme value index estimators have been studied, are those in Hill (2010). He develops central limit theory for the Hill (1975) estimator under L_2-extremal-near epoch dependence (L_2-E-NED). Similar to the mixing conditions of Hsing (1991), dependence is restricted only in the extremes. However, the NED property is often more easily verified (e.g., for ARMA-GARCH models) and offers more generality, whereas mixing conditions are typically harder to verify and some simple time series models fail to be mixing (e.g., Andrews, 1984).

For the following introduction to E-NED, it will be illustrative to keep an ARMA(p, q)-GARCH$(\overline{p}, \overline{q})$ model $\{X_i\}$ in mind. It is generated by the ARMA(p, q) structure

$$X_i = \mu + \sum_{t=1}^{p} \phi_t X_{i-t} + \sum_{t=1}^{q} \theta_t \epsilon_{i-t} + \epsilon_i,$$

which is driven by a GARCH$(\overline{p}, \overline{q})$ process $\{\epsilon_i\}$, i.e.,

$$\epsilon_i = \sigma_i U_i, \quad \text{where} \quad \sigma_i^2 = \omega + \sum_{t=1}^{\overline{p}} \alpha_t \epsilon^2_{i-t} + \sum_{t=1}^{\overline{q}} \beta_t \sigma^2_{i-t}.$$

In the following, dependence is restricted separately in the errors $\{\epsilon_i\}$ and the actual (observed) process $\{X_i\}$.

Consider a process $\{\epsilon_i\}$ (the GARCH process in the above example) and a possibly vector-valued functional of it, $\{E_{n,i}\}_{n\in\mathbb{N};i=1,\dots,n}$. The array nature of $E_{n,i}$ allows for tail functionals, such as $E_{n,i} = I_{\{\epsilon_i > a_{n,i}\}}$ for some triangular array $a_{n,i} \to \infty$ as $n \to \infty$. The $E_{n,i}$ induce σ-fields $\mathcal{F}^t_{n,s} = \sigma\left(E_{n,i} : s \leq i \leq t\right)$ (where $E_{n,i} = 0$ for $i \notin \{1, \dots, n\}$), which can be used to restrict dependence in $\{\epsilon_i\}$ using the mixing coefficients

$$\varepsilon_{n,q_n} := \sup_{A\in\mathcal{F}^i_{n,-\infty},\ B\in\mathcal{F}^\infty_{n,i+q_n}:\ i\in\mathbb{Z}} \left|P(A\cap B) - P(A)P(B)\right|,$$

$$\omega_{n,q_n} := \sup_{A\in\mathcal{F}^i_{n,-\infty},\ B\in\mathcal{F}^\infty_{n,i+q_n}:\ i\in\mathbb{Z}} \left|P(B|A) - P(B)\right|.$$

Here, $\{q_n\} \subset \mathbb{N}$ is a sequence of integer displacements with $1 \leq q_n < n$ and $q_n \to \infty$. We then say that $\{\epsilon_i\}$ is *F-strong (uniform) mixing with size* $\lambda > 0$ if

$$(n/k_n)q_n^\lambda \varepsilon_{n,q_n} \underset{(n\to\infty)}{\longrightarrow} 0 \qquad \left((n/k_n)q_n^\lambda \omega_{n,q_n} \underset{(n\to\infty)}{\longrightarrow} 0 \right).$$

Given $\{\epsilon_i\}$ thus restricted, it remains to restrict dependence in the observed series $\{X_i\}$ (the ARMA-GARCH process in the above example). Hill (2010) shows that the asymptotics of the Hill (1975) estimator can be grounded on tail arrays $\left\{I_{\{X_i>b_n e^u\}}\right\}$, where $b_n = U(1-k_n/n)$. Hence, dependence in $\{X_i\}$ need only be restricted via $\left\{I_{\{X_i>b_n e^u\}}\right\}$. This is achieved by assuming that, for some $p>0$, $\{X_i\}$ is L_p*-E-NED on* $\left\{\mathcal{F}_{n,1}^i\right\}$ *with size* $\lambda>0$, i.e.,

$$\left\| I_{\{X_i>b_n e^u\}} - P\left\{X_i > b_n e^u \;\middle|\; \mathcal{F}_{n,i-q_n}^{i+q_n}\right\}\right\|_p \leq f_{n,i}(u)\cdot\psi_{q_n},$$

where $f_{n,i} : [0,\infty) \to [0,\infty)$ is Lebesgue measurable, $\sup_{i=1,\ldots,n}\sup_{u\geq 0} f_{n,i}(u) = \mathcal{O}\left((k_n/n)^{1/p}\right)$, and $\psi_{q_n} = o(q_n^{-\lambda})$. For more on this dependence concept, we refer to Hill (2009, 2010, 2011).

Assumption 2. $\{X_i\}$ *is* L_2*-E-NED on* $\left\{\mathcal{F}_{n,1}^i\right\}$ *with size* $\lambda = 1/2$*. The constants* $f_{n,i}(u)$ *are integrable on* $[0,\infty)$ *with* $\sup_{i=1,\ldots,n}\int_0^\infty f_{n,i}(u)\mathrm{d}u = \mathcal{O}(\sqrt{k_n/n})$*. The base* $\{\epsilon_i\}$ *is either F-uniform mixing with size* $r/[2(r-1)]$*,* $r\geq 2$*, or F-strong mixing with size* $r/(r-2)$*,* $r>2$*.*

The final assumption we require is

Assumption 3. *The covariance matrix of*

$$\begin{pmatrix} \frac{1}{\sqrt{k_n}}\sum_{i=1}^n\left[\log(X_i/b_n)_+ - \mathrm{E}\log(X_i/b_n)_+\right] \\ \frac{1}{\sqrt{k_n}}\sum_{i=1}^n\left[I_{\left\{X_i>b_n e^{u/\sqrt{k_n}}\right\}} - P\left\{X_i > b_n e^{u/\sqrt{k_n}}\right\}\right] \end{pmatrix}$$

is positive definite uniformly in $n\in\mathbb{N}$ *for all* $u\in\mathbb{R}$*.*

Assumptions 2 and 3 are identical to Assumptions A.2 and D in Hill (2010), whereas Assumption 1 is stronger than the corresponding Assumption B in Hill (2010). Assumption 3 is used to show consistency of estimates of the asymptotic variance of the Hill (1975) estimator in Hill (2010, Thm. 3). This estimator, $\widehat{\sigma}_{k_n}^2$, appears in Theorem 1, because the asymptotics of $\widehat{x}_{p_n}$ are grounded on those of $\widehat{\gamma}$; see the proof of Theorem 1 and in particular the proof of Theorem 4.3.9 in de Haan and Ferreira (2006). The strengthening of Assumption B of Hill (2010) in Assumption 1 is required to derive limit theory for $\widehat{x}_{p_n}$ (see the proof of de Haan and Ferreira, 2006, Theorem 4.3.9).

Theorem 1. *Let* $a_1,\ldots,a_J$ *be positive and* $a_{J+1}=1$*. Assume that*

$$np_n = o(k_n) \qquad \textit{and} \qquad \log(np_n) = o(\sqrt{k_n}). \tag{11}$$

Suppose that Assumption 1 is met for $0 < \gamma < \max\{a_1, \ldots, a_{J+1}\}$. Suppose further that Assumptions 2 and 3 are met. Then

$$\frac{1}{\widehat{\sigma}_{k_n}} \frac{\sqrt{k_n}}{\log d_n} \left[\left(\frac{\widehat{\mathrm{CTM}}_{a_j}(p_n)}{\mathrm{CTM}_{a_j}(p_n)} - 1 \right)_{j=1,\ldots,J}, \left(\frac{\widehat{x}_{p_n}}{x_{p_n}} - 1 \right) \right]' \tag{12}$$

converges in distribution to a zero-mean Gaussian limit with covariance matrix $\boldsymbol{\Sigma} = (a_i a_j)_{i,j \in \{1,\ldots,J+1\}}$ and

$$\widehat{\sigma}^2_{k_n} := \frac{1}{k_n} \sum_{i,j=1}^{n} w\left(\frac{s-t}{\gamma_n}\right) \left[\log\left(\max\left\{ \frac{X_i}{X_{(k_n+1)}}, 1 \right\} \right) - \frac{k_n}{n} \widehat{\gamma} \right] \left[\log\left(\max\left\{ \frac{X_j}{X_{(k_n+1)}}, 1 \right\} \right) - \frac{k_n}{n} \widehat{\gamma} \right]$$

is a kernel-variance estimator with Bartlett kernel $w(\cdot)$, bandwidth $\gamma_n \to \infty$ with $\gamma_n = o(n)$, and $k_n/\sqrt{n} \to \infty$.

Remark 2. Condition (11) restricts the decay of $p_n \to 0$. Here, $p_n = o(k_n/n)$ describes the upper bound, required for the EVT approach to make sense, whereas $\log((n/k_n)p_n) = o(\log(np_n)) = o(\sqrt{k_n})$ prohibits p_n from decaying to zero too fast and thus describes the boundary, where extrapolation becomes infeasible.

Remark 3. The estimator $\widehat{\sigma}^2_{k_n}$ is due to Hill (2010, Sec. 4). Other possible choices for the kernel $w(\cdot)$ include the Parzen, quadratic spectral and Tukey-Hanning kernel.

Remark 4. It is interesting to contrast Theorem 1 with the fixed-p result in Linton and Xiao (2013). There, replacing the estimate $X_{(\lfloor (1-p)n \rfloor)}$ with the true quantile x_p in (1) does not change the limit of $n^{1-\gamma}(\widehat{\mathrm{ES}}_p - \mathrm{ES}_p)$ and the joint distribution of the VaR and the ES estimate is asymptotically independent (Linton and Xiao, 2013, pp. 778-779). In our case where $p = p_n \to 0$, the ES estimate *is* essentially the VaR estimate by (10) and the limit distributions of both estimates are perfectly linearly dependent by (A.3) in the Appendix.

Remark 5. The result of Theorem 1 is sufficient to deliver weak limit theory not only for VaR and ES, but also for a wide range of risk measures, e.g., the conditional tail variance, conditional tail skewness, conditional VaR. For terminology and more detail, we refer to El Methni *et al.* (2014).

Remark 6. It may be instructive to compare the rate of convergence from Theorem 1 for our ES estimator $\widehat{\mathrm{CTM}}_1(p_n)$ with the rates of $\widehat{\mathrm{ES}}_p$ and $\widehat{\mathrm{ES}}_p^{(2)}$. As pointed out in Remark 4, for $\gamma \in (1/2, 1)$ Linton and Xiao (2013) obtained a rate of $n^{1-\gamma}$ for $\widehat{\mathrm{ES}}_p$. Up to slowly varying terms, Hill (2015a) improves this rate to $\sqrt{n}$ for $\widehat{\mathrm{ES}}_p^{(2)}$ and general $\gamma < 1$. Recalling from Section 2.1 that $\mathrm{CTM}_1(p_n) \sim$

$U(1/p_n)/(1-\gamma)$, Theorem 1 implies, for $\gamma<1$,

$$\frac{1-\gamma}{\widehat{\sigma}_{k_n}}\frac{\sqrt{k_n}}{(\log d_n)U(1/p_n)}\left(\widehat{\mathrm{CTM}}_1(p_n)-\mathrm{CTM}_1(p_n)\right)\underset{(n\to\infty)}{\overset{\mathcal{D}}{\longrightarrow}}N(0,1).$$

To maximize the rate, we choose k_n as large as allowed by Assumption 1, i.e., $k_n=n^{1-\delta}/g(n)=o(n^{-2\rho/(1-2\rho)})$ for $\delta=1/(1-2\rho)$. Here, $g(\cdot)$ is a slowly varying function with $g(n)\underset{(n\to\infty)}{\longrightarrow}\infty$ as slowly as desired; e.g., $g(n)=\log n$ or $g(n)=\log(\log n)$. Since $p_n\to 0$ (such that $U(1/p_n)\to\infty$) in our framework, $\mathrm{CTM}_1(p_n)$ is at a disadvantage compared with $\widehat{\mathrm{ES}}_p$ and $\widehat{\mathrm{ES}}_p^{(2)}$, where $p\in(0,1)$ is fixed. So to make the comparison fairer, we choose the largest possible rate for p_n allowed by $np_n=o(k_n)$ from (11). Concretely, we set $p_n=k_n/(n\cdot g(n))=1/(n^{\delta}g(n))$. So the rate in our case is given by

$$\frac{\sqrt{k_n}}{(\log d_n)U(1/p_n)}\overset{(6)}{=}\sqrt{n}\frac{n^{-\delta/2}p_n^{\gamma}}{\sqrt{g(n)}(\log k_n/(np_n))L_U(1/p_n)}=\sqrt{n}\frac{n^{-\delta(1/2+\gamma)}}{\sqrt{g(n)}(\log g(n))L_U(1/p_n)g(n)^{\gamma}}.$$

Hence, up to terms of slow variation, the rate is given by $\sqrt{n}n^{-\delta(1/2+\gamma)}$. Two intuitive observations can be made. First, the larger γ (i.e., the heavier the tail), the slower the rate of convergence. This is to be expected, because the Hill estimate $\widehat{\gamma}$ (upon which our asymptotic results rest) has larger variance for larger γ – everything else being equal. For instance, for the t_ν-distribution with $\gamma=1/\nu$ and $\rho=-2$, one may choose $k_n=o(n^{-2\rho/(1-2\rho)})=o(n^{4/5})$ irrespective of the degrees of freedom ν (recall Remark 1). Then, for i.i.d. observations with d.f.s satisfying Assumption 1, de Haan and Ferreira (2006, Thm. 3.2.5) implies $\sqrt{k_n}(\widehat{\gamma}_k-\gamma)\overset{\mathcal{D}}{\longrightarrow}N(0,\gamma^2)$, as $n\to\infty$.

Second, the more negative ρ, the smaller $\delta=1/(1-2\rho)>0$ and hence the better the rate. This result is also expected, since a more negative ρ implies a better fit to true Pareto behavior; see Remark 1. So the heavier the tail (the larger γ), the better our method can be expected to work relative to the non-parametric estimate $\widehat{\mathrm{ES}}_p$.

So under the caveat that $\widehat{\mathrm{CTM}}_1(p_n)$ is at a disadvantage, a direct comparison of the convergence rates reveals the following. While the $\sqrt{n}$-rate (up to terms of slow variation) of $\widehat{\mathrm{ES}}_p^{(2)}$ cannot be obtained, the $n^{1-\gamma}$-rate of $\widehat{\mathrm{ES}}_p$ can be improved upon. For instance, for the t_ν-distribution (where $\gamma=1/\nu$ and $\rho=-2$) we obtain a rate of $\sqrt{n}n^{-\delta(1/2+\gamma)}=n^{1/5(2-\gamma)}$, which is faster (slower) than $n^{1-\delta}$ for $\gamma>3/4$ ($\gamma<3/4$).

2.2 Simultaneous inference on VaR

Working with VaR conditioned on past returns, Wang and Zhao (2016) and Francq and Zakoïan (2016) argue that it is desirable in risk management to be able to draw simultaneous inference on VaR at multiple risk levels. Theorem 2 below shows that in our (unconditional) extreme value context this is particularly easy. Heuristically, if the assumptions of Theorem 1 are met for some sequence $p_n\to 0$, then this also holds for the sequence $p_n(t) := p_n t$ for $t \in \left[\underline{t}, \overline{t}\right]$ $(0 < \underline{t} < \overline{t} < \infty)$, which suggests that $\widehat{x}_{p_n}(t) := X_{(k_n+1)}(k/(np_n(t)))^{\widehat{\gamma}}$ and $\widehat{x}_{p_n}$ should behave very similarly. Note that $\widehat{x}_{p_n} = \widehat{x}_{p_n}(1)$.

Theorem 2. *Under the conditions of Theorem 1 we have that for* $0 < \underline{t} < \overline{t} < \infty$

$$\sup_{t\in[\underline{t},\overline{t}]} \left| \frac{1}{\widehat{\sigma}_{k_n}} \frac{\sqrt{k_n}}{\log d_n(t)} \log\left(\frac{\widehat{x}_{p_n}(t)}{x_{p_n}(t)}\right) \right| \underset{(n\to\infty)}{\overset{\mathcal{D}}{\longrightarrow}} |Z|,$$

where $Z \sim \mathcal{N}(0,1)$, $d_n(t) = k_n/(np_n(t))$ *and* $x_{p_n}(t) = F^{\leftarrow}(1 - p_n(t))$.

Then uniform convergence in $t \in [\underline{t}, \overline{t}]$ of Theorem 2 suggests the following $(1-\beta)$-confidence corridor for VaR with levels between $p_n(\underline{t})$ and $p_n(\overline{t})$:

$$\widehat{x}_{p_n}(t) \exp\left\{-\Phi\left(1 - \frac{\beta}{2}\right) \frac{\log(d_n(t))}{\sqrt{k_n}}\right\} \leq x_{p_n}(t) \leq \widehat{x}_{p_n}(t) \exp\left\{\Phi\left(1 - \frac{\beta}{2}\right) \frac{\log(d_n(t))}{\sqrt{k_n}}\right\}. \tag{13}$$

It is surprising that the width of the confidence corridor for $x_{p_n}(t)$ does not depend on the values of $\underline{t}$ and $\overline{t}$. Indeed, the confidence corridor is simply obtained by calculating pointwise confidence intervals for $\widehat{x}_{p_n}(t)$. This can be explained by the Pareto-approximation that pins down the tail very precisely by extrapolation. Clearly, in finite sample one may not choose $\overline{t}$ too large, because then the quality of the Pareto-approximation will suffer, rendering confidence corridors (13) imprecise. Also, in actual applications one may not choose $\underline{t}$ too small, as this would push the boundaries of extrapolation too far. So in practice a judicious choice of $\underline{t}$ and $\overline{t}$ (and p_n) is required. In an application in Section 4, some guidance on this issue is given. A similar, yet non-uniform, version of Theorem 2 is given under a more restrictive β-mixing condition in Drees (2003, Thm. 2.2).

Remark 7. Gomes and Pestana (2007, Sec. 3.4) found in simulations that the finite-sample distribution of $\log\left(\widehat{x}_{p_n}/x_{p_n}\right)$ is in better agreement with the asymptotic distribution than $(\widehat{x}_{p_n}/x_{p_n} - 1)$. This may be due to $\log(\widehat{x}_{p_n}) = \widehat{\gamma}\log(d_n) + \log(X_{(k_n+1)})$ being a linear function of $\widehat{\gamma}$, upon which the asymptotic results rest (see the proof of de Haan and Ferreira, 2006, Thm. 4.3.9).

Remark 8. A close inspection of the proofs of Theorems 1 and 2 reveals that the methodology of this section may also be applied to conditional tail moments. For instance, for our ES estimator we obtain

$$\sup_{t\in[\underline{t},\overline{t}]} \left| \frac{1}{\widehat{\sigma}_{k_n}} \frac{\sqrt{k_n}}{\log d_n(t)} \log\left(\frac{\widehat{\mathrm{CTM}}_1(p_n(t))}{\mathrm{CTM}_1(p_n(t))}\right) \right| \underset{(n\to\infty)}{\overset{\mathcal{D}}{\longrightarrow}} |Z|,$$

where $Z \sim \mathcal{N}(0,1)$.

3 Simulations

This section compares the root mean squared error (RMSE) of our ES estimator $\widehat{\mathrm{CTM}}_1(p_n)$ with the optimally bias-corrected estimator $\widehat{\mathrm{ES}}_{p_n}^{(2)}$ of Hill (2015a). In his comparison of the finite-sample performance of $\widehat{\mathrm{ES}}_{p_n}^{(2)}$ and the untrimmed $\widehat{\mathrm{ES}}_{p_n}$, Hill (2015a, p. 21) finds that 'trimming does not impose a detectable penalty in terms of small sample mean-squared-error.' So for brevity we only report the results for $\widehat{\mathrm{ES}}_{p_n}^{(2)}$. We carry out the comparison for realistic models of financial and insurance data. As models for financial time series we use an AR(1)-GARCH$(1,1)$ with skewed-t innovations and a GARCH$(1,1)$ model with t-noise, both from Bücher *et al.* (2015, Sec. 5.2). Bücher *et al.* (2015) found that these two stationary and heavy-tailed models provide a good fit to the NASDAQ and DJIA log-returns from January 4, 1984 to December 31, 1990. We use the resulting parameter estimates from Bücher *et al.* (2015, Table 7). To the best of our knowledge, no results on the regular variation of AR(1)-GARCH$(1,1)$ processes exist. Yet, as both AR(1)-ARCH(1) and GARCH$(1,1)$ processes have regularly varying tails (see Fasen *et al.*, 2010, and the references therein), the same property is likely to hold for AR(1)-GARCH$(1,1)$ models as well. Verifying the second-order Assumption 1 is notoriously difficult for time series models, so it is frequently treated as a given (Shao and Zhang, 2010; Hill, 2015b).

As models for insurance data we use i.i.d. draws from a Burr distribution with survivor function

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

This is a popular class of distributions in insurance, because it offers more flexibility than the Pareto distribution (e.g., Burnecki *et al.*, 2011). Its tail index is given by $\alpha = \tau\lambda$ and the slowly varying function $A(\cdot)$ in Assumption 1 can be chosen as a constant multiple of $x^{-\tau}$. Hence, the larger $\tau > 0$, the faster the convergence to true Pareto behavior in (7). In insurance applications one often finds for the tail index that $\alpha \in (1,2)$ (see, e.g., Resnick, 2007), which motivates our choices of $\tau = 2$ and $\lambda = 0.75$, and $\tau = 3$ and $\lambda = 0.5$, both resulting in $\alpha = 1.5$. For the latter choice where τ is larger (and hence the Pareto approximation more accurate), we expect improved performance of our estimator relative to $\widehat{\mathrm{ES}}_{p}^{(2)}$, which only partially takes into account the Pareto-type tail for bias correction.

Both estimators $\widehat{\mathrm{CTM}}_1(p_n)$ and $\widehat{\mathrm{ES}}_{p_n}^{(2)}$ depend on a sequence k_n that is only specified asymptotically. Hence, some guidance for the choice of k_n in finite-samples is required. For $\widehat{\mathrm{ES}}_{p_n}^{(2)}$, Hill (2015a, Sec. 3) proposes to choose the intermediate sequence $k_n = \min\left\{1, \lfloor 0.25 n^{2/3}/(\log n)^{2\cdot 10^{-10}} \rfloor\right\}$, a fixed function of n. However, for the bias correction term $\widehat{\mathcal{R}}_n^{(2)}$ in $\widehat{\mathrm{ES}}_{p_n}^{(2)}$, which is a function of the Hill (1975) estimator, he uses a data-dependent choice of the intermediate sequence. We follow Hill's (2015a) recipe in the simulations for $\widehat{\mathrm{ES}}_{p_n}^{(2)}$.

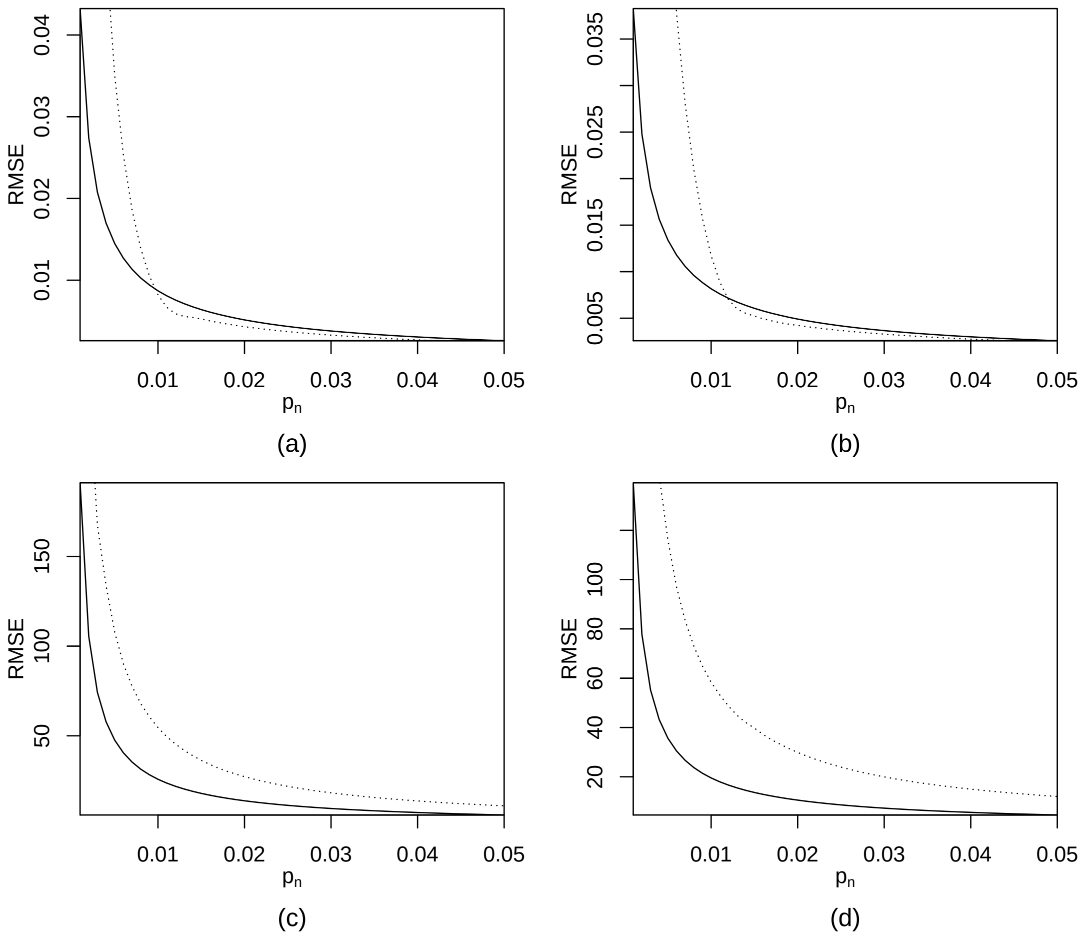

Figure 1: RMSE of $\widehat{\mathrm{CTM}}_1(p_n)$ (solid line) and $\widehat{\mathrm{ES}}^{(2)}_{p_n}$ (dashed line) for AR(1)-GARCH$(1,1)$ model in (a), GARCH$(1,1)$ in (b), i.i.d. draws from the Burr distribution with $\tau = 2$ and $\lambda = 0.75$ in (c), and with $\tau = 3$ and $\lambda = 0.5$ in (d).

For the choice of $k = k_n$ in $\widehat{\mathrm{CTM}}_1(p_n)$ we again take a different tack and modify a data-adaptive algorithm recently proposed by Daníelsson *et al.* (2016). Their method is based on the following considerations. Replacing p_n by j/n in (9), the Pareto-type tail suggests – similarly as before – the following estimate of the $(1 - j/n)$-quantile: $\widehat{x}_{j/n} = (k/j)^{\widehat{\gamma}_k} X_{(k+1)}$. The quality of the Pareto-approximation for this particular choice of k may now be judged by $\sup_{j=1,\ldots,k_{\max}} |X_{(j+1)} - \widehat{x}_{j/n}|$, i.e., a comparison of empirical quantiles and quantiles estimated using the Pareto-approximation. Here, $k_{\max}$ indicates the range over which the fit is assessed. These considerations motivate the choice

$$k^*_{\mathrm{VaR}} = \underset{k_{\min},\ldots,k_{\max}}{\arg\min} \left[\sup_{j=1,\ldots,k_{\max}} \left| X_{(j+1)} - \widehat{x}_{j/n} \right| \right], \tag{14}$$

where $k_{\min}$ is the smallest choice of k one is willing to entertain (see also below). While the choice k^*_{VaR} is well-suited conceptually for quantile estimation and $\widehat{\mathrm{CTM}}_1(p_n)$ is essentially a scaled quantile estimate, it may occasionally happen that $\widehat{\gamma}_{k^*_{\mathrm{VaR}}} \geq 1$, rendering ES estimates $\widehat{\mathrm{CTM}}_1(p_n)$ to be of different sign than quantile estimates.

To avoid such a nonsensical result, we adapt the general idea behind the choice of k^*_{VaR} to our particular task of ES estimation. Instead of assessing the fit of the Pareto-motivated quantile estimates to (nonparametric) empirical quantiles, we now assess the fit of Pareto-motivated ES estimates, $\widehat{\mathrm{CTM}}_1(j/n) = \widehat{x}_{j/n}/(1-\widehat{\gamma}_k)$, to the nonparametric estimates $\widehat{\mathrm{ES}}_{j/n}$ from (1). Then, by analogy, we choose

$$k^*_{\mathrm{ES}} = \operatorname*{arg\,min}_{k_{\min},\ldots,k_{\max}} \left[\sup_{j=1,\ldots,k_{\max}} \left| \widehat{\mathrm{ES}}_{j/n} - \widehat{\mathrm{CTM}}_1(j/n) \right| \right]. \tag{15}$$

With this particular choice, an estimate $\widehat{\gamma}_{k^*_{\mathrm{ES}}} \geq 1$ was always avoided in our simulations. Since the largest level we use is $p_n = 0.05$, the requirement $np_n/k_n = o(1)$ from (11) suggests $k_{\min} = \lfloor 0.05 \cdot n \rfloor$. Furthermore, we use $k_{\max} = \lfloor n^{0.9} \rfloor$. Following Hill (2010), we use the bandwidth $\gamma_n = (k^*_{\mathrm{ES}})^{0.25}$ for $\widehat{\sigma}^2_{k_n}$.

The RMSEs (calculated based on 10,000 replications) for time series of length $n = 2000$ are displayed in Figure 1 for $p_n = 0.001, 0.002, \ldots 0.05$. The RMSEs for the (AR-)GARCH models in panels (a) and (b) are similar.[1] For levels p_n between roughly 0.01 and 0.05, the estimator $\widehat{\mathrm{ES}}^{(2)}_{p_n}$ is slightly more accurate, possibly because the empirical distribution function is sufficiently informative in this range. For smaller p_n-values exploiting the Pareto form of the tails pays off with RMSEs up to 10 times smaller for $p_n = 0.001$. Panels (c) and (d) show the results for i.i.d. draws from the Burr distribution. Here, the Pareto approximation holds quite accurately over a wide range of the support, whence lower RMSEs result for all $p_n = 0.001, \ldots, 0.05$. In (d), where $\tau = 3$, the relative advantage of $\widehat{\mathrm{CTM}}_1(p_n)$ over $\widehat{\mathrm{ES}}^{(2)}_{p_n}$ is larger, as expected due to the better fit to the Pareto approximation when τ is larger.

Figure 1 suggests that for levels $p_n \leq 0.01$ the estimator $\widehat{\mathrm{CTM}}_1(p_n)$ generally is to be preferred. Hence, we investigate coverage of our confidence corridors for the value $p_n = 0.01$ and $t \in [0.1, 1]$, such that all quantiles in the range between 0.001 and 0.01 are covered. Following the suggestion of Daníelsson *et al.* (2016) for the choice of k_n in (14) (and using a bandwidth of $\gamma_n = (k^*_{\mathrm{VaR}})^{0.25}$), for 10,000 replications we have calculated coverage probabilities of the 90%-confidence corridor (13) (where $\beta = 0.1$) for the above processes. For the AR(1)-GARCH(1, 1) model coverage was 71.5%, for the pure GARCH(1, 1) 73.7%, for the Burr distribution with $\tau = 2$ ($\tau = 3$) 84.7% (89.1%). Coverage is somewhat off target for the (AR-)GARCH models. However, in other applications of extreme quantile estimation, *pointwise* confidence intervals have displayed some marked undercoverage on par with the

[1] The true value of the expected shortfall was calculated in all cases as in Hill (2015a, p. 17).

values observed here (e.g., Drees, 2003; Chan *et al.*, 2007). In view of this, the coverage of our *uniform* confidence intervals is rather encouraging.

To shed further light on this, we also investigate pointwise coverage of VaR for $p = 0.01$, also using Theorem 2. In this case, coverage is only slightly better with values 77.4%, 80.9%, 86.9% and 91.1%, respectively. This suggests that much of the estimation uncertainty lies in estimating the smallest quantile ($x_{0.01}$ in this case) and the extrapolation to smaller levels does not significantly affect coverage. We thus conclude that the Pareto tail pins down the actual tail behavior very well, particularly for the Burr distribution.

4 An application to extreme returns of VW shares

In this section we illustrate the use of Theorems 1 and 2 by calculating VaR corridors and ES estimates. We do so for the $n = 3490$ log-losses of the German auto maker VW's stock from March 27, 1995 to October 24, 2008 downloaded from *finance.yahoo.com.* (If P_i denotes the adjusted closing prices, the log-losses are defined as $X_i = \log(P_{i-1}/P_i)$. A similar analysis could of course be carried out for the log-returns $-X_i$.) This period was chosen to precede the tumultuous week of trading in VW shares from October 27, 2008 to October 31, 2008. Preceding this week, the sports car maker Porsche built up a huge position in VW shares in a takeover attempt that ultimately failed. Porsche announced on Sunday – October 26, 2008 – that it had indirect control of 74.1% of VW. Since the German state of Lower Saxony owned another 20.2% of VW, this left short-sellers scrambling to buy the remaining shares to close their positions. The shares closed at €210.85 on Friday, October 24, more than doubling on the next trading day – Monday, October 27 – to €520, and again almost doubling to €945 on Tuesday. During a few minutes of trading on Tuesday, VW was the world's most valuable company. Wednesday then saw the shares almost halve in value, closing at €517.

The magnitude of the log-returns from Monday, Tuesday and Wednesday of 0.904, 0.597 and -0.603, respectively, is very large indeed if compared with previous historical returns, which are displayed in Figure 2. In fact, a log-loss of 0.603 has not been observed before. Thus, one must assess the magnitude of a previously unseen event, which provides a natural application of the extreme value methods proposed in this paper.

To get a better sense of the significance of the log-loss of 0.603 we apply the methodology developed in this paper. Before doing so, we check that Theorems 1 and 2 may reasonably be applied. To this end we fit a standard AR(1)-GARCH(1, 1) model with skewed-t distributed innovations to the time series. Visual inspection and standard Ljung-Box tests of the (raw and squared) standardized residuals reveal that they may reasonably be considered i.i.d. and thus an adequate fit of our model. Under quite general conditions AR(1)-GARCH$(1, 1)$ models are stationary and L_2-E-NED (Hill, 2011, Sec. 4). At this point one may argue that it is sufficient to estimate the model parameters and simulate long sample paths of the estimated model often enough to obtain an estimate of VaR and ES. However, this approach is dangerous in our extreme value setting. Drees (2008) has shown in the context of extreme VaR estimation that even a slight misspecification of the model, that is not detectable by statistical tests, can lead to distorted estimates. Thus, the main point of our model fitting exercise is to show that a stationary time series model (here an AR(1)-GARCH$(1, 1)$ process) provides a good fit to the data at hand.

To the best of our knowledge, the Pareto-type tail assumption (4) has only been verified for the smaller class of AR(1)-ARCH(1) models by Borkovec and Klüppelberg (2001), so it seems worthwhile to check it empirically. To do so, we use the Pareto quantile plot of Beirlant *et al.* (1996). The idea is to use (6), i.e., $U(x) = x^{\gamma} L_U(x)$. Since $\log L_U(x)/\log x \to 0$ as $x \to \infty$ (de Haan and Ferreira, 2006, Prop. B.1.9.1), we obtain $\log U(x) \sim \gamma \log x$. Thus, for small j, the plot of

$$\left(-\log\left(\frac{j}{n+1}\right), \log X_{(j)} \approx \log U((n+1)/j) \right), \qquad j = 1, \dots, n,$$

should be roughly linear with positive slope $\gamma > 0$, if (4) holds with positive extreme value index. Since some log-losses are negative, rendering $\log X_{(j)}$ to be undefined, we only use the positive log-

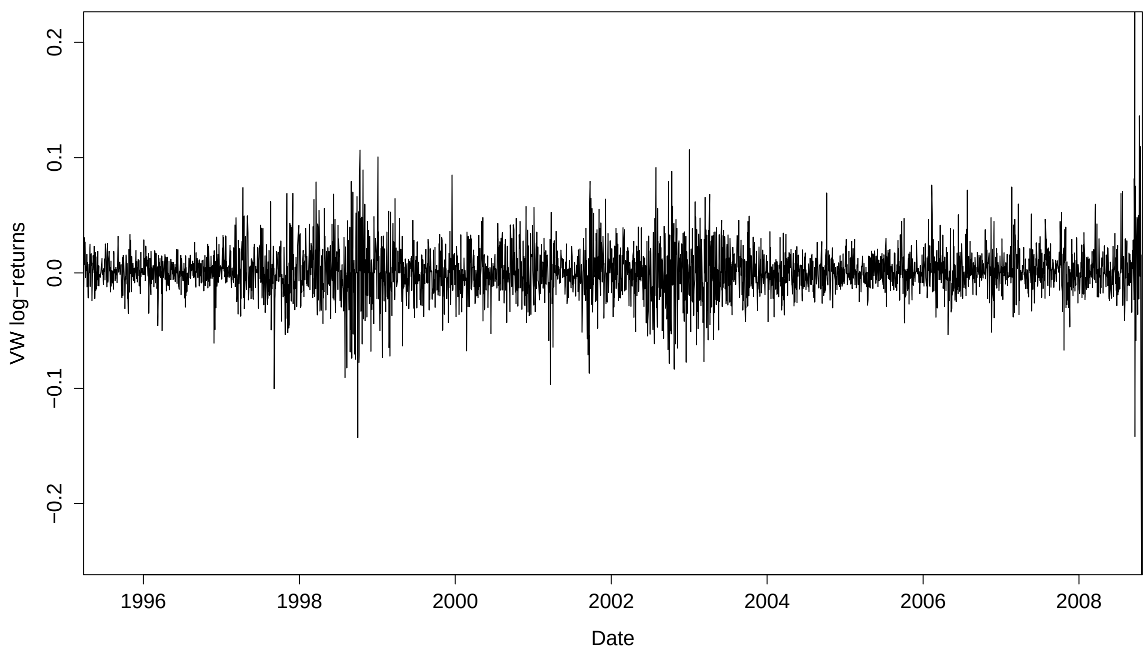

Figure 2: VW log-returns from March 27, 1995 to October 24, 2008

losses for the Pareto quantile plot in panel (a) in Figure 3. A roughly linear behavior with positive slope can be discerned from $-\log(j/(n+1)) = 2$ onwards, but it is not quite satisfactory, as the Hill plot of $k_n \mapsto \widehat{\gamma}_{k_n}$ in panel (b) is highly unstable. A better approximation to linearity in the Pareto quantile plot and more stable Hill estimates can often be obtained by a slight shift of the data. Here, a positive shift of 0.05 sufficed, as the plots in (c) and (d) for the shifted data reveal. The positive slope of the roughly linear portion in the Pareto quantile plot and the strictly positive and very stable Hill estimates for k_n up to 1000 strongly suggest a Pareto-type tail with positive tail index for the VW log-losses. From the stable portion of the Hill plot in panel (d) we read off an estimate of the extreme value index of $\widehat{\gamma} = 0.2$. The 95%-confidence intervals for γ for different values of k_n are indicated by the shaded area in panel (d). They were computed using Hill (2010, Thm. 2) and $\widehat{\sigma}_{k_n}$; see also Equation (A.1) in the Appendix. The null hypothesis $\gamma = 1$, which would invalidate our analysis for ES, is clearly rejected for k_n. All in all, we are confident that Theorems 1 and 2 can be applied.

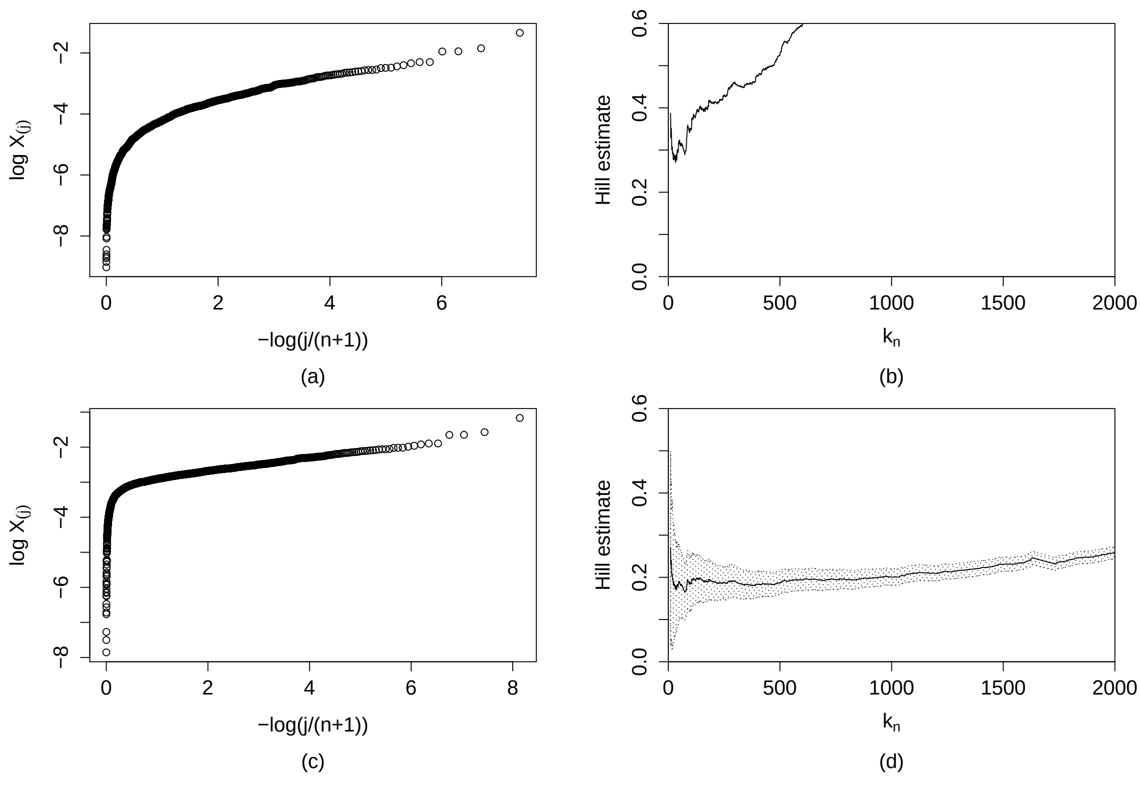

Figure 3: Pareto quantile plot and Hill plot for raw log-losses (in (a) and (b)) and for log-losses shifted by 0.05 (in (c) and (d)). The shaded area around the Hill estimates in panel (d) signifies 95%-confidence intervals.

Figure 4 displays the results, i.e., the VaR and ES estimates for levels between $p_n = 0.05$ and 0.0001. In view of the much more stable Hill estimates (upon which our VaR and ES estimators are based) for the shifted data in Figure 3, we carry out VaR and ES calculations for the shifted data and then subtract 0.05 from the results to arrive at estimates for the original series of log-losses. Because choosing k_n according to (15) ensures $\widehat{\gamma} < 1$, we use $k^*_{\text{ES}} = 1060$ to compute VaR and ES estimates. Incidentally, from the Hill plot in panel (d) of Figure 3 the use of k_n around a similar value of around 1000 seems sensible, because smaller values of k_n lead to roughly the same estimate (yet a slower rate) and for larger values the Hill plot is slightly upward trending, suggesting a possible bias. The choice of $p_n = 0.05$ is compatible with the theory requirement $np_n = o(k_n)$, since $np_n = 3490 \cdot 0.05 = 174.5$ is small relative to $k_n = k^*_{\text{ES}} = 1060$.

In more detail, Figure 4 displays VaR estimates (solid line). As is customary in extreme value theory, the risk level p_n is not plotted directly, but rather the m-year return level; see, e.g., Coles (2001, Sec. 4.4.2). Since there are approximately 250 trading days in a year, a probability of $p_n = 1/250$ corresponds to a return period of 1 year. Thus, the return level with return period of 1 year is, on average, only exceeded once a year. Similarly, the 2-year return period corresponds to $p_n = 1/500$, and so forth. As is also customary, we plot the return period on a log-scale to zoom in on the very large return periods that are of particular interest in risk management. The estimated and empirical data (calculated simply as $X_{(\lfloor np_n \rfloor+1)}$) are in reasonable agreement, strengthening further the belief that our methods are appropriate.

Most empirical estimates lie within the 95%-confidence corridor for VaR at different levels (grey area in Figure 4) calculated from Theorem 2. It has the interpretation that the null hypothesis that the true $x_{p_n}(t)$ lies in this grey area (for $t = [0.002, 1]$ and $p_n = 0.05$) cannot be rejected at the 5% level. In this sense, it provides an informative description of the tail region.

The dashed line in Figure 4 indicates ES estimates. As the expected loss given a VaR exceedance, the ES estimates provide further insight on the tail behavior. All in all, nothing in Figure 4 suggests that a log-loss of 0.603 was to be expected. Even ES estimates for a return period of 40 years do not come close to this value. Of course, further extrapolation of VaR and ES estimates in Figure 4 would be possible to see for which return period a return level of 0.603 is obtained. However, in view of the restriction on p_n imposed by (11) (see also Remark 2) and related applications of extreme value theory (Drees, 2003), we feel that extrapolation well beyond a level of $p_n = 0.0001 \approx 1/(2.87 \cdot n)$ is no longer justified.

5 Summary

Our first main contribution is to derive central limit theory for a wide range of popular risk measures, including VaR and ES, in time series. As in Linton and Xiao (2013) and Hill (2015a), we do so under a Pareto-type tail assumption. Yet, we exploit the Pareto approximation to motivate an estimator

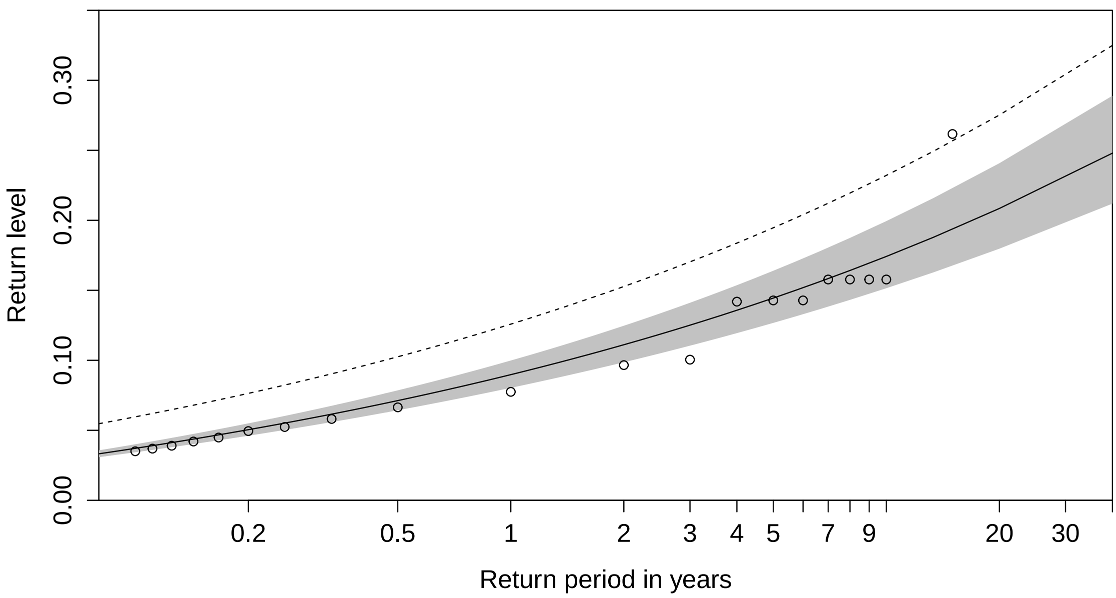

Figure 4: Return level plot for VW log-losses (solid line). Grey area indicates 95%-confidence corridor for return levels. ES estimates shown as the dashed line.

of (among other risk measures) ES, whereas Linton and Xiao (2013) consider a non-parametric ES estimator and Hill (2015a) only uses the Pareto assumption for bias correction of his tail-trimmed ES estimator. Asymptotic theory is derived under an E-NED property, which is significantly more general than the geometrically α-mixing assumption of Linton and Xiao (2013) and Hill (2015a). It is shown in simulations that our estimator (which fully takes into account the regularly varying tail) provides better estimates in terms of RMSE than Hill's (2015a) proposal (which only does so partially). Our second main contribution is to derive uniform confidence corridors for VaR and also the other risk measures covered by our analysis. Furthermore, we propose a method for choosing the sample fraction k_n used in the estimation of ES, which is used in the simulations. Finally, we illustrate our procedure with VW log-losses prior to the takeover attempt by Porsche.

References

Andrews D. 1984. Non-strong mixing autoregressive processes. *Journal of Applied Probability* **21**: 930–934.

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

Beirlant J, Vynckier P, Teugels J. 1996. Tail index estimation, Pareto quantile plots, and regression diagnostics. *Journal of the American Statistical Association* **91**: 1659–1667.

Borkovec M, Klüppelberg C. 2001. The tail of the stationary distribution of an autoregressive process with ARCH(1) errors. *The Annals of Applied Probability* **11**: 1220–1241.

Bücher A, Jäschke S, Wied D. 2015. Nonparametric tests for constant tail dependence with an application to energy and finance. *Journal of Econometrics* **187**: 154–168.

Burnecki K, Janczura J, Weron R. 2011. Building loss models. In Čížek P, Härdle W, Weron R (eds.) *Statistical Tools for Finance and Insurance*, Berlin: Springer, 2 edn., pages 363–370.

Chan N, 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.

Chen S. 2008. Nonparametric estimation of expected shortfall. *Journal of Financial Econometrics* **6**: 87–107.

Coles S. 2001. *An Introduction to Statistical Modeling of Extreme Values*. London: Springer.

Daníelsson J. 2011. *Financial Risk Forecasting*. Chichester: Wiley.

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

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

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

Drees H. 2008. Some aspects of extreme value statistics under serial dependence. *Extremes* **11**: 35–53.

Einmahl J, de Haan L, Zhou C. 2016. Statistics of heteroscedastic extremes. *Journal of the Royal Statistical Society: Series B* **78**: 31–51.

El Methni J, Gardes L, Girard S. 2014. Non-parametric estimation of extreme risk measures from conditional heavy-tailed distributions. *Scandinavian Journal of Statistics* **41**: 988–1012.

Engle R, Bollerslev T. 1986. Modelling the persistence of conditional variances. *Econometric Reviews* **5**: 1–50.

Fasen V, Klüppelberg C, Schlather M. 2010. High-level dependence in time series models. *Extremes* **13**: 1–33.

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.

Gomes M, 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 J. 2009. On functional central limit theorems for dependent, heterogeneous arrays with applications to tail index and tail dependence estimation. *Journal of Statistical Planning and Inference* **139**: 2091–2110.

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

Hill J. 2011. Tail and nontail memory with applications to extreme value and robust statistics. *Econometric Theory* **27**: 844–884.

Hill J. 2013. Least tail-trimmed squares for infinite variance autoregressions. *Journal of Time Series Analysis* **34**: 168–186.

Hill J. 2015a. Expected shortfall estimation and Gaussian inference for infinite variance time series. *Journal of Financial Econometrics* **13**: 1–44.

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

Hoga Y. 2017+a. Change point tests for the tail index of β-mixing random variables. Forthcoming in *Econometric Theory* (doi: 10.1017/S0266466616000189) : 1–40.

Hoga Y. 2017+b. Testing for changes in (extreme) VaR. Forthcoming in *Econometrics Journal* (doi: 10.1111/ectj.12080) : 1–29.

Hoga Y, Wied D. 2017. Sequential monitoring of the tail behavior of dependent data. *Journal of Statistical Planning and Inference* **182**: 29–49.

Hong J, Elshahat A. 2010. Conditional tail variance and conditional tail skewness. *Journal of Financial and Economic Practice* **10**: 147–156.

Hsing T. 1991. On tail index estimation using dependent data. *The Annals of Statistics* **19**: 1547–1569.

Ibragimov R, Jaffee D, Walden J. 2009. Non-diversification traps in markets for catastrophic risk. *Review of Financial Studies* **22**: 959–993.

Ibragimov R, Walden J. 2011. Value at risk and efficiency under dependence and heavy-tailedness: Models with common shocks. *Annals of Finance* **7**: 285–318.

Ling S. 2007. Self-weighted and local quasi-maximum likelihood estimators for ARMA-GARCH/IGARCH models. *Journal of Econometrics* **140**: 849–873.

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

Pan X, Leng X, Hu T. 2013. The second-order version of Karamata's theorem with applications. *Statistics & Probability Letters* **83**: 1397–1403.

Resnick S. 2007. *Heavy-Tail Phenomena: Probabilistic and Statistical Modeling.* New York: Springer.

Scaillet O. 2004. Nonparametric estimation and sensitivity analysis of expected shortfall. *Mathematical Finance* **14**: 115–129.

Shao X, Zhang X. 2010. Testing for change points in time series. *Journal of the American Statistical Association* **105**: 1228–1240.

Valdez E. 2005. Tail conditional variance for elliptically contoured distributions. *Belgian Actuarial Bulletin* **5**: 26–36.

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

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

Appendix

Proof of Theorem 1: From Hill (2010, Thm. 2) we get

$$\frac{\sqrt{k_n}}{\sigma_{k_n}}\left(\widehat{\gamma}-\gamma\right) \underset{(n\to\infty)}{\overset{\mathcal{D}}{\longrightarrow}} \mathcal{N}(0,1). \tag{A.1}$$

Note that Hill's (2010) Assumption B (required in his Thm. 2) can be seen to be implied by Assumption 1. Concretely, write (7) in terms of the slowly varying function $L(\cdot)$ from (5) to obtain

$$\lim_{x\to\infty} \frac{\frac{L(\lambda x)}{L(x)} - 1}{A(x)} = \frac{\lambda^{\rho/\gamma} - 1}{\gamma\rho},$$

where $A(\cdot)$ is a function with bounded increase due to $A(\cdot) \in RV_{\rho/\gamma}$ for $\rho/\gamma < 0$ (de Haan and Ferreira, 2006, Thm. B.3.1). Also note that $\liminf_{n\to\infty} \sigma_{k_n} > 0$ by arguments in Hill (2010, Sec. 3.2).

Hence, from (A.1) and arguments in the proof of de Haan and Ferreira (2006, Thm. 4.3.9), we get

$$\frac{1}{\sigma_{k_n}} \frac{\sqrt{k_n}}{\log d_n} \left(\frac{\widehat{x}_{p_n}}{x_{p_n}} - 1 \right) \underset{(n\to\infty)}{\xrightarrow{\mathcal{D}}} \mathcal{N}(0,1). \tag{A.2}$$

Here we have also used that

$$\sqrt{k_n} \left(\frac{X_{(k_n+1)}}{U(n/k_n)} - 1 \right) = O_P(1)$$

from Hill (2010, Lem. 3) and the fact that $\log(x) \sim x - 1$, as $x \to 1$. Next we show that

$$\frac{\sqrt{k_n}}{\log d_n} \left(\frac{\widehat{\mathrm{CTM}}_a(p_n)}{\mathrm{CTM}_a(p_n)} - 1 \right) = \frac{\sqrt{k_n}}{\log d_n} \left(\frac{\widehat{x}^a_{p_n}}{x^a_{p_n}} - 1 \right) + o_P(1). \tag{A.3}$$

To do so expand

$$\frac{\sqrt{k_n}}{\log d_n} \left(\frac{\widehat{\mathrm{CTM}}_a(p_n)}{\mathrm{CTM}_a(p_n)} - 1 \right) = \frac{\sqrt{k_n}}{\log d_n} \left(\frac{\widehat{x}^a_{p_n}}{x^a_{p_n}} \cdot \frac{1-a\gamma}{1-a\widehat{\gamma}} \cdot \frac{\frac{x^a_{p_n}}{1-a\gamma}}{\mathrm{CTM}_a(p_n)} - 1 \right) \tag{A.4}$$

By (A.1),

$$\frac{1-a\gamma}{1-a\widehat{\gamma}} = 1 + \mathcal{O}_P(1/\sqrt{k_n}). \tag{A.5}$$

From Pan *et al.* (2013, Thm. 4.2),

$$\lim_{n\to\infty} \frac{1}{A\left(U(1/p_n)\right)} \left(\frac{\mathrm{CTM}_a(p_n)}{x^a_{p_n}} - \frac{1}{1-a\gamma} \right) = \frac{a}{(1/\gamma - a)(1/\gamma - a - \rho)}.$$

Since $U(n/k_n) = \mathcal{O}\left(U(1/p_n)\right)$ (due to $np_n = o(k_n)$ from (11) and monotonicity of $U(\cdot)$), we have

$$A\left(U(1/p_n)\right) = \mathcal{O}\left(A\left(U(n/k_n)\right)\right) = o(1/\sqrt{k_n}),$$

implying together that

$$\frac{\mathrm{CTM}_a(p_n)}{\frac{x^a_{p_n}}{1-a\gamma}} - 1 = o\left(\frac{1}{\sqrt{k_n}}\right). \tag{A.6}$$

Combining (A.4) – (A.6), (A.3) follows.

In view of (A.3) and $\left|\widehat{\sigma}^2_{k_n} - \sigma^2_{k_n}\right| = o_P(1)$ (Hill, 2010, Thm. 3), it suffices to prove the claim of the theorem for the sequence of random vectors

$$\frac{1}{\sigma_{k_n}} \frac{\sqrt{k_n}}{\log d_n} \left[\left(\frac{\widehat{x}_{p_n}^{a_j}}{x_{p_n}^{a_j}} - 1 \right)_{j=1,\dots,J}, \left(\frac{\widehat{x}_{p_n}}{x_{p_n}} - 1 \right) \right]'.$$

Let $b_1, \dots, b_{J+1} \in \mathbb{R}$. Then, using a Cramér-Wold device, it suffices to consider

$$\frac{1}{\sigma_{k_n}} \frac{\sqrt{k_n}}{\log d_n} \sum_{j=1}^{J+1} b_j \left(\frac{\widehat{x}_{p_n}^{a_j}}{x_{p_n}^{a_j}} - 1 \right).$$

(Recall $a_{J+1} = 1$.) Invoking a Skorohod construction (e.g., de Haan and Ferreira, 2006, Thm. A.0.1) similarly as in de Haan and Ferreira (2006, Example A.0.3), we may assume that the convergence in (A.2) holds almost surely (a.s.) on a different probability space:

$$\frac{1}{\sigma_{k_n}} \frac{\sqrt{k_n}}{\log d_n} \left(\frac{\widehat{x}_{p_n}}{x_{p_n}} - 1 \right) \underset{(n\to\infty)}{\overset{\text{a.s.}}{\longrightarrow}} Z \sim \mathcal{N}(0,1).$$

(Note the slight abuse of notation here.) A Taylor expansion of the functions $f_j(x) = x^{a_j}$ around 1 thus implies

$$\frac{1}{\sigma_{k_n}} \frac{\sqrt{k_n}}{\log d_n} \sum_{j=1}^{J+1} b_j \left(\frac{\widehat{x}_{p_n}^{a_j}}{x_{p_n}^{a_j}} - 1 \right) \underset{(n\to\infty)}{\overset{\text{a.s.}}{\longrightarrow}} \sum_{j=1}^{J+1} b_j a_j Z.$$

Going back to the original probability space, the conclusion follows. ■

Proof of Theorem 2: Since $\log(1+x) \sim x$ as $x \to 0$, it suffices to show

$$\sup_{t \in [\underline{t}, \overline{t}]} \left| \frac{1}{\widehat{\sigma}_{k_n}} \frac{\sqrt{k_n}}{\log d_n(t)} \left(\frac{\widehat{x}_{p_n}(t)}{x_{p_n}(t)} - 1 \right) \right| \underset{(n\to\infty)}{\overset{\mathcal{D}}{\longrightarrow}} |Z|.$$

Due to $\widehat{x}_{p_n}(t) = \widehat{x}_{p_n} t^{-\widehat{\gamma}}$ and $\log d_n(t)/\log d_n = 1 + o(1)$ uniformly in $t \in [\underline{t}, \overline{t}]$, we can expand

$$\frac{\sqrt{k_n}}{\log d_n(t)} \left(\frac{\widehat{x}_{p_n}(t)}{x_{p_n}(t)} - 1 \right) = (1+o(1)) \frac{\sqrt{k_n}}{\log d_n} \left(\frac{\widehat{x}_{p_n}}{x_{p_n}} t^{\gamma - \widehat{\gamma}} \frac{x_{p_n}}{x_{p_n}(t)} t^{-\gamma} - 1 \right). \tag{A.7}$$

Apply the mean value theorem with $(\partial/\partial x) t^x = t^x \log(t)$ to derive $t^{\gamma-\widehat{\gamma}} = 1 + (\gamma - \widehat{\gamma}) t^{\nu(\gamma-\widehat{\gamma})}$ for some $\nu \in [0,1]$. Since $\gamma - \widehat{\gamma} = \mathcal{O}_P(1/\sqrt{k_n})$, this implies

$$t^{\gamma-\widehat{\gamma}} = 1 + \mathcal{O}_P(1/\sqrt{k_n}) \qquad \text{uniformly in } t \in [\underline{t}, \overline{t}]. \tag{A.8}$$

Writing (7) in terms of the quantile function $U(\cdot)$, we obtain from de Haan and Ferreira (2006, Thm. 2.3.9) that uniformly in $t \in [\underline{t}, \overline{t}]$,

$$\left| \frac{x_{p_n}(t)}{x_{p_n}} - t^{-\gamma} \right| = \left| \frac{U\big(1/(p_n t)\big)}{U\big(1/p_n\big)} - t^{-\gamma} \right| = \mathcal{O}\big(A(U(1/p_n))\big) = \mathcal{O}\big(A(U(n/k_n))\big) = o(1/\sqrt{k_n}). \tag{A.9}$$

Here we have also used that $n/k_n = o(1/p_n)$ by (11). Combining (A.7) with (A.8) and (A.9) gives

$$\frac{\sqrt{k_n}}{\log d_n(t)} \left(\frac{\widehat{x}_{p_n}(t)}{x_{p_n}(t)} - 1 \right) = \frac{\sqrt{k_n}}{\log d_n(t)} \left(\frac{\widehat{x}_{p_n}}{x_{p_n}} - 1 \right) + o_P(1) \qquad \text{uniformly in } t \in [\underline{t}, \overline{t}].$$

The conclusion now follows from Theorem 1. ■