When is Tail Mean Estimation More Efficient Than Tail Median? Answers and Implications for Quantitative Risk Management

Roger W. Barnard* Kent Pearce† A. Alexandre Trindade‡

May 16, 2017

Abstract

We investigate the relative efficiency of the empirical “tail median” vs. ”tail mean” as estimators of location when the data can be modeled by an exponential power distribution (EPD), a flexible family of light-tailed densities. By considering appropriate probabilities so that the quantile of the untruncated EPD (tail median) and mean of the left-truncated EPD (tail mean) coincide, limiting results are established concerning the ratio of asymptotic variances of the corresponding estimators. The most remarkable finding is that in the limit of the right tail, the asymptotic variance of the tail median estimate is approximately 36% larger than that of the tail mean, irrespective of the EPD shape parameter. This discovery has important repercussions for quantitative risk management practice, where the tail median and tail mean correspond to *value-at-risk* and *expected shortfall*, respectively. To this effect, a methodology for choosing between the two risk measures that maximizes the precision of the estimate is proposed. From an extreme value theory perspective, analogous results and procedures are discussed also for the case when the data appear to be heavy-tailed.

Keywords: value-at-risk; expected shortfall; generalized quantile function; exponential power distribution; extreme value theory; asymptotic relative efficiency; rate of convergence.

1 Introduction

The quantification of *risk* is an increasingly important exercise carried out by risk management professionals in a variety of disciplines. These may range from assessing the likelihood of structural failure in machined parts, and catastrophic floods in hydrology and storm management, to predicting economic loss in insurance, portfolio management, and credit lending companies. In the area of (financial) quantitative risk management (QRM), increasing demand by regulatory bodies such as the Basel Committee on Banking Supervision[1] to implement methodically sound credit risk assessment practices, has fostered much recent research into measures of risk. See McNeil et al. (2005) for a classic treatment of QRM, and Embrechts and Hofert (2014) and Embrechts et al. (2014) for recent surveys on QRM and risk measures relating to the Basel documents, respectively.

As the term *risk* is synonymous with *extreme event*, it is naturally desirable to determine or estimate high percentiles of the (often potential) distribution of losses. To be useful in quick

*Texas Tech University, Department of Mathematics & Statistics, Lubbock, U.S.A.

†Texas Tech University, Department of Mathematics & Statistics, Lubbock, U.S.A.

‡Texas Tech University, Department of Mathematics & Statistics, Lubbock, TX 79409-1042, U.S.A. Phone: 806-834-6164; Fax: 806-742-1112; E-mail: alex.trindade@ttu.edu. (Corresponding Author.)

[1][http://www.bis.org/bcbs/]: A forum for regular cooperation on worldwide banking supervisory matters based in Basel (Switzerland).

decision making, a single number is required, and two common measures in use by credit lenders include *Value-at-Risk* (VaR), and *Conditional Value-at-Risk* (CVaR), along with percentiles typically in excess of the 95th (Basel III, 2013). The term CVaR was coined by Rockafellar and Uryasev (2000), but synonyms (and closely related variants) for it also in common usage include: "expected shortfall" (Acerbi and Tasche, 2002), "tail-conditional expectation" (Artzner et al., 1999), and "average Value-at-Risk" (Chun et al., 2012), or simply "tail mean". Since the term *expected shortfall* (ES) appears to have become the most prevalent synonym for CVaR, we will adhere to the term ES.

VaR and ES are widely used to measure and manage risk in the financial industry (Jorion, 2003; Duffie and Singleton, 2003). With the random variable X representing the distribution of losses, given a probability level $0 < \beta < 1$, VaR_β is simply the β quantile of X, and hence is the value beyond which higher losses only occur with probability $1 - \beta$. Using similar notation, ES_β is the average of this $1 - \beta$ fraction of worst losses. To define these precisely, let X be a continuous real-valued random variable defined on some probability space $(\Omega, \mathcal{A}, P)$, with cumulative distribution function (CDF) $F(\cdot)$ and probability density function (PDF) $f(\cdot)$. The quantities μ and σ^2 will denote, respectively, the mean and variance of X, and both are assumed to be finite. The VaR and ES of X at probability level β are then defined as follows.

Definition 1 (VaR al level β)**.**

$$\mathrm{VaR}_\beta(X) \equiv \xi_\beta(X) = F^{-1}(\beta). \tag{1}$$

Definition 2 (ES at level β)**.**

$$\mathrm{ES}_\beta(X) \equiv \mu_\beta(X) = \mathbb{E}(X|X \geq \xi_\beta) = \frac{1}{1-\beta}\int_{\xi_\beta}^{\infty} x f(x)dx = \frac{1}{1-\beta}\int_{\beta}^{1} F^{-1}(u)du. \tag{2}$$

When no ambiguity arises we write simply ξ_β and μ_β. Note that although the quantile function $F^{-1}(\cdot)$ is well-defined here due to the assumed strict monotonicity on $F(\cdot)$, in general ξ_β has to be defined through the *generalized quantile function* $F^{-}(\cdot)$ (Embrechts and Hofert, 2013). In addition, the definition of μ_β implicitly assumes the existence of the first absolute moment, $\mathbb{E}|X|$.

There has been much debate and research over the past decade over which of these two measures should be the industry standard[2]. Some of the issues at stake concern trade-offs between conceptual simplicity (VaR), versus better axiomatic adherence and mathematical properties (ES), since these measures are often used in intricate optimization schemes; see e.g., Follmer and Schied (2011), and Pflug and Romisch (2007). While there is a general consensus that ES is more easily optimized (primarily due to convexity), Yamai and Yoshiba (2002, 2005) report that "...expected shortfall needs a larger size of sample than VaR for the same level of accuracy.". This basic fact is echoed in the review paper of Embrechts and Hofert (2014) when comparing confidence limits for VaR vs. ES based on extreme value theory (EVT). This an expected consequence of their definition that ensures $\xi_\beta \leq \mu_\beta$, and as such we would point out that it is not an entirely fair comparison since for a given β the two measures are in effect measuring different parts of the tail of X.

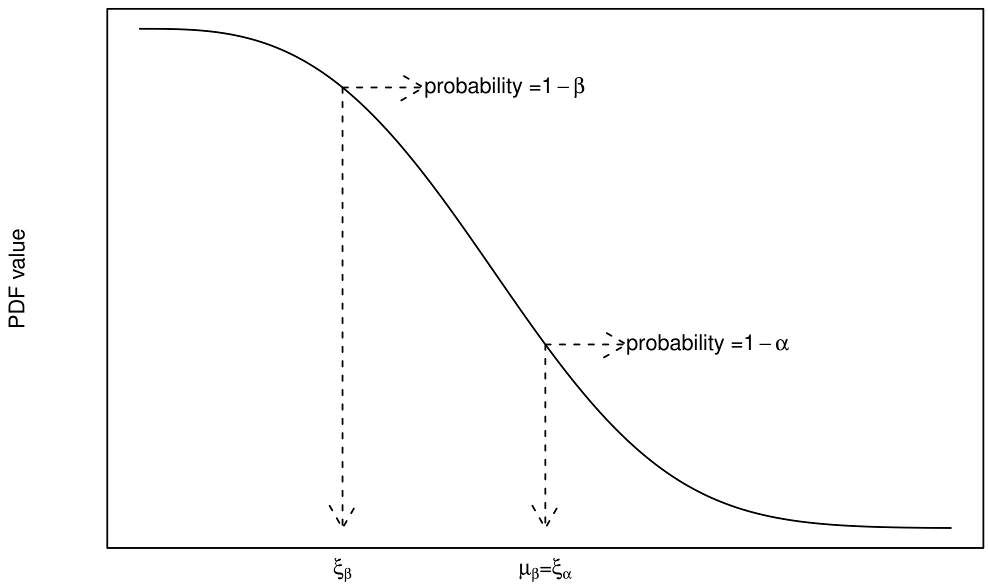
A fairer comparison of the estimation "accuracy" of VaR versus ES would result if the two measures were forced to coincide, a study that to the best of our knowledge has not yet been undertaken. This situation is depicted in Figure 1, which shows the relative positions of the quantiles ξ_β and ξ_α for a PDF with $\beta < \alpha$ so that $\mu_\beta = \xi_\alpha$. This implicitly assumes the existence of a function $g(\cdot)$, such that

$$\alpha = g(\beta) \equiv g_\beta. \tag{3}$$

[2]See for example [http://gloria-mundi.com/], a website serving as a resource for Value-at-Risk and more generally financial risk management.

Loosely speaking, we then have the "tail mean" (μ_β) coinciding with the "tail median" (ξ_α), where the usage of these terms is meant to convey that μ_β is a "mean-like" quantity and ξ_α a "median-like" quantity. Note that we are effectively adjusting the probability levels so that the mean of the left-truncated PDF at ξ_β coincides with the untruncated α-quantile.

Figure 1: Illustration of relative positions of $VaR_\beta = \xi_\beta$ and $VaR_\alpha = \xi_\alpha$ in the right tail of a PDF, so that $ES_\beta = \mu_\beta$ coincides with VaR_α.



The primary aim of this paper is to shed light on this matter, by considering the asymptotic relative efficiencies of the empirical (or nonparametric) estimators. When a random sample of size n with order statistics $X_{(1)} \leq \cdots \leq X_{(n)}$ is available, consistent estimators of VaR and ES are respectively,

$$\hat{\xi}_\beta = X_{(k_\beta)}, \qquad \text{and} \qquad \hat{\mu}_\beta = \frac{1}{n - k_\beta + 1} \sum_{r=k_\beta}^{n} X_{(r)}, \tag{4}$$

where $k_\beta = [n\beta]$ can denote either of the two integers closest to $n\beta$ (or any interpolant thereof). In the process we will also generalize the age-old statistical quandary of which of the sample median or sample mean is the "best" estimator of centrality, since these estimators are obtained in the limit as $\beta \to 0$ in (4). Nowadays the accepted solution to this quandary is to use either an *M-estimator* or an *L-statistic*, the key idea being to combine elements of the efficiency of the mean with the robustness of the median (Maronna, 2011). However, these solutions involve subjective choices, and are therefore not as "clean" as the simple mean or median.

To effect this comparison, we will consider sampling from the *exponential power distribution* (EPD); a flexible model for distributions with exponentially declining tails. The EPD is also variously called Subbotin, Generalized Error Distribution (Mineo and Ruggieri, 2005), and Generalized Normal Distribution (Nadarajah, 2005), with slight differences in the parametrizations.

The version we adopt here is similar to the EPD of Gomez et al. (1998), but following the simpler parametrization of Sherman (1997). This means the standard member of the family has mean and median zero, with a PDF[3] given by

$$f(x;p) = \frac{p}{2\Gamma(1/p)} \exp\{-|x|^p\}, \qquad p \in (0,\infty). \tag{5}$$

Remark 1. *As we shall see later, there is no loss of generality by restricting attention to the standard member of the EPD family. This is because both VaR and ES satisfy the properties of* translation invariance *and* positive homogeneity *(Pflug, 2007), whereby, for X the standard member, $a \in \mathbb{R}$ and $b > 0$ with $Y = a + bX$, we have that $\xi_\beta(Y) = a + b\xi_\beta(X)$ and $\mu_\beta(Y) = a + b\mu_\beta(X)$. These properties transfer also to the empirical estimators in (4), since they are L-statistics. From this it follows that the ratio of their variances, they key metric we will be looking at through equation (20), is invariant to location-scale shifts.*

The EPD has finite moments of all orders, and is hence "light-tailed" in EVT parlance. Parameter p controls the shape, so that we obtain for $p = 2$ a normal with variance $1/2$, and for $p = 1$ a classical Laplace with variance 2. For $p < 2$ and $p > 2$ we obtain respectively, leptokurtic (heavier than Gaussian) and platikurtic (lighter than Gaussian) tail characteristics. In the limit as $p \to \infty$ the EPD becomes $\mathcal{U}[-1,1]$, a uniform distribution on $[-1,1]$. As $p \to 0$ the limit is degenerate, as the PDF converges to zero everywhere on the real line.

The non-Gaussian members of the *elliptical* family of distributions (which includes EPD) have recently been investigated by Landsman and Valdez (2003) as providing, from a QRM perspective, a more realistic model than the Gaussian. However, the elliptical family covers a wide range of distributions all the way from light-tailed (e.g., EPD) to heavy-tailed (e.g., Student t), and thus have to be used with caution if appropriate moments are to exist. Rather, an EVT-based approach has recently guided much recent research with regard to the type of VaR vs. ES comparisons we are proposing. Recall that a distribution with CDF $F(x)$ is said to be *heavy-tailed*[4] if

$$\lim_{t\to\infty} \frac{1 - F(tx)}{1 - F(t)} = x^{-\theta}, \tag{6}$$

for some $\theta > 0$ known as the *tail index*. (Light-tailed distributions correspond to $\theta = 0$.) From the Basel III (2013) accord, it has become apparent that for light-tailed distributions, $\mathrm{VaR}_{0.99} \approx \mathrm{ES}_{0.975}$, a fact which can be generalized to $\mathrm{VaR}_{1-s} \approx \mathrm{ES}_{1-es}$, for small s (Danielsson and Zhou, 2016). This suggests that for high quantiles from light-tailed distributions, an approximation to the function in (3) is

$$\alpha = g(\beta) \approx \frac{e - 1 + \beta}{e}, \qquad \text{for large } \beta. \tag{7}$$

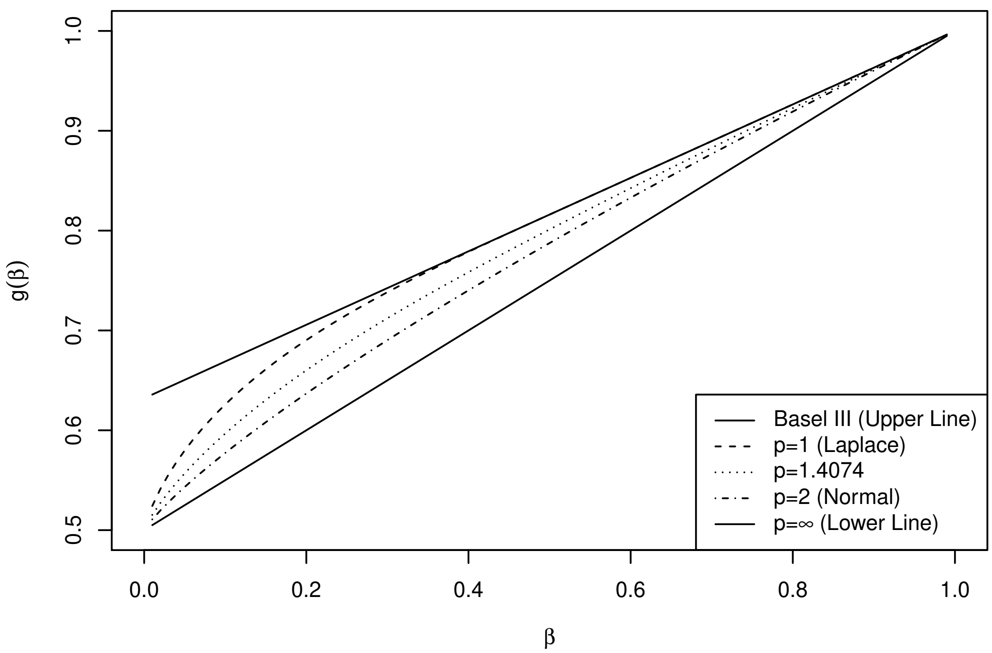
An illuminating plot of β vs. $g(\beta)$ is displayed in Figure 2, where $g(\beta) \equiv g(\beta, p)$ is given by (16) for the EPD. Naturally, we have $\lim_{\beta\to 0} g(\beta, p) = 0.5$ for all p, so that the mean (μ_0) coincides with the median ($\xi_{0.5}$) for the symmetric EPD family. The upper solid line in the plot corresponds to the Basel III function (7), whereas the lower solid line represents the limiting $\mathcal{U}[-1,1]$ distribution when $p = \infty$. As expected, $g(\beta, p)$ converges to the Basel III line as $\beta \to 1$.

The basic questions we are posing concerning the relative efficiency of estimators of VaR vs. ES, or tail median vs. tail mean, are interesting also from a purely academic point of view, and to statistical inference in general. As will be seen later, the central message of this paper is

[3]The PDF, CDF, quantiles, and random values from the standard EPD in (5) can be obtained with the "PE2" function from the R package `gamlss.dist`.

[4]When discussing tail characteristics, we will reserve the term "heavy" for the EVT sense according to (6) when $\theta > 0$, and "leptokurtic/platikurtic" to distinguish between grades of heaviness in relation to the Gaussian among the light-tailed distributions ($\theta = 0$).

Figure 2: Plot of the $g(\beta, p)$ function defined in (16) vs. β, for select values of p. The upper solid line is the Basel III $g(\beta)$ function defined in (7), and the lower solid line the limiting uniform distribution when $p = \infty$.



that for sufficiently "light" tails the tail mean is more efficient than the tail median, when the population quantiles that they are estimating coincide. This appears to remain true whether one is in a light-tailed ($\theta = 0$) or heavy-tailed ($\theta > 0$) regime. To obtain more details about the relative efficiency, it seems inevitable that one has to dive into a particular parametric family. The EPD was therefore chosen for reasons of tractability and flexibility in modeling the light-tailed portion of this spectrum of tail characteristics. To replicate this study for the heavy-tailed portion would at present appear to pose serious analytical challenges.

The rest of the paper is organized as follows. Notation and necessary preliminary results are established in Section 2. This is followed in Section 3 by statements of the main theorems and perspectives on their meaning. A discussion concerning the implications of these results for QRM practice is presented in Section 4. The proofs of the theorems appear in Appendix Sections C and D.

2 Preliminary Results

In this section we derive some preliminary results that will be useful in subsequent sections. Many results will rely on variants of the Gamma function. Let $\gamma(a, x)$ and $\Gamma(a, x)$ denote the incomplete Gamma functions defined by

$$\gamma(a, x) = \int_0^x e^{-t} t^{a-1} dt, \qquad \text{and} \qquad \Gamma(a, x) = \int_x^\infty e^{-t} t^{a-1} dt, \tag{8}$$

and note that $\gamma(a,x)+\Gamma(a,x)=\Gamma(a)$, the usual (complete) Gamma function. The properties of these functions are extensively documented in e.g., Abramowitz and Stegun (1972).

For notational convenience, define for $f(x)$ the EPD PDF in (5) and $n=0,1,2,\ldots$,

$$F_n(x)=\int_{-\infty}^{x} t^n f(t)dt, \qquad G_n(x)=\int_{x}^{\infty} t^n f(t)dt, \qquad A_n=\int_{-\infty}^{\infty} t^n f(t)dt. \tag{9}$$

Using this notation, the EPD CDF satisfies $F\equiv F_0$, and consequently $A_0=1$. We note that for all x we have

$$F_n(x)+G_n(x)=A_n \tag{10}$$

Now, for n odd or even, we can, by making the change of variable $t^p=u$, use (8) to rewrite G_n as

$$G_n(x)=\frac{1}{2\Gamma(1/p)}\Gamma\left(\frac{n+1}{p},x^p\right), \qquad \text{for } x>0, \tag{11}$$

and note that $G_n(0)=[2\Gamma(1/p)]^{-1}\Gamma((n+1)/p)$. For the case that n is odd, the integrand in (9) is odd, and hence, we have

$$G_n(x)=\int_{x}^{-x} t^n f(t)dt+G_n(-x)=\frac{1}{2\Gamma(1/p)}\Gamma\left(\frac{n+1}{p},|x|^p\right), \qquad \text{for } x<0. \tag{12}$$

For the case that n is even, the integrand in (9) is even, and hence $G_n(0)=\frac{1}{2}A_n$. Now, using (10) and (11) yields

$$F_n(x) = \begin{cases} G_n(-x), & \text{for all } x; \\ A_n-G_n(x)=\frac{\Gamma(\frac{n+1}{p})}{\Gamma(1/p)}-\frac{1}{2\Gamma(1/p)}\Gamma\left(\frac{n+1}{p},x^p\right), & \text{for } x>0; \\ G_n(-x)=\frac{1}{2\Gamma(1/p)}\Gamma\left(\frac{n+1}{p},|x|^p\right), & \text{for } x<0, \end{cases} \tag{13}$$

$$G_n(x) = A_n-F_n(x)=\frac{\Gamma\left(\frac{n+1}{p}\right)}{\Gamma(1/p)}-\frac{1}{2\Gamma(1/p)}\Gamma\left(\frac{n+1}{p},|x|^p\right), \qquad \text{for } x<0. \tag{14}$$

Consequently, and noting that $F_0(\xi_\beta)=\beta$, we can, by making the change of variable $u=F_0(t)$, rewrite μ_β defined in (2) as

$$\mu_\beta=\frac{1}{1-F_0(\xi_\beta)}G_1(\xi_\beta)=\frac{G_1(\xi_\beta)}{G_0(\xi_\beta)}, \tag{15}$$

whence the functional relationship between α and β in (3) can be written in closed form as

$$\alpha=g(\beta,p)=F_0(\mu_\beta)=F_0\left(\frac{G_1(\xi_\beta)}{G_0(\xi_\beta)}\right). \tag{16}$$

Notational expediency will often prompt us to write $g_\beta\equiv g(\beta,p)$.

Denote by $\mathbb{1}_A(x)$ the indicator function for set A, which takes on the value 1 if $x\in A$, and 0 otherwise. Define now the random variable obtained by upper tail truncation of X at ξ_β, $Y_\beta\equiv X \mid X\geq\xi_\beta$, and note that μ_β in (15) is in fact the mean of Y_β, since its PDF is $f_{Y_\beta}(y)=(1-\beta)^{-1}f(y)\mathbb{1}_{[\xi_\beta,\infty)}(y)$. The variance of Y_β, which will appear in subsequent sections, is then given by

$$\sigma_\beta^2=\frac{1}{1-\beta}\int_{\xi_\beta}^{\infty}(x-\mu_\beta)^2 f(x)dx=\frac{1}{1-\beta}\int_{\beta}^{1}[F^{-1}(u)-\mu_\beta]^2du, \tag{17}$$

and using (15), (16), and (9), we can rewrite it as

$$\sigma_\beta^2 = \frac{1}{G_0(\xi_\beta)} \int_{\xi_\beta}^{\infty} \left(x - \frac{G_1(\xi_\beta)}{G_0(\xi_\beta)} \right)^2 f(x)dx = \left(\frac{G_2(\xi_\beta)}{G_0(\xi_\beta)} \right) - \left(\frac{G_1(\xi_\beta)}{G_0(\xi_\beta)} \right)^2. \tag{18}$$

Defining for $n = 1, 2$,

$$h_n = \frac{G_n(\xi_\beta)}{G_0(\xi_\beta)}, \tag{19}$$

leads to notationally simpler expressions for several of the above equations.

3 Asymptotic Efficiency of VaR Relative to ES

Appealing to well-known results for order statistics (David and Nagaraja, 2003) we obtain asymptotic normality for $\hat{\xi}_\alpha$ in the *central* (or *quantile*) case of interest here ($k_\alpha/n \to \alpha$ as $n \to \infty$). One way to state this result is that for large n, $\sqrt{n}(\hat{\xi}_\alpha - \xi_\alpha)$ is approximately Gaussian with mean zero and variance $\alpha(1-\alpha)/f^2(\xi_\alpha)$, the *asymptotic variance*. A corresponding result holds for $\sqrt{n}(\hat{\mu}_\beta - \mu_\beta)$ (Trindade et al., 2007) with the asymptotic variance given by the inverse of the second fraction on the right of (20). (See also Giurcanu and Trindade (2007) for joint asymptotic normality of $(\hat{\xi}_\beta, \hat{\mu}_\beta)$.) Appealing to these results we obtain the *asymptotic relative efficiency* (ARE) of the empirical estimator of VaR_α with respect to that of ES_β, as

$$\text{ARE}(\hat{\xi}_\alpha, \hat{\mu}_\beta) = \frac{\text{asymptotic variance of } \hat{\xi}_\alpha}{\text{asymptotic variance of } \hat{\mu}_\beta} = \frac{\alpha(1-\alpha)}{f^2(\xi_\alpha)} \cdot \frac{1-\beta}{\sigma_\beta^2 + \beta(\mu_\beta - \xi_\beta)^2}, \tag{20}$$

where σ_β^2 is as defined in (17).

For the purpose of simplifying subsequent discussions, we define $\text{ARE}(\hat{\xi}_\alpha, \hat{\mu}_\beta) \equiv H(\beta, p)$, where the function $H : \mathcal{D} \mapsto \mathbb{R}$ highlights the fact that the ARE is ultimately dependent on just these two parameters, and its domain is given by:

$$\mathcal{D} = (0,1) \times (0, \infty). \tag{21}$$

Then, choosing $\alpha = g_\beta$ as in (3) to force VaR_α and ES_β to coincide, leads to the following closed form expression for the ARE under random sampling from the EPD.

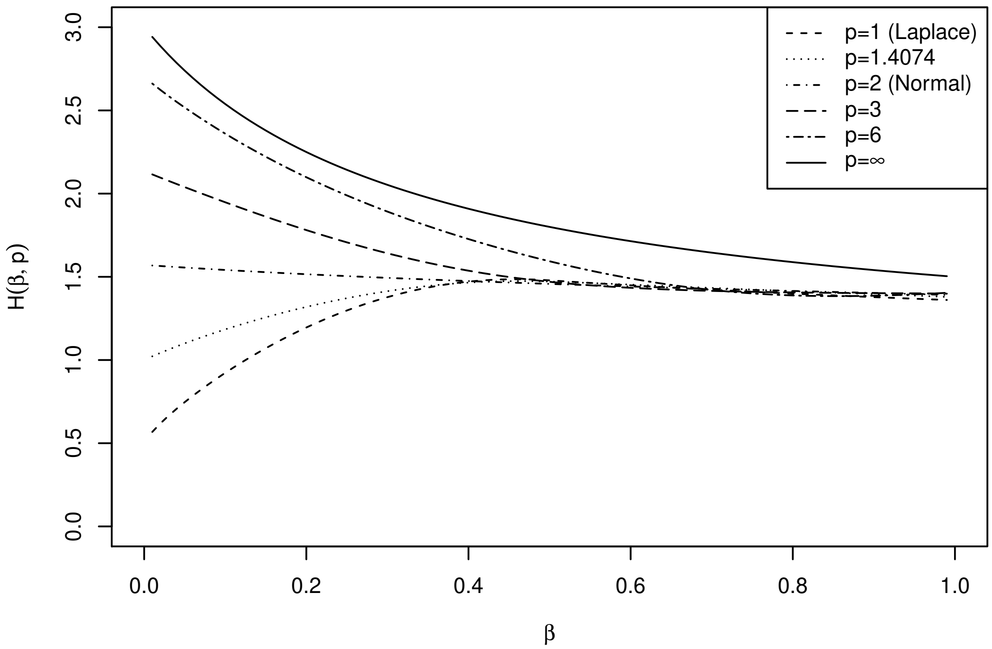
Lemma 1. *If $\alpha = g_\beta$ is as in (16), then the ARE of $\hat{\xi}_\alpha$ with respect to $\hat{\mu}_\beta$ as given in (20) is*

$$ARE(\hat{\xi}_\alpha, \hat{\mu}_\beta) \equiv H(\beta, p) = \frac{F_0(h_1) G_0(h_1)}{\left(\frac{p}{2\Gamma(1/p)}\right)^2 \exp\{-2\,(h_1)^p\}} \cdot \frac{G_0(\xi_\beta)}{h_2 - (h_1)^2 + F_0(\xi_\beta)\,(h_1 - \xi_\beta)^2}. \tag{22}$$

Proof. See Appendix Section B. □

Several interesting questions now arise, but primarily, when is $H(\beta, p) > 1$ so that ES is a more efficient estimator than VaR? Figure 3 shows a plot of $H(\beta, p)$ as a function of $0 < \beta < 1$, for different values of p. Although neither estimator is uniformly better, we do see that for p greater than about 1.4 (solid line), ES is uniformly more efficient. This uniformity is difficult to establish rigorously; but we can make concrete statements by investigating the limiting values of $H(\beta, p)$ separately for each variable while holding the other fixed. Consider first the cases $\beta \to 0$ and $\beta \to 1$, for fixed p. This yields the interesting results of the following theorem.

Figure 3: Plot of the $H(\beta, p)$ function defined in Lemma 1 vs. β, for select values of p.



Theorem 2 (Limiting behavior of the ARE in β)**.** *With $H(\beta, p)$ as defined in Lemma 1, and for $p \in (0, \infty)$, we have that:*

$$\lim_{\beta \to 0} H(\beta, p) \equiv H(0, p) = \frac{3\Gamma^3(1 + 1/p)}{\Gamma(1 + 3/p)}, \qquad \textit{and} \qquad \lim_{\beta \to 1} H(\beta, p) \equiv H(1, p) = \frac{e}{2}.$$

Proof. See Appendix Section C. □

The result for $H(0, p)$ agrees with Sherman (1997), where his $\text{eff}(\bar{x}, \tilde{x})$ corresponding to the ARE of the sample mean relative to the sample median, is our $1/H(0, p)$. We also note (as did Sherman, 1997) that $H(0, p)$ is an increasing function of p which maps the interval $(0, \infty)$ onto the interval $(0, 3)$. Consequently, there exists a unique $p^* \approx 1.4074$ such that for $p < p^*$, $H(0, p) < 1$ and for $p > p^*$, $H(0, p) > 1$. Thus, for leptokurtic (platikurtic) distributions, the median (mean) is a more efficient estimator, where the boundary between these two is exactly $p = p^*$.

On the other hand, the result $H(1, p) = e/2 \approx 1.36$ is remarkable! It says that in the limit of the right tail, the asymptotic variance of the tail median (VaR) is approximately 36% larger than that of the tail mean (ES); a result that holds uniformly for all p. Equating "efficiency" with "reduction in variance", this translates equivalently into the tail mean being approximately 26% more efficient than the tail median. The implications of this finding are clear: ES is a more efficient estimator than VaR for the typically high quantiles β that are used in practice.

Although not as interesting from a practical standpoint, the next set of results considers also the cases $p \to 0$ and $p \to \infty$, for fixed β. In the latter case, we are able to establish the following theorem.

Theorem 3 (Limiting behavior of the ARE as $p \to \infty$)**.** *With $H(\beta, p)$ as defined in Lemma 1, and for $\beta \in (0, 1)$, we have that:*

$$\lim_{p\to\infty} H(\beta, p) \equiv H(\beta, \infty) = \frac{1+\beta}{1/3+\beta}.$$

Proof. See Appendix Section D. $\square$

In the former case, we are unable to rigorously prove the particular result in question, but the arguments and graphical evidence presented in Case 3 of Appendix Section D strongly suggest that the following conjecture holds.

Conjecture 1 (Limiting behavior of the ARE as $p \to 0$)**.** *With $H(\beta, p)$ as defined in Lemma 1, and for $\beta \in (0, 1)$, we conjecture that:*

$$\lim_{p\to 0} H(\beta, p) \equiv H(\beta, 0) = 0.$$

Interestingly, and in parallel with the second result of Theorem 2, the limiting expression for $H(\beta, 0)$ also appears to be independent of the other parameter (β). Note that the expression for $H(\beta, \infty)$ is a decreasing function of β mapping the interval $(0, 1)$ onto the interval $(3/2, 3)$. Thus, for the limiting $\mathcal{U}[-1, 1]$ distribution obtained when $p \to \infty$, the tail mean would be the (uniformly) preferred measure for all β. Table 1 summarizes the limiting values of $H(\beta, p)$ along the 4 edges defined by $\{(\beta, p) : \beta = 0, 1 \text{ and } p = 0, \infty\}$. The values obtained at the corresponding 4 corners are shown inside the table, as the limit is approached either horizontally along p for fixed β, or vertically along β for fixed p. From this it is apparent that the function is continuous everywhere along $\beta = 0$, but discontinuous at each of the two corners: $(\beta = 1, p = 0)$ and $(\beta = 1, p = \infty)$. This points the way to extending the domain of definition of $H(\beta, p)$ from that given by (21), to its closure (minus the two corners)

$$\bar{\mathcal{D}} = \{(\beta, p) \in [0, 1] \times [0, \infty] \mid (\beta, p) \neq (1, 0) \text{ and } (\beta, p) \neq (1, \infty)\}, \tag{23}$$

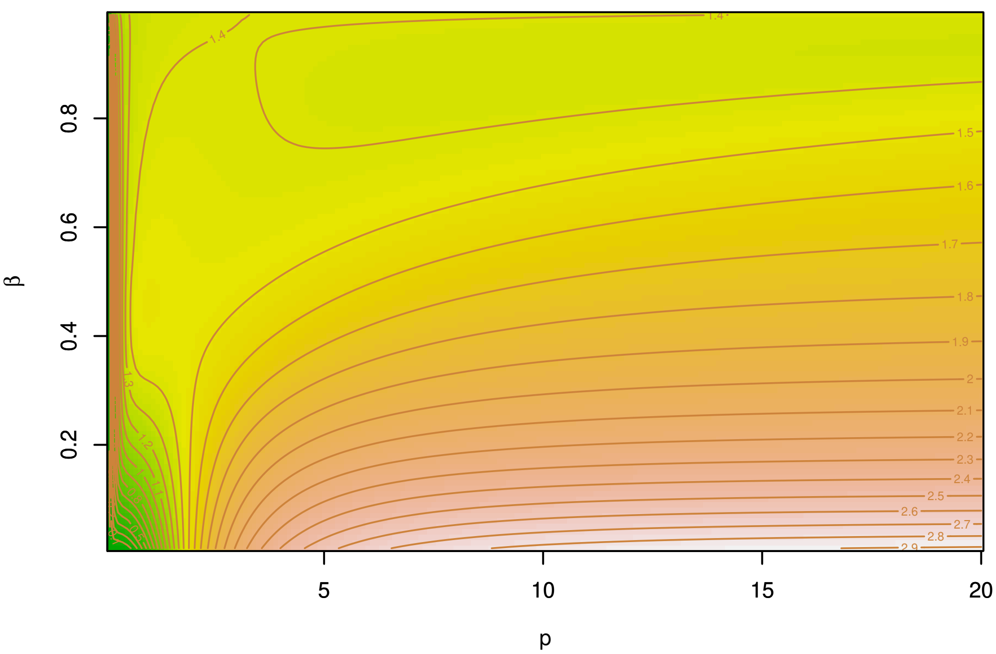
where we refrain from defining the function at each of the problematic corners in order to circumvent any non-uniqueness issues.

Table 1: Limiting values of the $H(\beta, p)$ function.

		$H(1,p)=\frac{e}{2}$		
		$p=0$	$p=\infty$	
$H(\beta,0)=0$	$\beta=1$	$\lim_{p\to 0} H(1,p)=e/2$ $\lim_{\beta\to 1} H(\beta,0)=0$	$\lim_{p\to\infty} H(1,p)=e/2$ $\lim_{\beta\to 1} H(\beta,\infty)=3/2$	$H(\beta,\infty)=\frac{1+\beta}{1/3+\beta}$
	$\beta=0$	$\lim_{p\to 0} H(0,p)=0$ $\lim_{\beta\to 0} H(\beta,0)=0$	$\lim_{p\to\infty} H(0,p)=3$ $\lim_{\beta\to 0} H(\beta,\infty)=3$	
		$H(0,p)=\frac{3\Gamma^3(1+1/p)}{\Gamma(1+3/p)}$		

Finally, Figure 4 displays a contour plot of $H(\beta, p)$ as a function of its two arguments. The overwhelming message from this plot is the fact that apart from the steep "cliff" at the left in the region $p < p^*$, the function is always larger than unity. Thus the tail mean is uniformly more

Figure 4: Contour plot of $H(\beta, p)$ as a function of its arguments.



efficient (regardless of β) than the tail median for distributions with $p > p^*$ (platikurtic). We can refine this further by considering a particular β, e.g., if $\beta = 0.975$ (Basel III), then $H(\beta, p) > 1$ for $p > 0.23$.

Remark 2. *In EVT terminology, a result for the ARE in (20) corresponding to heavy-tailed distributions with tail index $\theta > 0$ in the* intermediate *case where $k_\beta/n \to 1$ as $n \to \infty$, was derived by Danielsson and Zhou (2016):*

$$ARE(\hat{\xi}_\alpha, \hat{\mu}_\beta) = \frac{1-\beta}{1-\alpha} \cdot \frac{\theta - 2}{2(\theta - 1)}.$$

By comparison, recall that the type of convergence for the number of order statistics at which the summand (4) *starts in the* central *case, $k_\beta = [n\beta]$, is that $k_\beta/n \to \beta$ as $n \to \infty$, which is fundamentally different, and in our view more relevant for practical QRM.*

4 Discussion

We established limiting results concerning the ratio of asymptotic variances of the classical empirical estimators of location, tail median vs. tail mean, in the context of the flexible EPD family of distributions. The most remarkable result concerned the fact that in the limit of the right tail, the asymptotic variance of the tail median is approximately 36% larger than that of the tail mean, irrespective of the EPD shape parameter. Equating "efficiency" with "reduction in asymptotic variance", this translates equivalently into the tail mean being approximately 26%

more efficient than the tail median. The findings also offer a generalization of the solution to the age-old statistical quandary concerning which of the sample median vs. sample mean is the most efficient estimator of centrality.

The central tenet of this paper is that for sufficiently "light" tails, the tail mean is a more efficient estimator than the tail median, when the population quantiles that they are estimating coincide. This appears to remain true whether one is in a light-tailed (tail index $\theta = 0$) or heavy-tailed (tail index $\theta > 0$) regime. From a practical perspective, this message may have important repercussions for QRM practitioners with regard to choice of risk measure, VaR or ES, as follows:

- If the data on hand is believed to follow a light-tailed distribution ($\theta = 0$). Proceed by fitting an EPD via, e.g., maximum likelihood. Guided by efficiency considerations, the corresponding choice of risk measure in implementing Basel III (where $\alpha = 0.99$ and $\beta = 0.975$) would then be dictated by ascertaining whether or not the estimated value of the EPD shape parameter, p, exceeds 0.23. For other risk quantile levels α, the corresponding β can be determined from Figure 2 or equation (16), whence the appropriate choice can be made from Figure 4 or equation (22), bearing in mind that for $p > 1.4074$ ES is always more efficient.

- If the data follows a heavy-tailed distribution ($\theta > 0$). Remark 2 can be used to determine which of VaR/ES is more efficient, given an estimate of θ. However, relating α and β to yield the same value of the risk measures like we did through the function $g(\beta, p)$ for the light-tailed case of EPD, still requires distribution-specific knowledge. Thus, for example, setting $(1-\beta)/(1-\alpha) = 2.5 \approx e$ as suggested by Basel III, we see that $\text{ARE}(\hat{\xi}_\alpha, \hat{\mu}_\beta) > 1$ only for $\theta > 6$ (Danielsson and Zhou, 2016), and so for really heavy tails it is VaR that is less variable than ES at these specific quantiles. A further problem with this EVT approach is the fact that comparisons are made for the intermediate rather than central quantile case, whereas we argue the latter is more realistic than the former in practical applications since the probability corresponding to VaR_β is converging to β instead of 1 (Remark 2). Rather, the intermediate quantile case is considered merely because it leads to a general tractable solution.

There is therefore room for improvement in both approaches. The light-tailed case could benefit from a more general result for $\lim_{p\to 0} H(1,p)$ that would be independent of the distributional family, of the type mentioned for the heavy-tailed case. On the other hand, the heavy-tailed situation might benefit from a similar analysis as was done for light-tailed, where the focus is the central rather than intermediate quantile case. At present, both extensions would seem to be offer substantial analytical challenges.

A Overview of Proof Techniques

Throughout the proofs in the remaining sections, we will make liberal use of certain mathematical results. We document the primary ones here in order to add transparency to the proofs.

(i) "Big O" and "little o" notation. For sequences of real numbers $\{u_n\}$ and $\{v_n\}$, recall that $u_n = O(v_n)$ if and only if u_n/v_n is bounded, and $u_n = o(v_n)$ if and only if $\lim_{n\to\infty} u_n/v_n = 0$. Rules for manipulating these can be found in any advanced text. In particular, note that if $x_n = o(u_n)$ and $y_n = o(v_n)$, then $x_n + y_n = o(\max\{u_n, v_n\})$, and $x_n y_n = o(u_n v_n)$.

(ii) Asymptotic equivalence of (non-random) sequences. For real-valued sequences $a(x)$ and $b(x)$, we write $a(x) \approx b(x)$ if and only if $a(x)/b(x) \to 1$ as $x \to \infty$. An equivalent way of stating this definition (which points the way to arithmetic manipulations) is:

$$a(x) \approx b(x) \qquad \Longleftrightarrow \qquad \frac{a(x) - b(x)}{a(x)} = o(1).$$

Note however that for bounded sequences this simplifies: $a(x) \approx b(x) \Leftrightarrow a(x) - b(x) = o(1)$.

(iii) Establishing limits of ratios. For sufficiently small $x \downarrow 0$, recall from the geometric series expansion that

$$\frac{1}{1+x} = \frac{1}{1-(-x)} = 1 - x + x^2 + o(x^{-2}).$$

Suppose now that $\eta(z) = 1 + a/z + b/z^2 + o(z^{-2})$, where $z \to \infty$. A standard technique for dealing with limits of ratios is to set $x = a/z + b/z^2 + o(z^{-2})$, and to then employ the representation

$$\frac{1}{\eta(z)} = \frac{1}{1+x} = 1 - \left(\frac{a}{z} + \frac{b}{z^2} + o(z^{-2})\right) + \frac{a^2}{z^2} + o(z^{-2}) = 1 - \frac{a}{z} + \frac{a^2 - b}{z^2} + o(z^{-2}). \quad (24)$$

B Proof of Lemma 1

Define the numerators (V_1, V_2) and denominators (W_1, W_2) by writing:

$$H(\beta, p) = \frac{\alpha(1-\alpha)}{f^2(\xi_\alpha)} \cdot \frac{1-\beta}{\sigma_\beta^2 + \beta(\mu_\beta - \xi_\beta)^2} \equiv \frac{V_1}{W_1} \cdot \frac{V_2}{W_2}.$$

Now compute each of these 4 terms as follows.

V_1**:** From (16) and (19) we have $\alpha = F_0(\mu_\beta) = F_0(h_1)$, whence noting that $A_0 = 1$, we obtain, in view of (10), $1 - \alpha = A_0 - F_0(h_1) = G_0(h_1)$. Putting these together gives: $V_1 = \alpha(1-\alpha) = F_0(h_1)G_0(h_1)$.

V_2**:** By definition, $\beta = F_0(\xi_\beta)$, whence the fact that $A_0 = 1$ and (10) gives: $V_2 = 1 - \beta = A_0 - F_0(\xi_\beta) = G_0(\xi_\beta)$.

W_1**:** $\xi_\alpha = F^{-1}(\alpha)$, thus since it follows by (16) that $\alpha = g(\beta, p) = F_0(h_1)$, we have $\xi_\alpha = F^{-1}(F(h_1)) = h_1 \geq 0$, since $\xi_\alpha = \mu_\beta \geq 0$ (see Figure 1). Therefore, substituting this into (5) gives: $\sqrt{W_1} = f(\xi_\alpha) = p\exp\{-(h_1)^p\}/[2\Gamma(1/p)]$.

W_2**:** From (18) and (19), $\sigma_\beta^2 = h_2 - h_1^2$, and from (15) and (19), $\mu_\beta - \xi_\beta = h_1 - \xi_\beta$. These then give: $W_2 = \sigma_\beta^2 + \beta(\mu_\beta - \xi_\beta)^2 = h_2 - h_1^2 + F_0(\xi_\beta)(h_1 - \xi_\beta)^2$.

C Proof of Theorem 2

To assess the limiting behavior of $H(\beta, p)$, we take the approach of considering the limits of its individual components, separately for the cases $\beta \to 0$ (Case 1) and $\beta \to 1$ (Case 2). A key idea is to make the change of variable $\beta = F_0(\xi_\beta)$ early on in each case. This sheds light on the connections to the Gamma and incomplete Gamma, thus naturally allowing one to invoke the properties and asymptotics of these functions. In all cases, the basic strategy will be to ascertain the limits of each piece of $H(\beta, p)$ as given in Lemma 1, which are then combined to obtain the desired result.

Case 1: $\beta \to 0$ **(for fixed** p**)**

We can rewrite (20), using $\alpha = g_\beta$ so that $\mu_\beta = \xi_\alpha$, as $H(\beta, p) = P_1 \cdot Q_1$, where

$$P_1 = \frac{\alpha(1-\alpha)}{f^2(\xi_\alpha)}, \qquad Q_1 = \frac{1-\beta}{\sigma_\beta^2 + \beta(\mu_\beta - \xi_\beta)^2}.$$

If we let $\beta \to 0$, then $\mu_\beta \to 0$ which is the average value of F_0^{-1} over $(0,1)$. Hence, as $\beta \to 0$, we have $g_\beta \to 1/2$ which is the value of F_0 at $x = 0$. Clearly as $\beta \to 0$, $f(\mu_\beta) \to f(0) = \frac{p}{2\Gamma(1/p)}$. Thus, $P_1 \to \frac{1}{4} / \left[\frac{p}{2\Gamma(1/p)}\right]^2$ as $\beta \to 0$.

Define ξ_β by $F_0(\xi_\beta) = \beta$. Then, $\beta \to 0$ implies $\xi_\beta = F_0^{-1}(\beta) \to -\infty$. Considering the term $\beta(\mu_\beta - \xi_\beta)^2$ in Q_1, we have, since $\mu_\beta \to 0$ as $\beta \to 0$,

$$\lim_{\beta \to 0} \beta(\mu_\beta - \xi_\beta)^2 \quad = \quad \lim_{\beta \to 0} -2\mu_\beta \beta F_0^{-1}(\beta) + \beta[F_0^{-1}(\beta)]^2. \tag{25}$$

Making a change of variable $\beta = F_0(\xi_\beta)$ and applying l'Hopital's rule, we note, for $k = 1, 2$,

$$\lim_{\beta \to 0} \beta[F_0^{-1}(\beta)]^k = \lim_{\xi_\beta \to -\infty} F_0(\xi_\beta)\xi_\beta^k = \lim_{\xi_\beta \to -\infty} \frac{F_0(\xi_\beta)}{\xi_\beta^{-k}} = \lim_{\xi_\beta \to -\infty} \frac{\frac{p}{2\Gamma(1/p)} e^{-|\xi_\beta|^p}}{-k\xi_\beta^{-k-1}} = 0.$$

Hence, the limit in (25) is 0. Finally, we note that as $\beta \to 0$, $\sigma_\beta^2 \to \sigma_0^2 = \int_0^1 [F_0^{-1}(u)]^2 du$. Making a change of variable, $u = F_0(t)$, we can write

$$\sigma_0^2 = \int_{-\infty}^{\infty} \frac{p}{2\Gamma(1/p)} t^2 e^{-|t|^p} dt = \frac{p}{\Gamma(1/p)} \int_0^\infty t^2 e^{-t^p} dt = \frac{\Gamma(3/p)}{\Gamma(1/p)}.$$

Thus, $Q_1 \to 1/[\Gamma(3/p)/\Gamma(1/p)]$ as $\beta \to 0$.

Hence, we have as $\beta \to 0$,

$$\lim_{\beta \to 0} H(\beta, p) = \lim_{\beta \to 0} P_1 \cdot Q_1 = \frac{1/4}{\left[\frac{p}{2\Gamma(1/p)}\right]^2} \cdot \frac{1}{\frac{\Gamma(3/p)}{\Gamma(1/p)}} = \frac{1}{p^2} \frac{[\Gamma(1/p)]^3}{\Gamma(3/p)} = \frac{3[\Gamma(1+1/p)]^3}{\Gamma(1+3/p)}. \tag{26}$$

Case 2: $\beta \to 1$ **(for fixed** p**)**

We can rewrite (22) as $H(\beta, p) = P_2 \cdot Q_2$ where

$$P_2 = \frac{G_0(h_1) G_0(\xi_\beta)}{\left(\frac{1}{2\Gamma(1/p)}\right)^2 \exp\{-2\,(h_1)^p\}}, \quad Q_2 = \frac{F_0(h_1)}{p^2[h_2 - (h_1)^2 + F_0(\xi_\beta)\,(h_1 - \xi_\beta)^2]}.$$

Define ξ_β by $F_0(\xi_\beta) = \beta$. Then, $\beta \to 1$ implies $\xi_\beta \to \infty$. In particular, we have, since $\xi_\beta > 0$, from (11) that

$$G_n(\xi_\beta) = \frac{\Gamma(\frac{n+1}{p}, \xi_\beta^p)}{2\Gamma(1/p)}, \qquad h_n = \frac{G_n(\xi_\beta)}{G_0(\xi_\beta)} = \frac{\Gamma(\frac{n+1}{p}, \xi_\beta^p)}{\Gamma(1/p, \xi_\beta^p)}, \qquad \text{for } n = 1, 2.$$

Defining,

$$s(a, z) = 1 + \frac{a-1}{z} + \frac{(a-1)(a-2)}{z^2} + o\left(\frac{1}{z^2}\right), \qquad z \to \infty, \tag{27}$$

we can represent $\Gamma(a,z)$, see Erdelyi et al. (1953, p. 135, (6)), as

$$\Gamma(a,z) = \frac{z^a e^{-z}}{z} s(a,z), \qquad z \to \infty. \tag{28}$$

Consequently, we have, using (27) in the second step,

$$\begin{aligned} G_n(\xi_\beta) &= \frac{1}{2\Gamma(1/p)} \frac{(\xi_\beta^p)^{(n+1)/p} e^{-\xi_\beta^p}}{\xi_\beta^p} s\left(\frac{n+1}{p}, \xi_\beta^p\right) \\ &= \frac{1}{2\Gamma(1/p)} \frac{\xi_\beta^{n+1} e^{-\xi_\beta^p}}{\xi_\beta^p} \left(1 + \frac{(n+1)/p - 1}{\xi_\beta^p} + o\left(\frac{1}{\xi_\beta^p}\right)\right), \end{aligned}$$

whence, employing the technique of (24) to deal with the term $1/G_0(\xi_\beta)$, gives

$$\begin{aligned} h_n &= \frac{G_n(\xi_\beta)}{G_0(\xi_\beta)} = \xi_\beta^n \, s\left(\frac{n+1}{p}, \xi_\beta^p\right) \Big/ s\left(\frac{1}{p}, \xi_\beta^p\right) \\ &= \xi_\beta^n \left(1 + \frac{(n)/p}{\xi_\beta^p} + o\left(\frac{1}{\xi_\beta^p}\right)\right). \end{aligned} \tag{29}$$

In particular,

$$\begin{aligned} G_0(\xi_\beta) &= \frac{1}{2\Gamma(1/p)} \frac{\xi_\beta e^{-\xi_\beta^p}}{\xi_\beta^p} \left(1 + \frac{1/p - 1}{\xi_\beta^p} + o\left(\frac{1}{\xi_\beta^p}\right)\right), \\ h_1 &= \xi_\beta \left(1 + \frac{1/p}{\xi_\beta^p} + o\left(\frac{1}{\xi_\beta^p}\right)\right), \end{aligned}$$

which implies $G_0(\xi_\beta) \to 0$, $F_0(\xi_\beta) \to 1$, and $h_1 \to \infty$, as $\xi_\beta \to \infty$. Furthermore, we have, once again using (24), that

$$G_0(h_1) = \frac{1}{2\Gamma(1/p)} \frac{h_1 e^{-(h_1)^p}}{(h_1)^p} \left(1 + \frac{1/p - 1}{h_1^p} + o\left(\frac{1}{h_1^p}\right)\right). \tag{30}$$

Consequently, we have $G_0(h_1) \to 0$ and $F_0(h_1) \to 1$ as $\xi_\beta \to \infty$. Thus, we have, as $\xi_\beta \to \infty$,

$$\begin{aligned} P_2 = \frac{G_0(h_1) G_0(\xi_\beta)}{(\frac{1}{2\Gamma(1/p)})^2 \exp\{-2\,(h_1)^p\}} &\approx \frac{h_1}{e^{(h_1)^p} (h_1)^p} \frac{\xi_\beta}{e^{\xi_\beta^p} \xi_\beta^p} e^{2(h_1)^p} = (h_1)^{1-p} \xi_\beta^{1-p} e^{(h_1)^p - \xi_\beta^p} \\ &\approx \xi_\beta^{2-2p} e^{(h_1)^p - \xi_\beta^p}. \end{aligned} \tag{31}$$

Now, note that for the exponent of the exponential in (31) we have

$$(h_1)^p - \xi_\beta^p = \xi_\beta^p \left(1 + \frac{1/p}{\xi_\beta^p} + o\left(\frac{1}{\xi_\beta^p}\right)\right)^p - \xi_\beta^p = \xi_\beta^p \left(1 + p\frac{1/p}{\xi_\beta^p} + o\left(\frac{1}{\xi_\beta^p}\right)\right) - \xi_\beta^p = 1 + o(1),$$

whence, $e^{(h_1)^p - \xi_\beta^p} \to e$ as $\xi_\beta \to \infty$. Thus, we have as $\xi_\beta \to \infty$

$$P_2 \approx {\xi_\beta}^{2-2p} \cdot e \tag{32}$$

Considering the terms in Q_2, we have, as $\xi_\beta \to \infty$,

$$\begin{aligned} (h_1 - \xi_\beta)^2 &= \left[\xi_\beta \left(1 + \frac{1/p}{\xi_\beta^p} + o\left(\frac{1}{(\xi_\beta)^p}\right)\right) - \xi_\beta\right]^2 = \xi_\beta^2 \left(\frac{1/p}{\xi_\beta^p} + o\left(\frac{1}{(\xi_\beta)^p}\right)\right)^2 \\ &= \frac{\xi_\beta^{2-2p}}{p^2} \left(1 + o\left(\frac{1}{(\xi_\beta)^p}\right)\right), \end{aligned}$$

and finally, using (29),

$$h_2 - (h_1)^2 = \xi_\beta^2 \left[\frac{s(3/p, \xi_\beta^p)s(1/p, \xi_\beta^p) - s(2/p, \xi_\beta^p)s(2/p, \xi_\beta^p)}{s(1/p, \xi_\beta^p)s(1/p, \xi_\beta^p)} \right]. \tag{33}$$

Since, using the second order terms in (27),

$$\begin{aligned} s(3/p, z)s(1/p, z) - s(2/p, z)s(2/p, z) &= \frac{1/p^2}{z^2} + o\left(\frac{1}{z^2}\right), \\ s(1/p, z)s(1/p, z) &= 1 + \frac{2/p - 2}{z} + o\left(\frac{1}{z}\right), \end{aligned}$$

and, again using (24) to deal with the inversion of the denominator,

$$\frac{s(3/p, z)s(1/p, z) - s(2/p, z)s(2/p, z)}{s(1/p, z)s(1/p, z)} = \frac{1/p^2}{z^2} + o\left(\frac{1}{z^2}\right),$$

we have, as $\xi_\beta \to \infty$,

$$h_2 - (h_1)^2 \approx \frac{\xi_\beta^{2-2p}}{p^2}.$$

Thus, we have as $\xi_\beta \to \infty$,

$$Q_2 \approx \frac{1}{p^2[\frac{{\xi_\beta}^{2-2p}}{p^2} + \frac{{\xi_\beta}^{2-2p}}{p^2}]} \tag{34}$$

Combining, (32) and (34), we have as $\xi_\beta \to \infty$,

$$H(\beta, p) = P_2 \cdot Q_2 \approx \frac{[1-0]\xi_\beta^{2-2p} \cdot e}{p^2 \left[\frac{\xi_\beta^{2-2p}}{p^2} + (1-0)\frac{\xi_\beta^{2-2p}}{p^2}\right]}, \tag{35}$$

and thus $H(\beta, p) \to \dfrac{e}{2}$ as $\xi_\beta \to \infty$ ($\Leftrightarrow \beta \to 1$).

D Discussion of Conjecture 1 and Proof of Theorem 3

For continuity with Cases 1 and 2 above, let Case 3 denote the limiting result of $p \to 0$ given by Conjecture 1, and Case 4 the limiting result of $p \to \infty$ stated in Theorem 3. In Case 3 we are only able to rigorously show that the limiting EPD PDF is zero everywhere, whence it follows that the corresponding CDF is converging to 1/2. For the resulting convergence of $H(\beta, p)$ to zero as $p \to 0$, we have only graphical evidence. Case 4 is tackled by considering the limiting distribution that results when $p \to \infty$. The computation of $H(\beta, \infty)$ is then straightforward for the limiting $\mathcal{U}[-1, 1]$. Usage of the (well-defined) generalized quantile function to invert the CDF permits the interchange of limits and integrals, via the Lebesgue Dominated Convergence theorem, thus justifying the $\mathcal{U}[-1, 1]$ computation.

Case 3: $p \to 0$ (for fixed β)

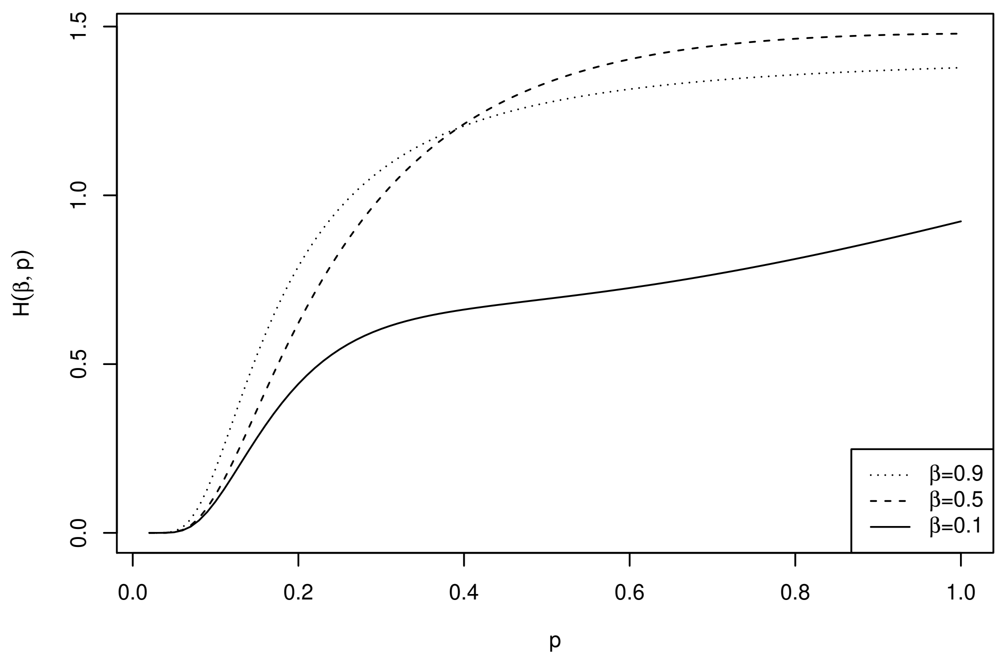
Note that since the EPD PDF, $f(x; p) = p[2\Gamma(1/p)]^{-1} \exp\{-|x|^p\}$, is the product of term 1 and term 2, where term 2 is the exponential, it converges (uniformly) to 0 as $p \to 0$. This follows from the fact that term 2 is uniformly bounded by 1 for all x and all $p > 0$, whereas term 1 converges to 0 as $p \to 0$. Since the total area under the PDF is 1 (for any p), then for smaller p the PDF has to spread further out to capture this total area. Since half the area occurs for $x \in (-\infty, 0)$ and half for $x \in (0, \infty)$, this in turn forces the CDF $F(x; p)$ to converge to 1/2 for each x as $p \to 0$.

It is not obvious how this fact implies that $H(\beta, p) \to 0$ as $p \to 0$. A critical complicating factor for constructing an analytical proof for this case is the fact that although β is fixed, ξ_β is not fixed (as a function of p), and, indeed, exhibits the following limiting behavior:

$$\lim_{p \to 0} \xi_\beta = \begin{cases} -\infty, & \text{if } \beta < 0.5, \\ 0, & \text{if } \beta = 0.5, \\ +\infty, & \text{if } \beta > 0.5. \end{cases}$$

Possibilities we investigated included efforts to establish asymptotic expansions for the incomplete Gamma functions in the representation of $G_n(\xi_\beta)$ given σ_β^2. Nevertheless the result seems to be true, as is apparent from Figure 5.

Figure 5: Plot of $H(\beta, p)$ as a function of p for select values of β.



Case 4: $p \to \infty$ (for fixed β)

Denote by $f(x; p)$ and $F(x; p)$ the EPD PDF and CDF, respectively, for given shape parameter p. Formally define $f(x; \infty) = \lim_{p\to\infty} f(x; p)$ and $F(x; \infty) = \lim_{p\to\infty} F(x; p)$. Note that $f(x; \infty) = \frac{1}{2}\mathbb{1}_{[-1,1]}(x)$ is a uniform distribution on $[-1, 1]$. Thus, $F(x; \infty) = \frac{1+x}{2}\mathbb{1}_{[-1,1]}(x) + \mathbb{1}_{(1,\infty)}(x)$, and we have from the Lebesgue Dominated Convergence theorem (Royden, 1988) that:

$$\lim_{p\to\infty} F(x; p) = F(x; \infty), \qquad \text{for } x \in \mathbb{R}. \tag{36}$$

Throughout, we will let $F^-(u;\infty)$ denote the generalized inverse of $F(x;\infty)$, also known as the generalized quantile function (Embrechts and Hofert, 2013). Thus, we have $F^-(u;\infty) = 2u-1$ for $0 < u \le 1$. Noting (now and for the remainder of the proof) that in the limit when $p=\infty$ we are dealing with a $\mathcal{U}[-1,1]$ distribution, straightforward calculations yield

$$\mu_\beta(\infty) \equiv \mu(\beta,\infty) = \frac{1}{1-\beta}\int_\beta^1 F^-(u;\infty)du = \beta, \qquad g_\beta(\infty) \equiv g(\beta,\infty) = F(\mu(\beta;\infty),\infty) = \frac{1+\beta}{2},$$

$$\sigma_\beta^2(\infty) \equiv \sigma^2(\beta,\infty) = \frac{1}{1-\beta}\int_\beta^1 [F^-(u;\infty) - \mu(\beta,\infty)]^2 du = \frac{(1-\beta)^2}{3},$$

and hence,

$$H(\beta,\infty) \quad \equiv \quad \frac{g_\beta(\infty)(1-g_\beta(\infty))}{[f(\mu_\beta(\infty);\infty)]^2}\cdot\frac{1-\beta}{\sigma_\beta^2(\infty)+\beta[\mu_\beta(\infty)-F^-(\beta;\infty)]^2} = \frac{1+\beta}{1/3+\beta}.$$

Having thus demonstrated that $H(\beta,\infty)$ exists, we will now show that indeed $\lim_{p\to\infty} H(\beta,p) = H(\beta,\infty)$, and thus verify Theorem 3. Recall from Section 2 that

$$\mu_\beta = \mu(\beta,p) = \frac{1}{1-\beta}\int_\beta^1 F^-(u;p)du, \qquad g_\beta = g(\beta,p) = F(\mu(\beta,p);p),$$

$$\sigma_\beta^2 = \sigma^2(\beta,p) = \frac{1}{1-\beta}\int_\beta^1 [F^-(u;p) - \mu(\beta,p)]^2 du.$$

For $0 < \delta < 1$, define

$$e_n(\delta) = \lim_{p\to\infty}\int_\delta^\infty t^n f(t,p)dt,$$

and note that, since $\lim_{x\to 0} x\Gamma(x) = \lim_{x\to 0}\Gamma(1+x) = 1$, we have, setting $x = 1/p$, that $\lim_{p\to\infty}\frac{p}{2\Gamma(1/p)} = 1/2$. Since, for $n = 0,1,2$ and for $p \ge 2$ and $t \ge 1$, we have $t^n e^{-t^p} \le e^{-t^{p/2}} \le e^{-t}$, we see that

$$e_n(\delta) = \lim_{p\to\infty}\int_\delta^\infty t^n f(t,p)dt = 1/2\int_\delta^\infty t^n \mathbb{1}_{[-1,1]}(t)dt = \frac{1-\delta^{n+1}}{2(n+1)}. \tag{37}$$

(This follows by noting that since the integrand on the left hand side of (37) is bounded by the (integrable) function e^{-t}, we can, applying the Lebesgue Dominated Convergence theorem, bring the limit inside the integrand.) Now define $B = B(p)$ by $F(B;p) = \beta$. Then, making a change of variable $u = F(t;p)$ we obtain

$$\mu_\beta = \mu(\beta,p) = \frac{1}{1-\beta}\int_{B(p)}^\infty tf(t;p)dt, \qquad g_\beta = g(\beta,p) = F(\mu(\beta,p);p),$$

$$\sigma_\beta^2 = \sigma^2(\beta,p) = \frac{1}{1-\beta}\int_{B(p)}^\infty (t-\mu(\beta,p))^2 f(t;p)dt.$$

Since $e_n(\delta)$ given by (37) is continuous in δ and (by an analogous argument) $F(x;p)$ is also continuous in x, the convergence in (36) (in the appropriate topology) for $x = \mu(\beta;p)$, where $0 < \beta < 1$, implies that

$$\lim_{p\to\infty} B(p) = \lim_{p\to\infty} F^-(\beta;p) = F^-(\beta;\infty) = 2\beta - 1,$$

whence we then have that, assembling these and earlier results,

$$\begin{aligned}
\lim_{p\to\infty}\mu(\beta,p) &= \lim_{p\to\infty}\frac{1}{1-\beta}e_1(B(p)) = \frac{1}{1-\beta}e_1(2\beta-1) = \beta,\\
\lim_{p\to\infty}g(\beta,p) &= \lim_{p\to\infty}F(\mu(\beta,p);p) = F(\beta;\infty) = \frac{1+\beta}{2},\\
\lim_{p\to\infty}\sigma^2(\beta,p) &= \lim_{p\to\infty}\frac{1}{1-\beta}\left[e_2(B(p)) - 2\mu(\beta,p)e_1(B(p)) + \mu^2(\beta,p)e_0(B(p))\right]\\
&= \frac{1}{1-\beta}\left[e_2(2\beta-1) - 2\beta e_1(2\beta-1) + \beta^2 e_0(2\beta-1)\right] = \frac{(1-\beta)^2}{3}.
\end{aligned}$$

Thus, we obtain

$$\lim_{p\to\infty}H(\beta,p) = H(\beta,\infty) = \frac{1+\beta}{1/3+\beta}. \tag{38}$$

Conflict of Interest

The authors declare that they have no conflict of interest.

Acknowledgements

We are indebted to detailed comments from two anonymous referees that led to vast improvements in the paper.

References

[1] Abramowitz, M., Stegun, I.A., eds. (1972). *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*, New York: Dover Publications.

[2] Acerbi, C., and Tasche, D. (2002). "On the coherence of expected shortfall", *Journal of Banking & Finance*, 26, 1487–1503.

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

[4] Basel III (2013). "Fundamental review of the trading book: A revised market risk framework", *Technical report*, Basel Committee on Banking Supervision, October 2013.

[5] Chun, S.Y., Shapiro A., and Uryasev, S. (2012). "Conditional Value-at-Risk and Average Value-at-Risk: Estimation and Asymptotics", *Operations Research*, 60, 739–756.

[6] Danielsson, J., and Zhou, C. (2016). "Why Risk Is So Hard to Measure", *DNB Working Paper No. 494*, De Nederlandsche Bank NV, Amsterdam (The Netherlands).

[7] David, H., and Nagaraja, H. (2003). *Order Statistics*, 3rd ed., Hoboken: Wiley.

[8] Duffie, D., Singleton, K.J. (2003). *Credit Risk: Pricing, Measurement, and Management*, Princeton: Princeton University Press.

[9] Embrechts, P., and Hofert, M. (2013). "A note on generalized inverses", *Mathematical Methods in Operations Research*, **77**, 423–432.

[10] Embrechts, P., and Hofert, M. (2014). "Statistics and Quantitative Risk Management for Banking and Insurance", *Annual Review of Statistics and Its Application*, **1**, 493–514.

[11] Embrechts, P., Puccetti, G., Rüschendorf, L., Wang, R., and Beleraj, A. (2014). "An academic response to Basel 3.5", *Risks*, 2, 25–48.

[12] Erdelyi, A., Magnus, W., Oberhettinger, F., and Tricomi, F.G. (1953). *Higher Transcendental Functions*, Vol. 1, New York: McGraw-Hill.

[13] Follmer, H., and Schied, A. (2011). *Stochastic Finance: An Introduction in Discrete Time*, 3rd ed., Berlin: de Gruyter.

[14] Giurcanu, M. and Trindade, A.A. (2007). "Establishing Consistency of M-Estimators Under Concavity with an Application to Some Financial Risk Measures," *Journal of Probability and Statistical Science*, 5, 123–136.

[15] Gomez, E., Gomez-Villegas, M.A., Marin, J.M. (1998). "A multivariate generalization of the power exponential family of distributions", *Communications in Statistics*, A27, 589–600.

[16] Jorion, P. (2003). *Financial Risk Manager Handbook*, 2nd ed., New York: Wiley.

[17] Landsman, Z., and Valdez, E. (2003). "Tail conditional expectations for elliptical distributions", *North American Actuarial Journal*, 7, 55–71.

[18] Maronna, R. (2011). "Robust Statistical Methods", in *International Encyclopedia of Statistical Science*, M. Lovric (ed.), pp. 1244–1248. Berlin, Heidelberg: Springer.

[19] McNeil, A.J., Frey, R., and Embrechts, P. (2005). *Quantitative Risk Management: Concepts, Techniques, Tools*, Princeton, NJ: Princeton Univ. Press.

[20] Mineo, A.M., and Ruggieri, M. (2005). "A software tool for the exponential power distribution: the normalp package", *Journal of Statistical Software*, 12(4), 1–23.

[21] Nadarajah, S. (2005). "A generalized normal distribution", *Journal of Applied Statistics*, 32, 685–694.

[22] Pflug, G., and Romisch, W. (2007). *Modeling, Measuring, and Managing Risk*, London: World Scientific.

[23] Prudnikov, A.P., Brychkov, Y.A. and Marichev, O.I. (1986). *Integrals and Series (Volume Two: Special Functions)*, Amsterdam: Overseas Publishers Association.

[24] Rockafellar, R., and Uryasev, S., (2000). "Optimization of conditional value-at-risk", *Journal of Risk*, 2, 21–41.

[25] Royden, H.L. (1988). *Real Analysis*. Englewood Cliffs: Prentice Hall.

[26] Sherman, M. (1997). "Comparing the Sample Mean and the Sample Median: An Exploration in the Exponential Power Family", *The American Statistician*, **51**, 52–54.

[27] Temme, N.M. (1996). *Special Functions: An Introduction to the Classical Functions of Mathematical Physics*, New York: Wiley.

[28] Trindade, A.A., Uryasev, S., Shapiro, A., and Zrazhevsky, G. (2007). "Financial Prediction with Constrained Tail Risk", *Journal of Banking and Finance*, 31, 3524–3538.

[29] Yamai, Y., and Yoshiba, T. (2002). “Comparative Analyses of Expected Shortfall and Value-at-Risk: Their Estimation Error, Decomposition, and Optimization”, *Monetary and Economic Studies (Bank of Japan)*, 20, 87–121.

[30] Yamai, Y., and Yoshiba, T. (2005). “Value-at-risk versus expected shortfall: A practical perspective”, *Journal of Banking & Finance*, 29, 997–1015.