Report

Abstract

In this paper, a tail-based estimator of Expected Shortfall (the *adjusted tail-based normal approximation*) is discussed. The objective is to compare this estimator with the Arithmetic Average estimator through MSE, variance, and bias metrics. The original paper does this for the *t*, *Gamma, Weibull*, and other distributions with various parameters. In our replication, we will see if we achieve similar results - in particular, if the *adjusted tail-based normal approximation* substantially outperforms the Arithmetic Average estimator.

1 Introduction

In the paper *A Simple and Robust Approach for Expected Shortfall Estimation* [2], an adjusted tail-based normal approximation is proposed and compared to other ES estimators.
Expected Shortfall (ES), otherwise known as Conditional VaR (CVAR), is a measure of risk quantifying tail risk that takes a weighted average of losses beyond the VaR cutoff.

That is, given $\beta \in (0,1)$, $ES_\beta(L) = \frac{1}{1-\beta}\int_\beta^1 VaR_\phi(L)d_\phi$ where $VaR_\phi(L) = \inf\{z \in \mathbb{R} | P(L \le z) \ge \phi\}$

In ES estimation, global-based approaches are typically considered where all sample points are used to estimate parameters (see [2]). However, since ES depends on tail behaviours, global-based approximations often underestimate the true ES. [2] In this paper, a tail-based approach which uses only tail sample points, is proposed.
In particular, it is applied to a normal distribution for reasons of having useful properties and sufficiently accurate results.[2]

2 Preliminaries

Let $y_1, \cdots, y_N$ be a sample of losses. Given $\alpha \in (0,1)$, define A_α, the α-th quantile as $A_\alpha = (\lfloor N\alpha \rfloor + 1 - N\alpha)y_{(\lfloor N\alpha \rfloor)} + (N\alpha - \lfloor N\alpha \rfloor)y_{(\lfloor N\alpha \rfloor + 1)}$ [2]. Choose a threshold α such that $\alpha < \beta$. Define $X \sim N(\mu, \sigma^2)$ to approximate the sample right tail beyond A_α. Let $z_\alpha = \Phi^{-1}(\alpha)$ denote the z-score of a standard normal distribution.
From mathematical calculations in the paper, the following sample approximations of μ and σ^2 are obtained: [2]

$\hat{\sigma}^2 = [z_\alpha + 1 - \frac{z_\alpha}{(1-\alpha)\sqrt{2\pi}} e^{-z_\alpha^2/2}]^{-1} [\frac{\sum_{n=1}^{N} (y_n - A_\alpha)^2 1_{y_n > A_\alpha}}{\sum_{n=1}^{N} 1_{y_n > A_\alpha}}]$

$\hat{\mu} = A_\alpha - \hat{\sigma} z_\alpha$

Let $\mathrm{VaR}_\beta(X)$ and $\mathrm{ES}_\beta(X)$ be the estimated VaR and ES at level β

Then it follows immediately from the fact that *X* is normally distributed:

$\mathrm{VaR}_\beta(X) = \hat{\mu} + \hat{\sigma} z_\beta$,

$\mathrm{ES}_\beta(X) = \mathbb{E}[X | X > \mathrm{VaR}_\beta] = \hat{\mu} + \frac{\hat{\sigma}}{(1-\beta)\sqrt{2\pi}} e^{-z_\beta^2/2}$

Note that distributions of financial data are usually heavy-tailed, so in this paper, we consider distributions such as *t*, Weibull, and Generalized Pareto Distributions. The goal is to calculate several metrics (MSE, variance, and bias) for various estimators for these heavy-tailed distributions, to see the metrics of which estimator is the lowest.

The paper notes that while this *tail-based normal approximation* performs better than global-based approaches, it can be noted that the errors have dependence on the shape parameters of the distributions. The *adjusted tail-based normal approximation* (which will be described below) adjusts the *tail-based normal approximation* by a factor related to tail weight statistics. This adjusted approximation is the main focus of the paper and of the project.

3 Adjusted Tail-based Normal Approximation

Let *W* be a loss random variable with a β-level ES estimate $\mathrm{ES}_\beta(W)$.

Define a ratio that measures estimation errors $R_{\alpha,\beta}$ where $R_{\alpha,\beta} = \frac{\mathrm{ES}_\beta(W) - A_\alpha}{\mathrm{ES}_\beta(X) - A_\alpha}$.

Recall that a measure for the tail weight is needed to adjust for the tail-based normal approximation. Here, the *conditional skewness* $\gamma_\alpha = \frac{E[(W-A_\alpha)^3 | W > A_\alpha]}{(E[(W-A_\alpha)^2 | W > A_\alpha])^{1.5}}$ is considered.

The idea is to build a regression model $\hat{R}_{\alpha,\beta} = f_{\alpha,\beta}(\gamma_\alpha)$ defined below. Consider the *t*-distribution: choose different degrees of freedom (df), yielding *t-distributions* with different *df* with respective conditional skewness γ_α and ratios $R_{\alpha,\beta}$. With these pairs of γ_α and $R_{\alpha,\beta}$, a nonlinear regression can be fit for $\alpha = 0.95, \beta = 0.99 \text{ or } 0.995$ resulting in:

$\hat{R}_{\alpha,\beta} = f_{\alpha,\beta}(\gamma_\alpha) = b_0 + b_1 e^{-b_2 y_\alpha} + b_3 \gamma_\alpha^{-1} + b_4 \gamma_\alpha^{-2}$ with the coefficients:

For $\alpha = 0.95, \beta = 0.99$,

$b_0 = 0.8611, b_1 = 0.5191, b_2 = 0.9747, b_3 = 0.6099, b_4 = -0.9413$

For $\alpha = 0.95, \beta = 0.995$,
$b_0 = 0.9919, b_1 = 0.6681, b_2 = 0.9607, b_3 = 0.6022, b_4 = -1.4623$
For *any W*, we can use this $f_{\alpha,\beta}(\gamma_\alpha)$ (which was calculated only for *t* distributions!) to adjust our tail-based approximation $\mathrm{ES}_\beta(X)$ [2].

In particular, the *adjusted tail-based* normal approximation is defined as:
$\hat{\mathrm{ES}}_\beta(X) = [\mathrm{ES}_\beta(X) - A_\alpha] f_{\alpha,\beta}(\gamma_\alpha) + A_\alpha$

In summary, the adjusted tail-based ES estimator at level β is obtained as follows: [2]
(i) Choose an α such that $\alpha < \beta$ and calculate A_α
(ii) Compute $\hat{\sigma^2}$ and $\hat{\mu}$, the parameters of the tail-based normal approximation
(iii) Compute $\gamma_\alpha = (\frac{\sum_{n=1}^N (y_n - A_\alpha)^3 1_{y_n > A_\alpha}}{\sum_{n=1}^N 1_{y_n > A_\alpha}})[\frac{\sum_{n=1}^N (y_n - A_\alpha)^2 1_{y_n > A_\alpha}}{\sum_{n=1}^N 1_{y_n > A_\alpha}}]^{-1.5}$, the *sample conditional skewness*.
(iv) Compute $f_{\alpha,\beta}(\gamma_\alpha)$, the adjustment factor
(v) Compute $\hat{\mathrm{ES}}_\beta(X)$, the adjusted tail-based ES estimator

4 AA ES estimator

The Arithmetic Average (AA) ES estimator will be compared with the *Adjusted Tail-based Normal Approximation* above through several metrics.
The AA ES estimator is defined in the following manner: [2]
Let $y_1, \cdots, y_N$ be a sample of losses. $\tilde{ES}_\beta = \frac{\sum_{n=\lceil N\beta \rceil}^N Y_{(n)}}{N+1-\lceil N\beta \rceil}$ is the AA ES estimator.

Note that in the paper, the Extreme value theory expected shortfall estimator is also compared but this project will leave that out due to time constraints.

5 Comparison: Adjusted Tail-based vs AA

For the comparison, 3 metrics will be compared - the MSE, variance, and bias. This is done through Monte Carlo simulation.
Let *M* be the number of replications and *n* be the number of observations *per sample.*

Then, letting $\mathrm{ES}_\beta(i)$ be either the i-th ES_β replication for either Adjusted Tail-based or AA estimator ($1 i M$), then we define the metrics as follows: [2]

(i) $\text{MSE of ES}_\beta = \frac{1}{M}\sum_{i=1}^{M}(\text{ES}_\beta(i) - \text{true ES}_\beta)^2$

(ii) $\text{variance of ES}_\beta = \frac{1}{M}\sum_{i=1}^{M}[\text{ES}_\beta(i) - \frac{1}{M}\sum_{i=1}^{M}\text{ES}_\beta(i)]^2$

(iii) $\text{bias of ES}_\beta = \frac{1}{M}\sum_{i=1}^{M}\text{ES}_\beta(i) - \text{true ES}_\beta$

6 Replication

In this section, I will go over the code for my replication (see Appendix).

- A_(alpha, y) is a function taking as inputs α, and our sample $y = (y_1, \ldots, y_n)$ and outputs $A_\alpha = (\lfloor N\alpha \rfloor + 1 - N\alpha)y_{(\lfloor N\alpha \rfloor)} + (N\alpha - \lfloor N\alpha \rfloor)y_{(\lfloor N\alpha \rfloor + 1)}$, the α -th quantile

- s2_(alpha, y, A) is a function taking as inputs α, our sample $y = (y_1, \ldots, y_n)$ and A_α. It then outputs our estimator $\hat{\sigma}^2 = [z_\alpha + 1 - \frac{z_\alpha}{(1-\alpha)\sqrt{2\pi}}e^{-z_\alpha^2/2}]^{-1}[\frac{\sum_{n=1}^{N}(y_n - A_\alpha)^2 1_{y_n > A_\alpha}}{\sum_{n=1}^{N} 1_{y_n > A_\alpha}}]$

- mu_(alpha, A, s2) is a function taking as inputs α, A_α, and $\hat{\sigma}^2$. It then outputs $\hat{\mu} = A_\alpha - \hat{\sigma}z_\alpha$

- skew_(y, A) is a function taking as inputs $y = (y_1, \ldots, y_n)$ and A_α. It then outputs the sample conditional skewness $\gamma_\alpha = (\frac{\sum_{n=1}^{N}(y_n - A_\alpha)^3 1_{y_n > A_\alpha}}{\sum_{n=1}^{N} 1_{y_n > A_\alpha}})[\frac{\sum_{n=1}^{N}(y_n - A_\alpha)^2 1_{y_n > A_\alpha}}{\sum_{n=1}^{N} 1_{y_n > A_\alpha}}]^{-1.5}$

- R_(skew, beta) is a taking as inputs the sample conditional skewness γ_α, and β. It then outputs the ratio $\hat{R}_{\alpha,\beta} = f_{\alpha,\beta}(\gamma_\alpha) = b_0 + b_1 e^{-b_2 y_\alpha} + b_3\gamma_\alpha^{-1} + b_4\gamma_\alpha^{-2}$

- ES_adj_(mu, s2, beta, A, R) is a function taking as inputs $\hat{\mu}, \hat{\sigma}^2, \beta, A_\alpha, \text{and } \hat{R}_{\alpha,\beta}$. It then outputs the adjusted tail-based ES estimator $\hat{\text{ES}}_\beta(X) = [\text{ES}_\beta(X) - A_\alpha]f_{\alpha,\beta}(\gamma_\alpha) + A_\alpha$

- ES_W_(beta, distr, df, shape, mu, s, scale, xi) is a function taking as inputs β, $\text{distr} \in$ {*t, gamma, lognormal, gpd, weibull}*, and also takes optional parameters *df, shape,* μ, σ, *scale*, and ξ which should be inputted depending on the distribution chosen. It then outputs the ES of the distribution at level β.
    For the GPD and Weibull distributions, ES can be expressed through
closed-form equations from the paper [1].

$Y \sim Weibull(\lambda, k),\ then\ \mathrm{ES}_\beta = \frac{\lambda}{1-\alpha}\Gamma_U(1+\frac{1}{k}, -ln(1-\alpha))\ where\ \Gamma_U(a,b) = \int_b^\infty p^{\alpha-1}e^{-p}dp$

$Y \sim GPD(\mu, s, \xi)\ with\ \xi \neq 0\ ,\ then\ \mathrm{ES}_\beta = \mu + s[\frac{(1-\alpha)^{-\xi}}{1-\xi} + \frac{(1-\alpha)^{-\xi}-1}{\xi}$

- AA_(beta, y) is a function taking as inputs β and $y = (y_1, \ldots, y_n)$ and outputs the Arithmetic Average ES estimator $\tilde{ES}_\beta = \frac{\sum_{n=\lceil N\beta\rceil}^{N} Y_{(n)}}{N+1-\lceil N\beta\rceil}$

- calculate_ES_adj(alpha, beta, y) is a function taking as inputs α, β, and $y = (y_1, \ldots, y_n)$. It then runs through all the functions mentioned above to obtain A_α, $\hat{\sigma}^2$, $\hat{\mu}$, γ_α, $\hat{R}_{\alpha,\beta}$, and finally returns the adjusted tail-based ES estimator $\hat{\mathrm{ES}}_\beta(X) = [\mathrm{ES}_\beta(X) - A_\alpha]f_{\alpha,\beta}(\gamma_\alpha) + A_\alpha$

Now, for the Monte Carlo simulations. To illustrate the process, suppose *n = 250* and $y = (y_1, \ldots, y_{250})$ is drawn from a *t*-distribution with df=3.5. (*i.e. y = rt(n=250, df=3.5)*)
Then, $\mathrm{ES}_\beta(W)$ is calculated (the ES of *t*-distributed random variable with df=3.5 at level β.
After that both $\hat{\mathrm{ES}}_\beta$ (the adjusted tail-based ES estimator) and *AA* (the Arithmetic Average ES estimator) are calculated.
The 3 metrics (MSE, variance, and bias) are then calculated for both $\hat{\mathrm{ES}}_\beta$ and AA.
The goal is then to compare these metrics and see which of $\hat{\mathrm{ES}}_\beta$ and AA yields the lowest MSE, variance, and bias.

I first created the function print_t(M), taking as input *M,* the number of Monte Carlo simulations. This function looks at the *t*-distribution with $df \in \{3.5, 5, 8\}$, $n \in \{250, 500\}$ and $\beta \in \{0.99, 0.995\}$ and prints out the MSE, variance, and bias of both the $\hat{\mathrm{ES}}_\beta$ and AA ES estimators through a Monte Carlo simulation of *M* replications.
Note that the seed is reset at 001 for each replication. This is so that each of the *M* replications of *rt(n=n, df=df)* are identical so that we are comparing the estimators on the same data.

I then created the function print_gpd(M), taking as input *M,* the number of Monte Carlo simulations. This function looks at the GPD(*location=0, scale=1*)-distribution with shape parameter $\xi \in \{0.3, 0.2, 0.1\}$, $n \in \{250, 500\}$ and $\beta \in \{0.99, 0.995\}$ and prints out the MSE, variance, and bias of both the $\hat{\mathrm{ES}}_\beta$ and AA ES estimators through a Monte Carlo simulation of M replications.

I then created the function print_weibull(M), taking as input *M*, the number of Monte Carlo simulations. This function looks at the *Weibull (shape, scale=1)*-distribution with shape parameter $shape \in \{0.6, 0.9, 1.4\}$, $n \in \{250, 500\}$ and $\beta \in \{0.99, 0.995\}$ and prints out the MSE, variance, and bias of both the $\hat{\mathrm{ES}}_\beta$ and AA ES estimators through a Monte Carlo simulation of *M* replications.

I then created the function print_t2(M), taking as input *M*, the number of Monte Carlo simulations. This function looks at the *t*-distribution with $df \in \{2.5, 3\}$, $n \in \{250, 500\}$ and $\beta \in \{0.99, 0.995\}$ and prints out the MSE, variance, and bias of both the $\hat{\mathrm{ES}}_\beta$ and AA ES estimators through a Monte Carlo simulation of *M* replications.

Finally, I created the function print_gpd2(M), taking as input *M*, the number of Monte Carlo simulations. This function looks at the GPD(*location=0, scale=1*)-distribution with shape parameter $\xi \in \{0.5, 0.35\}$, $n \in \{250, 500\}$ and $\beta \in \{0.99, 0.995\}$ and prints out the MSE, variance, and bias of both the $\hat{\mathrm{ES}}_\beta$ and AA ES estimators through a Monte Carlo simulation of M replications.

7 Results

In this section, the functions *print_t(2500)*, *print_gpd(2500), print_weibull(2500), print_t2(2500), and print_gpd2(2500)* are run in order to print out the Monte Carlo estimate of 2500 replications of MSE, variance, and bias for the *t*, *Generalized Pareto Distribution (GPD),* and *Weibull* distributed loss samples.
In my replications, after the results are printed out in the R Console, they are carefully reorganized into the tables below. For each set of parameters, the lowest of $\hat{\mathrm{ES}}$ and AA is bolded for every metric.
The fundamental question is: is MSE, variance, and bias lower for $\hat{\mathrm{ES}}$ as compared with AA?

Note that error estimates should also be considered in the simulations. This is important since it allows us to see if the differences between the source paper and my replication results are statistically significant, and whether or not the differences between ES and AA results are statistically significant. Due to time constraints, this was not calculated. The source paper also did not calculate any error estimates which can be considered a flaw in their methodology.

7.1 t-Distribution

The results from the paper are listed in the table below below for samples from the *t*-distribution with *df = 3.5, 5, 8* and $\beta = 0.99, 0.995$ with *n = 250.*

Source Paper (*t*-distribution) n = 250

n = 250		MSE $\beta = 0.99$	MSE $\beta = 0.995$	Var $\beta = 0.99$	Var $\beta = 0.995$	Bias $\beta = 0.99$	Bias $\beta = 0.995$
t, df = 3.5	$\hat{\mathrm{ES}}$	**2.514**	**5.243**	**2.280**	**3.897**	-0.484	-1.160
	AA	3.080	7.426	3.053	6.968	**-0.162**	**-0.676**
t, df = 5	$\hat{\mathrm{ES}}$	**0.949**	**1.821**	**0.901**	**1.521**	-0.219	-0.548
	AA	1.214	2.687	1.213	2.616	**-0.028**	**-0.266**
t, df = 8	$\hat{\mathrm{ES}}$	**0.356**	**0.645**	**0.340**	**0.552**	-0.126	-0.305
	AA	0.445	0.910	0.445	0.888	**-0.01**	**-0.147**

Results show that $\hat{\mathrm{ES}}$ outperforms AA in MSE and variance and only falls short in bias.
The table below shows a summary of the earlier results. As MSE is a very important metric in this experiment, it is bolded.

Source Paper (*t*-distribution) n = 250

	MSE	Var	Bias
$\hat{\mathrm{ES}}$	**6**	6	0
AA	**0**	0	6

In the paper, $\hat{\mathrm{ES}}$ has lower MSE for all 6 combinations of parameters and is massively better than AA.

The results from my replications are listed in the tables below.

My Replications (*t*-distribution) n = 250

n = 250		MSE $\beta = 0.99$	MSE $\beta = 0.995$	Var $\beta = 0.99$	Var $\beta = 0.995$	Bias $\beta = 0.99$	Bias $\beta = 0.995$
t, df = 3.5	$\hat{\mathrm{ES}}$	3.513701	6.918576	3.349879	5.931118	**-0.4064011**	**-0.9949018**
	AA	**3.045859**	**6.618685**	**2.757139**	**4.895043**	-0.5383521	-1.313621
t, df = 5	$\hat{\mathrm{ES}}$	0.9119978	**1.772814**	0.8458146	1.465247	**-0.2579177**	**-0.5551157**

	AA	**0.896156**	1.88101	**0.7838646**	**1.292381**	-0.3355666	-0.7675588
t, df = 8	$\hat{\mathrm{ES}}$	**0.3733118**	**0.6882679**	0.3424361	0.5728771	**-0.1761041**	**-0.3400294**
	AA	0.3821088	0.7564698	**0.3286057**	**0.5127364**	-0.2315912	-0.4939012

In the summary table, it is shown that $\hat{\mathrm{ES}}$ had the lower MSE for 3 of the parameter combinations while AA had the lower MSE for the 3 others - effectively a tie. Note that the AA all had lower variance while $\hat{\mathrm{ES}}$ all had lower bias.

Recall that $\hat{\mathrm{ES}}$ beats out AA 6-0 in MSE in the source paper replications which is inconsistent with my results.
Furthermore, in the source paper, $\hat{\mathrm{ES}}$ beats out AA 6-0 in variance while AA beats out $\hat{\mathrm{ES}}$ 6-0 in bias. But in my replications, AA beats out $\hat{\mathrm{ES}}$ 6-0 in variance while $\hat{\mathrm{ES}}$ beats out AA in bias. These are completely flipped!

My Replications (*t*-distribution) n = 250

	MSE	Var	Bias
$\hat{\mathrm{ES}}$	**3**	0	6
AA	**3**	6	0

Now moving on to *n = 500*.

Source Paper (*t*-distribution) n = 500

n = 500		MSE $\beta = 0.99$	MSE $\beta = 0.995$	Var $\beta = 0.99$	Var $\beta = 0.995$	Bias $\beta = 0.99$	Bias $\beta = 0.995$
t, df = 3.5	$\hat{\mathrm{ES}}$	1.683	**3.561**	1.606	**2.970**	-0.279	-0.769
	AA	**1.511**	4.904	**1.449**	4.878	**-0.248**	**-0.161**
t, df = 5	$\hat{\mathrm{ES}}$	0.533	**1.068**	**0.517**	**0.946**	-0.129	-0.348
	AA	**0.530**	1.583	0.518	1.583	**-0.108**	**-0.002**
t, df = 8	$\hat{\mathrm{ES}}$	**0.193**	**0.366**	**0.187**	**0.331**	-0.077	-0.187
	AA	0.201	0.521	0.197	0.521	**-0.060**	**-0.001**

	MSE	Var	Bias
$\hat{\mathrm{ES}}$	**4**	5	0
AA	**2**	1	6

Looking at the results from the paper, $\hat{\mathrm{ES}}$ still generally outperforms AA. In this case, for MSE, $\hat{\mathrm{ES}}$ beats out AA by 4-2. Similarly as in source paper results for *n=250*, $\hat{\mathrm{ES}}$ beats out AA for variance but AA beats out $\hat{\mathrm{ES}}$ for bias.

My Replications (*t*-distribution) n = 500

n = 500		MSE $\beta = 0.99$	MSE $\beta = 0.995$	Var $\beta = 0.99$	Var $\beta = 0.995$	Bias $\beta = 0.99$	Bias $\beta = 0.995$
t, df = 3.5	$\hat{\mathrm{ES}}$	2.32565	4.622969	2.311198	4.324478	**0.1239996**	**0.547924**
	AA	**1.683792**	**4.569174**	**1.514917**	**4.20693**	0.4116804	0.6032632
t, df = 5	$\hat{\mathrm{ES}}$	0.548986	**1.095874**	0.5352389	**0.9876675**	**0.1181574**	**0.329547**
	AA	**0.4980796**	1.152373	**0.42303**	1.005189	0.2742606	0.3841695
t, df = 8	$\hat{\mathrm{ES}}$	**0.1927528**	**0.3692167**	0.1858955	**0.3318638**	**-0.08325665**	**-0.1936121**
	AA	0.1933235	0.4038955	**0.1597503**	0.3446732	-0.183404	-0.2436394

	MSE	Var	Bias
$\hat{\mathrm{ES}}$	**3**	2	6
AA	**3**	4	0

In my replications, it's again a tie (3-3) between $\hat{\mathrm{ES}}$ and AA with regards to MSE. AA beats $\hat{\mathrm{ES}}$ for variance (4-2) while $\hat{\mathrm{ES}}$ beats AA for bias (6-0). Recall that in the source paper, $\hat{\mathrm{ES}}$ beats AA by 4-2 in MSE, with $\hat{\mathrm{ES}}$ beating AA (5-1) for variance and AA beating $\hat{\mathrm{ES}}$ (6-0) for bias.

Again, $\hat{\mathrm{ES}}$ performs better in the source paper compared to my replications. It should be noted that the variance and bias results are again flipped between source paper results and my replications. In the paper, $\hat{\mathrm{ES}}$ performs well with regards to variance and AA with regards to bias but in my replications, the opposite occurs.

7.2 Weibull Distribution

Now, onto the Weibull (*shape, scale=1*) distribution

Source Paper Weibull (*shape, scale=1*) n = 250

n = 250		MSE $\beta = 0.99$	MSE $\beta = 0.995$	Var $\beta = 0.99$	Var $\beta = 0.995$	Bias $\beta = 0.99$	Bias $\beta = 0.995$
shape=0.6	$\hat{\mathrm{ES}}$	**19.342**	**37.144**	**17.652**	**29.131**	-1.300	-2.831
	AA	24.114	50.042	23.998	47.898	**-0.341**	**-1.464**
shape=0.9	$\hat{\mathrm{ES}}$	**1.249**	**2.248**	**1.168**	**1.881**	-0.285	-0.606
	AA	1.576	3.005	1.574	2.914	**-0.049**	**-0.302**
shape=1.4	$\hat{\mathrm{ES}}$	**0.131**	**0.226**	**0.124**	**0.195**	-0.084	-0.176
	AA	0.164	0.297	0.164	0.289	**-0.014**	**-0.091**

Source Paper Weibull (*shape, scale=1*) n = 250

	MSE	Var	Bias
$\hat{\mathrm{ES}}$	**6**	6	0
AA	**0**	0	6

In the paper, $\hat{\mathrm{ES}}$ has lower MSE for all 6 combinations of parameters, beating AA by 6-0. Note that $\hat{\mathrm{ES}}$ beats AA 6-0 in variance, and vice versa for bias. This is identical to the source paper results for *t*-distribution for *n=250.*

My Replications Weibull (*shape, scale=1*) n = 250

n = 250		MSE $\beta = 0.99$	MSE $\beta = 0.995$	Var $\beta = 0.99$	Var $\beta = 0.995$	Bias $\beta = 0.99$	Bias $\beta = 0.995$
shape=0.6	$\hat{\mathrm{ES}}$	21.13662	**40.22232**	19.36554	32.91464	**-1.333727**	**-2.705706**
	AA	**20.8001**	42.5514	**18.02868**	**28.79762**	-1.666921	-3.710162
shape=0.9	$\hat{\mathrm{ES}}$	**1.249634**	**2.257493**	1.12784	1.846376	**-0.3496358**	**-0.6417602**
	AA	1.306666	2.532453	**1.100847**	**1.638437**	-0.4541577	-0.9458706
shape=1.4	$\hat{\mathrm{ES}}$	**0.127255**	**0.2189877**	0.1145921	0.1810451	**-0.1127329**	**-0.1949746**
	AA	0.1376682	0.2554103	**0.1137437**	**0.1608685**	-0.1548224	-0.3075811

My Replications Weibull (*shape, scale=1*) n = 250

	MSE	Var	Bias
$\hat{\mathrm{ES}}$	**5**	0	6
AA	**1**	6	0

In my replications, $\hat{\mathrm{ES}}$ beats AA 5-1 with regards to MSE. This is similar to the source paper result of 6-0 but $\hat{\mathrm{ES}}$ still performs slightly worse here.
Note however, AA beats $\hat{\mathrm{ES}}$ 6-0 in variance and vice versa for bias. This is again completely flipped from the source paper results!

Source Paper Weibull (*shape, scale=1*) n = 500

n = 500		MSE $\beta = 0.99$	MSE $\beta = 0.995$	Var $\beta = 0.99$	Var $\beta = 0.995$	Bias $\beta = 0.99$	Bias $\beta = 0.995$
shape=0.6	$\hat{\mathrm{ES}}$	**10.999**	**21.795**	**10.273**	**18.209**	-0.852	-1.894
	AA	11.155	29.882	10.720	29.822	**-0.659**	**-0.245**
shape=0.9	$\hat{\mathrm{ES}}$	**0.660**	**1.221**	**0.621**	**1.063**	-0.197	-0.396
	AA	0.702	1.675	0.678	1.674	**-0.154**	**-0.032**
shape=1.4	$\hat{\mathrm{ES}}$	**0.067**	**0.117**	**0.064**	**0.105**	-0.058	-0.110
	AA	0.073	0.157	0.071	0.157	**-0.049**	**-0.009**

Source Paper Weibull (*shape, scale=1*) n = 500

	MSE	Var	Bias
$\hat{\mathrm{ES}}$	**6**	6	0
AA	**0**	0	6

In the source paper results, $\hat{\mathrm{ES}}$ beats AA 6-0 for MSE. In addition, $\hat{\mathrm{ES}}$ beats AA for variance and vice versa for bias. This is identical to the source paper results for *n=250.*

My Replications Weibull (*shape, scale=1*) n = 500

n = 500		MSE $\beta = 0.99$	MSE $\beta = 0.995$	Var $\beta = 0.99$	Var $\beta = 0.995$	Bias $\beta = 0.99$	Bias $\beta = 0.995$
shape=0.6	$\hat{\mathrm{ES}}$	11.74906	**23.06104**	11.26571	**20.20684**	**-0.6984654**	**-1.691829**
	AA	**11.25699**	24.61849	**9.428502**	21.14317	-1.353611	-1.866489
shape=0.9	$\hat{\mathrm{ES}}$	**0.6679525**	**1.240388**	0.635357	**1.09569**	**-0.1812447**	**-0.3809666**

	AA	0.7109973	1.378274	**0.580438**	1.155482	-0.361651	-0.4724979
shape=1.4	$\hat{\text{ES}}$	**0.06647344**	**0.1170228**	0.06338532	**0.1050339**	**-0.05579845**	**-0.1096854**
	AA	0.0748008	0.1327834	**0.06002494**	0.1094981	-0.1216547	-0.1527384

My Replications Weibull (*shape, scale=1*) n = 500

	MSE	Var	Bias
$\hat{\text{ES}}$	**5**	3	6
AA	**1**	3	0

In my replications, $\hat{\text{ES}}$ beats AA by 5-1 for MSE. This is similar to the source paper result of 6-0 but $\hat{\text{ES}}$ still performs slightly worse here.
In addition, $\hat{\text{ES}}$ is tied with AA 3-3 with regards to variance but $\hat{\text{ES}}$ beats AA 6-0 for bias. Again, the results for variance and bias differ dramatically from that in the source paper.

7.3 Generalized Pareto Distribution

Now, we turn our attention to the Generalized Pareto Distribution with *location = 0* and *scale = 1*.

Source Paper Generalized Pareto Distribution ($\xi = 0.3, 0.2, 0.1$) n = 250

n = 250		MSE $\beta = 0.99$	MSE $\beta = 0.995$	Var $\beta = 0.99$	Var $\beta = 0.995$	Bias $\beta = 0.99$	Bias $\beta = 0.995$
$\xi = 0.3$	$\hat{\text{ES}}$	**21.634**	**49.003**	**17.271**	**29.052**	-2.089	-4.467
	AA	24.628	59.532	23.204	50.409	**-1.193**	**-3.021**
$\xi = 0.2$	$\hat{\text{ES}}$	**7.246**	**14.638**	**6.518**	**10.995**	-0.853	-1.909
	AA	8.947	20.088	8.857	18.984	**-0.300**	**-1.050**
$\xi = 0.1$	$\hat{\text{ES}}$	**2.207**	**4.183**	**2.037**	**3.368**	-0.412	-0.903
	AA	2.772	5.752	2.762	5.542	**-0.097**	**-0.458**

Source Paper Generalized Pareto Distribution ($\xi = 0.3, 0.2, 0.1$) n = 250

	MSE	Var	Bias
$\hat{\text{ES}}$	**6**	6	0
AA	**0**	0	6

In the source paper results, $\hat{\mathrm{ES}}$ beats AA 6-0 for MSE. In addition, $\hat{\mathrm{ES}}$ beats AA for variance 6-0 and vice versa for bias. This is identical to the source paper results for the *Weibull* distribution.

My Replications Generalized Pareto Distribution ($\xi = 0.3, 0.2, 0.1$) n = 250

n = 250		MSE $\beta = 0.99$	MSE $\beta = 0.995$	Var $\beta = 0.99$	Var $\beta = 0.995$	Bias $\beta = 0.99$	Bias $\beta = 0.995$
$\xi = 0.3$	$\hat{\mathrm{ES}}$	38.5636	75.88232	37.04943	65.80364	**-1.236523**	**-3.178836**
	AA	**31.10906**	**69.26451**	**28.36118**	**52.39458**	-1.661093	-4.109853
$\xi = 0.2$	$\hat{\mathrm{ES}}$	8.775228	**17.12799**	8.216385	14.30545	**-0.7497529**	**-1.681746**
	AA	**8.017563**	17.15953	**7.086977**	**12.14669**	-0.9661373	-2.240021
$\xi = 0.1$	$\hat{\mathrm{ES}}$	2.298071	**4.326176**	2.10109	3.535198	**-0.4447709**	**-0.890164**
	AA	**2.290295**	4.661933	**1.964367**	**3.119049**	-0.5715881	-1.242631

My Replications Generalized Pareto Distribution ($\xi = 0.3, 0.2, 0.1$) n = 250

	MSE	Var	Bias
$\hat{\mathrm{ES}}$	**2**	0	6
AA	**4**	6	0

In my replication, AA beats $\hat{\mathrm{ES}}$ 4-2 for MSE. This is surprising considering that $\hat{\mathrm{ES}}$ beats AA 6-0 in the source paper results.

In addition, AA beats $\hat{\mathrm{ES}}$ 6-0 for variance and vice versa for bias. This is identical to my replication results for Weibull (*shape, scale=1*) n = 250 and t-distribution n=250. However, these results for variance and bias are again completely flipped compared to the source paper results.

Source Paper Generalized Pareto Distribution ($\xi = 0.3, 0.2, 0.1$) n=500

n = 500		MSE $\beta = 0.99$	MSE $\beta = 0.995$	Var $\beta = 0.99$	Var $\beta = 0.995$	Bias $\beta = 0.99$	Bias $\beta = 0.995$
$\xi = 0.3$	$\hat{\mathrm{ES}}$	15.195	**34.398**	13.513	**24.882**	-1.297	-3.085
	AA	**13.469**	42.283	**11.991**	40.732	**-1.216**	**-1.245**
$\xi = 0.2$	$\hat{\mathrm{ES}}$	4.306	**8.990**	4.010	**7.302**	-0.544	-1.299
	AA	**4.124**	12.175	**3.910**	12.099	**-0.463**	**-0.275**

$\xi = 0.1$	$\hat{\text{ES}}$	**1.264**	**2.471**	**1.196**	**2.135**	-0.259	-0.580
	AA	1.286	3.499	1.244	3.497	**-0.204**	**-0.044**

Source Paper Generalized Pareto Distribution ($\xi = 0.3, 0.2, 0.1$) **n=500**

	MSE	Var	Bias
$\hat{\text{ES}}$	**4**	4	0
AA	**2**	2	6

In the source paper, $\hat{\text{ES}}$ beats AA 4-2 for MSE. In addition, $\hat{\text{ES}}$ beats AA 4-2 for variance while AA beats $\hat{\text{ES}}$ 6-0 in bias.

My Replications Generalized Pareto Distribution ($\xi = 0.3, 0.2, 0.1$) **n=500**

n = 500		MSE $\beta = 0.99$	MSE $\beta = 0.995$	Var $\beta = 0.99$	Var $\beta = 0.995$	Bias $\beta = 0.99$	Bias $\beta = 0.995$
$\xi = 0.3$	$\hat{\text{ES}}$	29.75435	59.14329	29.56326	55.50351	**-0.4504673**	**-1.913632**
	AA	**18.02656**	**51.4315**	**16.05085**	**47.03413**	-1.407881	-2.101471
$\xi = 0.2$	$\hat{\text{ES}}$	5.718911	11.47192	5.594679	10.39187	**-0.3556264**	**-1.041253**
	AA	**4.551092**	**11.32327**	**3.893655**	**9.993395**	-0.8117846	-1.154933
$\xi = 0.1$	$\hat{\text{ES}}$	1.344728	**2.624374**	1.291502	**2.324945**	**-0.2318252**	**-0.5480504**
	AA	**1.288242**	2.796771	**1.063535**	2.38474	-0.4744814	-0.6426391

My Replications Generalized Pareto Distribution ($\xi = 0.3, 0.2, 0.1$) **n=500**

	MSE	Var	Bias
$\hat{\text{ES}}$	**1**	1	6
AA	**5**	5	0

For my replications, AA beats $\hat{\text{ES}}$ 5-1. This is surprising because considering that $\hat{\text{ES}}$ beats AA 6-0 in the source paper results.
In addition, AA beats $\hat{\text{ES}}$ 5-1 for variance and $\hat{\text{ES}}$ beats AA 6-0 in bias. This is again more or less flipped compared to the source paper results.

7.4 Miscellaneous (i.e. distributions without 3rd moment)

In this section, distributions where 3rd moments do not exist are looked at. In the paper, the authors note that for distributions such as *t* (df=2.5,3) and GPD ($\xi = 0.35, 0.5$), 3rd moments do not exist. However, they exist in sample estimates although they can be very large. Here, source paper results and my replications for *t* (df=2.5,3) and GPD ($\xi = 0.35, 0.5$) will be looked at.

Source Paper *t*-distribution n = 250

n = 250		MSE $\beta = 0.99$	MSE $\beta = 0.995$	Var $\beta = 0.99$	Var $\beta = 0.995$	Bias $\beta = 0.99$	Bias $\beta = 0.995$
df = 2.5	$\hat{\text{ES}}$	**9.237**	**22.882**	**6.749**	**11.310**	-1.577	-3.402
	AA	10.199	25.784	9.125	18.964	**-1.036**	**-2.611**
df = 3	$\hat{\text{ES}}$	**4.290**	**9.646**	**3.588**	**6.120**	-0.838	-1.878
	AA	5.010	12.508	4.808	10.903	**-0.449**	**-1.267**

Source Paper *t*-distribution n = 250

	MSE	Var	Bias
$\hat{\text{ES}}$	**4**	4	0
AA	**0**	0	4

In the source paper, $\hat{\text{ES}}$ beats AA 4-0 in MSE. In addition, $\hat{\text{ES}}$ beats AA 4-0 for variance and vice versa for bias.

My Replications *t*-distribution n = 250

n = 250		MSE $\beta = 0.99$	MSE $\beta = 0.995$	Var $\beta = 0.99$	Var $\beta = 0.995$	Bias $\beta = 0.99$	Bias $\beta = 0.995$
df = 2.5	$\hat{\text{ES}}$	24.17287	46.92687	23.67937	42.43688	**-0.7092038**	**-2.122961**
	AA	**17.40103**	**39.55924**	**16.32441**	**32.09198**	-1.040747	-2.734978
df = 3	$\hat{\text{ES}}$	8.11987	15.89164	7.840544	13.96987	**-0.5314722**	**-1.388295**
	AA	**6.538331**	**14.35799**	**6.033168**	**11.16446**	-0.7124439	-1.788295

My Replications *t*-distribution n = 250

	MSE	Var	Bias
$\hat{\text{ES}}$	**0**	0	4
AA	**4**	4	0

In my replications, the results are the complete opposite of the source paper results.

Source Paper *t*-distribution n = 500

n = 250		MSE $\beta = 0.99$	MSE $\beta = 0.995$	Var $\beta = 0.99$	Var $\beta = 0.995$	Bias $\beta = 0.99$	Bias $\beta = 0.995$
df = 2.5	$\hat{\mathrm{ES}}$	6.570	**16.347**	5.414	**9.980**	-1.076	-2.523
	AA	**5.664**	18.168	**4.565**	16.193	**-1.048**	**-1.405**
df = 3	$\hat{\mathrm{ES}}$	2.774	**6.430**	2.459	**4.574**	-0.561	-1.362
	AA	**2.439**	7.894	**2.161**	7.548	**-0.528**	**-0.588**

Source Paper *t*-distribution n = 500

	MSE	Var	Bias
$\hat{\mathrm{ES}}$	**2**	2	0
AA	**2**	2	4

In the source paper, $\hat{\mathrm{ES}}$ and AA are tied 2-2 in MSE. In addition, $\hat{\mathrm{ES}}$ and AA are tied 2-2 for variance and AA beats $\hat{\mathrm{ES}}$ 4-0 in bias.

My Replications *t*-distribution n = 500

n = 500		MSE $\beta = 0.99$	MSE $\beta = 0.995$	Var $\beta = 0.99$	Var $\beta = 0.995$	Bias $\beta = 0.99$	Bias $\beta = 0.995$
df = 2.5	$\hat{\mathrm{ES}}$	26.92445	51.72448	26.93142	50.50161	**-0.06163187**	**-1.11493**
	AA	**11.45585**	**37.94064**	**10.74524**	**36.26734**	-0.8455263	-1.299154
df = 3	$\hat{\mathrm{ES}}$	6.286682	12.35365	6.275689	11.81623	**-0.1162061**	**-0.7363079**
	AA	**3.624881**	**10.70422**	**3.320491**	**10.00695**	-0.5529182	-0.8374194

My Replications *t*-distribution n = 500

	MSE	Var	Bias
$\hat{\mathrm{ES}}$	**0**	0	4
AA	**4**	4	0

In my replications, AA beats $\hat{\mathrm{ES}}$ 4-0 for MSE, AA beats $\hat{\mathrm{ES}}$ 4-0 for variance, and $\hat{\mathrm{ES}}$ beats AA 4-0 for bias.

Here, AA performs significantly better than $\hat{\mathrm{ES}}$ but in the source paper, it's more or less a tie.

Source Paper Generalized Pareto Distribution ($\xi = 0.5, 0.35$) n = 250

n = 250		MSE $\beta = 0.99$	MSE $\beta = 0.995$	Var $\beta = 0.99$	Var $\beta = 0.995$	Bias $\beta = 0.99$	Bias $\beta = 0.995$
$\xi = 0.5$	$\hat{\text{ES}}$	259.758	**753.222**	**132.640**	**223.358**	-11.275	-23.019
	AA	**257.813**	755.729	176.495	388.803	**-9.018**	**-19.155**
$\xi = 0.35$	$\hat{\text{ES}}$	**38.150**	**92.495**	**27.710**	**46.481**	-3.231	-6.783
	AA	41.887	104.238	37.500	79.897	**-2.095**	**-4.934**

Source Paper Generalized Pareto Distribution ($\xi = 0.5, 0.35$) n = 250

	MSE	Var	Bias
$\hat{\text{ES}}$	**3**	4	0
AA	**1**	0	4

In the source paper, $\hat{\text{ES}}$ beats AA 3-1 for MSE. In addition, $\hat{\text{ES}}$ beats AA 4-0 for variance and vice versa for bias.

My Replications Generalized Pareto Distribution ($\xi = 0.5, 0.35$) n = 250

n = 250		MSE $\beta = 0.99$	MSE $\beta = 0.995$	Var $\beta = 0.99$	Var $\beta = 0.995$	Bias $\beta = 0.99$	Bias $\beta = 0.995$
$\xi = 0.5$	$\hat{\text{ES}}$	1215.873	2306.649	1206.239	2168.002	**-3.180546**	**-11.81159**
	AA	**730.968**	**1691.3**	**702.3057**	**1465.593**	-5.379887	-15.04305
$\xi = 0.35$	$\hat{\text{ES}}$	85.8819	168.154	83.435	149.0199	**-1.574889**	**-4.381058**
	AA	**64.38802**	**145.5031**	**59.58475**	**114.0284**	-2.197068	-5.614301

My Replications Generalized Pareto Distribution ($\xi = 0.5, 0.35$) n = 250

	MSE	Var	Bias
$\hat{\text{ES}}$	**0**	0	4
AA	**4**	4	0

In my replications, AA beats $\hat{\text{ES}}$ 4-0 for MSE. In addition, AA beats $\hat{\text{ES}}$ 4-0 for variance and vice versa for bias.

Source Paper Generalized Pareto Distribution ($\xi = 0.5, 0.35$) **n = 500**

n = 250		MSE $\beta = 0.99$	MSE $\beta = 0.995$	Var $\beta = 0.99$	Var $\beta = 0.995$	Bias $\beta = 0.99$	Bias $\beta = 0.995$
$\xi = 0.5$	$\hat{\mathrm{ES}}$	259.758	**753.222**	**132.640**	**223.358**	-11.275	-23.019
	AA	**257.813**	755.729	176.495	388.803	**-9.018**	**-19.155**
$\xi = 0.35$	$\hat{\mathrm{ES}}$	**38.150**	**92.495**	**27.710**	**46.481**	-3.231	-6.783
	AA	41.887	104.238	37.500	79.897	**-2.095**	**-4.934**

Source Paper Generalized Pareto Distribution ($\xi = 0.5, 0.35$) **n = 500**

	MSE	Var	Bias
$\hat{\mathrm{ES}}$	**3**	4	0
AA	**1**	0	4

In the source paper, $\hat{\mathrm{ES}}$ beats AA 3-1 for MSE. In addition, $\hat{\mathrm{ES}}$ beats AA 4-0 for variance and vice versa for bias.

My Replications Generalized Pareto Distribution ($\xi = 0.5, 0.35$) **n = 500**

n = 500		MSE $\beta = 0.99$	MSE $\beta = 0.995$	Var $\beta = 0.99$	Var $\beta = 0.995$	Bias $\beta = 0.99$	Bias $\beta = 0.995$
$\xi = 0.5$	$\hat{\mathrm{ES}}$	1529.252	2890.392	1529.285	2857.976	**0.7606934**	**-5.793081**
	AA	**507.1064**	**1817.774**	**486.7762**	**1763.332**	-4.530439	-7.42612
$\xi = 0.35$	$\hat{\mathrm{ES}}$	73.68382	144.5934	73.52725	138.0959	**-0.4312581**	**-2.559823**
	AA	**38.10807**	**116.1246**	**34.63903**	**108.0333**	-1.866251	-2.852118

My Replications Generalized Pareto Distribution ($\xi = 0.5, 0.35$) **n = 500**

	MSE	Var	Bias
$\hat{\mathrm{ES}}$	**0**	0	4
AA	**4**	4	0

In our replications, AA beats $\hat{\mathrm{ES}}$ 4-0 for MSE. In addition, AA beats $\hat{\mathrm{ES}}$ 4-0 for variance and vice versa for bias. Again, variance and bias is flipped between source paper and my replications.

7.5 Comparison

In the paper, $\hat{\mathrm{ES}}$ massively outperformed AA (and EVT) estimators. We were unable to reproduce this through our simulations. In our simulations, AA actually performed much better than $\hat{\mathrm{ES}}$.

The sum over all distributions from 7.1 - 7.4 is now considered for the source paper and for my replications.

Source Paper [Sum over all distributions 7.1 - 7.4]

	MSE	Var	Bias
$\hat{\mathrm{ES}}$	**44 (84.62%)**	47 (90.38%)	0 (0%)
AA	**8 (15.38%)**	5 (9.62%)	52 (100%)

In the source paper, $\hat{\mathrm{ES}}$ outperformed AA by 44-8. In addition, $\hat{\mathrm{ES}}$ outperformed AA 47-5 for variance and AA outperformed $\hat{\mathrm{ES}}$ 52-0 for bias. In nearly all the sections, $\hat{\mathrm{ES}}$ significantly outperformed AA, often by 6-0.

My Replications [Sum over all distributions 7.1 - 7.4]

	MSE	Var	Bias
$\hat{\mathrm{ES}}$	**19 (36.54%)**	6 (11.54%)	52 (100%)
AA	**33 (63.46%)**	46 (88.46%)	0 (0%)

In my replications, AA outperformed $\hat{\mathrm{ES}}$ by 33-19. In addition, AA beat $\hat{\mathrm{ES}}$ 46-6 for variance and $\hat{\mathrm{ES}}$ beat AA 52-0 for bias.

In Section 7.1 (*t*-distribution df=3.5,5,8), $\hat{\mathrm{ES}}$ and AA performed around equally; in Section 7.2 (Weibull), $\hat{\mathrm{ES}}$ outperformed AA; and in Section 7.3 (Generalized Pareto $\xi = 0.3, 0.2, 0.1$) and 7.4 (Misc. *t* & Generalized Pareto), AA outperformed $\hat{\mathrm{ES}}$. Overall, AA definitely outperformed $\hat{\mathrm{ES}}$.

To gain more insights on my replication results, I plotted all the pdf's for all the distribution and parameter combinations on Desmos. The equations and graphs (see below) can be viewed here: https://www.desmos.com/calculator/fudnf7bhte.

The red lines denote the distribution/parameter combinations where $\hat{\mathrm{ES}}$ performed worse than AA. The orange lines denote the combinations where $\hat{\mathrm{ES}}$ were tied with AA. And the green lines denote the combinations where $\hat{\mathrm{ES}}$ outperformed AA. From the plot above, it appears that $\hat{\mathrm{ES}}$ outperformed AA when the tail was less heavy, and AA generally outperformed $\hat{\mathrm{ES}}$ when the distribution had heavier tails.

Note that for both the source paper and my replications, the bias was negative for both $\hat{\mathrm{ES}}$ and AA estimators; i.e. both the $\hat{\mathrm{ES}}$ and AA estimators both still underestimate the true ES. In fact, every single bias result in the tables in Sections 7.1-7.4 was negative, which shows that this is a major problem.

8 Conclusions

In this experiment, I looked at the adjusted tail-based estimator of ES that the paper developed (which we called $\hat{\mathrm{ES}}$). The AA estimator was then introduced. The objective is to compute the MSE, variance, and bias of the $\hat{\mathrm{ES}}$ and AA estimators to see which performs the best. The functions were then described and outlined.
In the source paper simulations, $\hat{\mathrm{ES}}$ massively outperformed AA in terms of MSE and variance and only fell short in bias. It was then concluded that $\hat{\mathrm{ES}}$ is a much better estimator than AA.
However, I could not replicate these conclusions in my Monte Carlo simulations. From my replications, AA outperformed $\hat{\mathrm{ES}}$ in terms of MSE and variance and only fell short in bias. This is exactly the opposite of the results from the paper. I am not sure exactly why the results between the source paper and my replications are so different. I thought that perhaps M = the number of Monte Carlo replications = 2500, was too low but trying with a much higher M did not yield results similar to those in

the source paper. In Section 7.5, I also showed that the main culprit was probably the heavier-tailed distributions where $\hat{\mathrm{ES}}$ performed particularly poorly. I believe that other approaches should be considered or improvements to the estimator should be made in terms of lowering the MSE/variance (especially with respect to heavier-tailed distributions) and to help resolve the consistent underestimation of the true ES.

9 References

[1] Norton, M., Khokhlov, V., & Uryasev, S. (2019). Calculating cvar and bpoe for common probability distributions with application to portfolio optimization and density estimation. *Annals of Operations Research*, *299*(1–2), 1281–1315. https://doi.org/10.1007/s10479-019-03373-1

[2] Pan, Z., Pang, T., & Zhao, Y. (2021). A simple and robust approach for expected shortfall estimation. *The Journal of Computational Finance*. https://doi.org/10.21314/jcf.2021.003

10 Appendix

The following is the R code I wrote for this report.

```r
if (!require("pacman")) install.packages("pacman")
pacman::p_load(cvar, TLMoments, ROOPSD)

A_ = function(alpha, y){
  N = length(y)
  A_alpha = (floor(N*alpha)+1-(N*alpha))*sort(y)[floor(N*alpha)] +
    (N*alpha - floor(N*alpha))*sort(y)[floor(N*alpha)+1]
  return(A_alpha)
}
s2_ = function(alpha, y, A){
  z = qnorm(alpha)
  x1 = (z^2 + 1 - (z/((1-alpha)*sqrt(2*pi)))*exp(-(z^2)/2))^-1
  x2 = sum((y-A)^2 * ifelse(y>A,1,0))/sum(ifelse(y>A,1,0))
  return(x1*x2)
}
mu_ = function(alpha, A, s2){
  z = qnorm(alpha)
```

```
  return(A - sqrt(s2)*z)
}
skew_ = function(y, A){
  x1 = sum((y-A)^3 * ifelse(y>A,1,0))/sum(ifelse(y>A,1,0))
  x2 = (sum((y-A)^2 * ifelse(y>A,1,0))/sum(ifelse(y>A,1,0)))^-1.5
  return(x1*x2)
}
R_ = function(skew, beta){
  if (beta==0.99){
    return(0.8611 + 0.5191*exp(-0.9747*skew) + 0.6099/skew -
             0.9413/(skew^2))
  }
  if (beta==0.995){
    return(0.9919 + 0.6681*exp(-0.9607*skew) + 0.6022/skew -
             1.4623/(skew^2))
  }
}
ES_adj_ = function(mu, s2, beta, A, R){
  z = qnorm(beta)
  ES_X = mu + sqrt(s2)/((1-beta)*sqrt(2*pi)) * exp(-(z^2)/2)
  ES_adj = (ES_X-A)*R + A
  return(ES_adj)
}
ES_W_ = function(beta, distr, df, shape, mu, s, scale, xi){
  if (distr == "t"){
    ES_W = ES(pt, p_loss = 1-beta, dist.type = "cdf", df=df)
  }
  if (distr == "gpd"){
    #xi != 0
    ES_W = mu + s*(((1-beta)^(-xi))/(1-xi) + ((1-beta)^(-xi) -
1)/xi)
  }
  if (distr == "weibull"){
    f = function(x) scale/(1-beta) * x^((1+1/shape)-1) * exp(-x)
    ES_W = (integrate(f,-log(1-beta),+Inf))$value
  }
  return(ES_W)
}
error_ = function(ES_W, ES_adj){
  return((ES_W - ES_adj)/(ES_W))
}
calculate_ES_adj = function(alpha, beta, y){
```

```
  A = A_(alpha,y)
  s2 = s2_(alpha, y, A)
  mu = mu_(alpha, A, s2)
  skew = skew_(y,A)
  R = R_(skew, beta)
  ES_adj = ES_adj_(mu, s2, beta, A, R)
  return(ES_adj)
}
calculate_error = function(alpha, beta, y, distr, ...){
  A = A_(alpha,y)
  s2 = s2_(alpha, y, A)
  mu = mu_(alpha, A, s2)
  skew = skew_(y,A)
  R = R_(skew, beta)
  ES_adj = ES_adj_(mu, s2, beta, A, R)
  ES_W = ES_W_(beta, distr, ...)
  error = error_(ES_W, ES_adj)
  return(error)
}

#Arithmetic Average
AA_ = function(beta, y){
  N = length(y)
  AA = sum(sort(y)[ceiling(N*beta):N])/(N+1-ceiling(N*beta))
  return(AA)
}
#Monte Carlo simulations
print_t = function(M){
  for (df in c(3.5,5,8)){
    for (n in c(250,500)){
      for (beta in c(0.99,0.995)){
        ES_W = ES_W_(beta=beta, distr="t", df=df)
        set.seed(001)
        print(mean(replicate(M, (calculate_ES_adj(alpha=0.95,
beta=beta,
                                                  y=rt(n=n, df=df)
- ES_W))^2)))
        set.seed(001)
        print(mean(replicate(M, (AA_(beta=beta,y=rt(n=n, df=df) -
                                       ES_W))^2)))
        set.seed(001)
        print(mean(var(replicate(M, calculate_ES_adj(alpha=0.95,
```

```
                                                          beta=beta,
y=rt(n=n, df=df))))))
        set.seed(001)
        print(mean(var(replicate(M, AA_(beta=beta, y=rt(n=n,
df=df))))))
        set.seed(001)
        print(mean(replicate(M, calculate_ES_adj(alpha=0.95,
beta=beta,
                                                      y=rt(n=n, df=df)
- ES_W))))
        set.seed(001)
        print(mean(replicate(M, AA_(beta=beta, y=rt(n=n, df=df) -
ES_W))))
      }
    }
  }
}
print_gpd = function(M){
  for (xi in c(0.3,0.2,0.1)){
    for (n in c(250,500)){
      for (beta in c(0.99,0.995)){
        ES_W = ES_W_(beta=beta, distr="gpd", mu=0, s=1, xi=xi)
        set.seed(001)
        print(mean(replicate(M, (calculate_ES_adj(alpha=0.95,
beta=beta,
                                                       y=rgpd(n=n,
loc=0, scale=1, shape=xi)) - ES_W)^2)))
        set.seed(001)
        print(mean(replicate(M, (AA_(beta=beta, y=rgpd(n=n, loc=0,
                                                            scale=1,
shape=xi)) - ES_W)^2)))
        set.seed(001)
        print(mean(var(replicate(M, calculate_ES_adj(alpha=0.95,
                                                          beta=beta,
y=rgpd(n=n, loc=0, scale=1, shape=xi))))))
        set.seed(001)
        print(mean(var(replicate(M, AA_(beta=beta, y=rgpd(n=n,
loc=0,
                                                               scale=1,
shape=xi))))))
        set.seed(001)
        print(mean(replicate(M, calculate_ES_adj(alpha=0.95,
```

```
beta=beta,
                                                        y=rgpd(n=n,
loc=0, scale=1, shape=xi)) - ES_W)))
        set.seed(001)
        print(mean(replicate(M, AA_(beta=beta, y=rgpd(n=n, loc=0,
scale=1,
                                                             shape=xi)) -
ES_W)))
      }
    }
  }
}
print_weibull = function(M){
  for (shape in c(0.6,0.9,1.4)){
    for (n in c(250,500)){
      for (beta in c(0.99,0.995)){
        ES_W = ES_W = ES_W_(beta=beta, distr="weibull",
shape=shape,
                            scale=1)
        set.seed(001)
        print(mean(replicate(M, (calculate_ES_adj(alpha=0.95,
beta=beta,
                                                         y=rweibull(n=n,
shape=shape, scale=1)) - ES_W)^2)))
        set.seed(001)
        print(mean(replicate(M, (AA_(beta=beta,y=rweibull(n=n,

shape=shape, scale=1)) - ES_W)^2)))
        set.seed(001)
        print(mean(var(replicate(M, calculate_ES_adj(alpha=0.95,
                                                            beta=beta,
y=rweibull(n=n, shape=shape, scale=1))))))
        set.seed(001)
        print(mean(var(replicate(M, AA_(beta=beta, y=rweibull(n=n,

shape=shape, scale=1))))))
        set.seed(001)
        print(mean(replicate(M, calculate_ES_adj(alpha=0.95,
beta=beta,
                                                        y=rweibull(n=n,
shape=shape, scale=1)) - ES_W)))
        set.seed(001)
```

```
        print(mean(replicate(M, AA_(beta=beta, y=rweibull(n=n,

shape=shape, scale=1)) - ES_W)))
      }
    }
  }
}
print_t2 = function(M){
  for (df in c(2.5,3)){
    for (n in c(250,500)){
      for (beta in c(0.99,0.995)){
        ES_W = ES_W_(beta=beta, distr="t", df=df)
        set.seed(001)
        print(mean(replicate(M, (calculate_ES_adj(alpha=0.95,
beta=beta,
                                                              y=rt(n=n, df=df)
- ES_W))^2)))
        set.seed(001)
        print(mean(replicate(M, (AA_(beta=beta,y=rt(n=n, df=df) -
                                               ES_W))^2)))
        set.seed(001)
        print(mean(var(replicate(M, calculate_ES_adj(alpha=0.95,
                                                                 beta=beta,
y=rt(n=n, df=df))))))
        set.seed(001)
        print(mean(var(replicate(M, AA_(beta=beta, y=rt(n=n,
df=df))))))
        set.seed(001)
        print(mean(replicate(M, calculate_ES_adj(alpha=0.95,
beta=beta,
                                                             y=rt(n=n, df=df)
- ES_W))))
        set.seed(001)
        print(mean(replicate(M, AA_(beta=beta, y=rt(n=n, df=df) -
ES_W))))
      }
    }
  }
}
print_gpd2 = function(M){
  for (xi in c(0.5,0.35)){
    for (n in c(250,500)){
```

```
      for (beta in c(0.99,0.995)){
        ES_W = ES_W_(beta=beta, distr="gpd", mu=0, s=1, xi=xi)
        set.seed(001)
        print(mean(replicate(M, (calculate_ES_adj(alpha=0.95,
beta=beta,
                                                  y=rgpd(n=n,
loc=0, scale=1, shape=xi)) - ES_W)^2)))
        set.seed(001)
        print(mean(replicate(M, (AA_(beta=beta, y=rgpd(n=n, loc=0,
                                                       scale=1,
shape=xi)) - ES_W)^2)))
        set.seed(001)
        print(mean(var(replicate(M, calculate_ES_adj(alpha=0.95,
                                                     beta=beta,
y=rgpd(n=n, loc=0, scale=1, shape=xi))))))
        set.seed(001)
        print(mean(var(replicate(M, AA_(beta=beta, y=rgpd(n=n,
loc=0,
                                                          scale=1,
shape=xi))))))
        set.seed(001)
        print(mean(replicate(M, calculate_ES_adj(alpha=0.95,
beta=beta,
                                                 y=rgpd(n=n,
loc=0, scale=1, shape=xi)) - ES_W)))
        set.seed(001)
        print(mean(replicate(M, AA_(beta=beta, y=rgpd(n=n, loc=0,
scale=1,
                                                      shape=xi)) -
ES_W)))
      }
    }
  }
}
print_t(2500)
print_gpd(2500)
print_weibull(2500)
print_t2(2500)
print_gpd2(2500)
```