Expected Shortfall Estimation and Gaussian Inference
for Infinite Variance Time Series

Jonathan B. Hill

Dept. of Economics, University of North Carolina

May 2013

ABSTRACT

We develop methods of non-parametric estimation for the Expected Shortfall of possibly heavy tailed asset returns that leads to asymptotically standard inference. We use a tail-trimming indicator to dampen extremes negligibly, ensuring standard Gaussian inference, and a higher rate of convergence than without trimming when the variance is infinite. Trimming, however, causes bias in small samples and possibly asymptotically when the variance is infinite, so we exploit a rarely used remedy to estimate and utilize the tail mean that is removed by trimming. Since estimating the tail mean involves estimation of tail parameters and therefore an added arbitrary choice of the number of included extreme values, we present weak limit theory for an ES estimator that optimally selects the number of tail observations by making our estimator arbitrarily close to the untrimmed estimator, yet still asymptotically normal. Finally, we apply the new estimators to financial returns data.

1 INTRODUCTION

Let \( \{ y_t : -\infty < t < \infty \} \) be a stationary stochastic process in \( \mathbb{R} \) where \( y_t \) has an absolutely continuous distribution, and let \( q_\alpha < 0 \) be the \( \alpha \)-quantile of \( y_t \) for some small \( \alpha > 0 \):

\[
P(y_t \leq q_\alpha) = \alpha \in (0, 1).
\]

This paper was previously circulated under the title "Robust Expected Shortfall Estimation for Infinite Variance Time Series".

Dept. of Economics, University of North Carolina-Chapel Hill, www.unc.edu/~jbhill, jbhill@email.unc.edu.

Key words and phrases: Expected Shortfall, heavy tails, robust estimation, bias correction.

JEL subject classifications. C13, C20, C22.

AMS subject classifications. Primary 62F35; secondary 62F07.

We thank two anonymous referees and Co-Editor Eric Renault for comments and suggestions that lead to an improved manuscript.
The $\alpha$-quantile Value-at-Risk [VaR] is $-q_\alpha > 0$, hence the probability a loss $-y_t$ exceeds $-q_\alpha$ is $\alpha$. The Expected Shortfall [ES] of $y_t$ is the expected loss given the loss exceeds the $\alpha$-level VaR:

$$ES_\alpha := E [ -y_t | y_t \geq -q_\alpha ] = -\frac{1}{\alpha} E [ y_t I (y_t \leq q_\alpha) ] > 0.$$ 

In order to justify the existence of the ES we assume integrability, thus $E|y_t| < \infty$.

ES is used in finance to provide valuable improvements over VaR as a coherent measure of risk (Artzner et al 1999, Acerbi and Tasche 2002, Scaillet 2004, Chen 2008, Danielsson 2011), and falls into a broad category of spectral measures of risk that includes the VaR (see Acerbi 2002 and Tasche 2002). See also McNeil et al (2005) and Christoffersen (2012) for textbook treatments. Recent studies focus on systemic or market risk, including expected shortfall for a portfolio, or the marginal impact of expected shortfall for a small change in the portfolio allocation. As another example, the systematic risk index of a particular institution is a function of the tail expectation of a firm’s loss conditional on a market event, which generalizes the ES self-conditional form $E[-y_t | y_t \geq -q_\alpha]$. A related measure is the VaR conditional on the state of an institution, and the spread between VaR’s conditional of normal and distress states of an institution. See Gourieroux et al (2000), Fermanian and Scaillet (2005), Brownlees and Engle (2010), Acharya et al (2010) and Adrian and Brunnermeier (2011) amongst many others.

Let $\{y_t\}_{t=1}^n$ be the observed sample of size $n \geq 1$. A simple nonparametric estimator of the ES is

$$\tilde{ES}_n = -\frac{1}{\alpha n} \sum_{t=1}^n y_t I (y_t \leq \tilde{q}_{n,\alpha}) ,$$

where $\tilde{q}_{n,\alpha}$ estimates $q_\alpha$. In the literature a finite second moment $E[y_t^2] < \infty$ is almost universally assumed in order to ensure a Gaussian limit for $\sqrt{n}(\tilde{ES}_n - ES_\alpha)$. See Scaillet (2004) and Chen (2008) and their references. See also Cai and Wang (2008) for use of nonparametric conditional distribution estimation for the ES, and for nonparametric estimation of the ES under netting agreements see Fermanian and Scaillet (2005).

Otherwise only Linton and Xiao (2012) tackle the heavy tail case $E[y_t^2] = \infty$ by considering those processes $y_t$ with a regularly varying distribution tail

$$P (|y_t| > c) = c^{-\kappa} L(c) \text{ with tail index } \kappa \in (0, 2),$$

and $L(c)$ is a slowly varying function.\footnote{Slow variation is defined by $\lim_{x \to \infty} L(\lambda x)/L(x) = 1 \forall \lambda > 0$. Classic examples are constants, and the natural log and its powers, e.g. $a(\ln(c))^b$ for any $a > 0$ and $b \geq 0$ (Resnick 1987).} Recall a tail index value $\kappa \leq 2$ implies $y_t$ has an infinite variance (see Leadbetter et al 1983 and Resnick 1987), hence Linton and Xiao (2012) only treat the infinite variance case. Linton and Xiao (2012) derive the non-Gaussian stable limit of $n^{1-1/\kappa} (\tilde{ES}_n - ES_\alpha)$ for geometrically $\alpha$-mixing $y_t$ with $\kappa \in (1, 2)$. Since the limit law is non-Gaussian and
depends on the unknown (but estimable) tail index $\kappa$, the authors present a subsampling method for inference. Notice they do not treat the hairline infinite variance case $\kappa = 2$ which neglects, for instance, IGARCH with Gaussian errors (e.g. Mikosch and Stårică 2000) and the popularly used RiskMetric volatility model which intrinsically has an IGARCH representation (cf. Christoffersen 2012: p. 23). Evidence for heavy tails in general and particularly conditional heteroscedasticity in financial and macroeconomic returns is substantial (e.g. Embrechts et al 1997, Rachev and Mittnik 2000, Davis 2010, Ibragimov et al 2010). Specifically, an infinite variance appears to exists for some exchange rates and asset returns, for example assets traded on emerging stock markets, depending on the time frame (Ling 2005, Ibragimov et al 2010), while evidence for IGARCH effects is even more substantial (see Morana 2002 and Caporale et al 2003 for examples).

In the case where tails are so heavy that $E|y_t| = \infty$ then the ES does not exist and it is no longer a valid solution to the lack of coherency of the VaR. See especially Garcia et al (2007) and Ibragimov (2009). Ibragimov et al (2009) and Ibragimov and Walden (2011) provide evidence of extremely heavy tailed economic losses following natural disasters. See also Nešlehová et al (2006), Powers (2010), Ibragimov and Walden (2011) and Mainik and Embrechts (2012) for infinite mean models and the implications for diversification.

Evidently there does not exist a unified estimation methodology for the ES that leads to classic inference even if $E[y_t^2] = \infty$. Since ES by construction conditions on a fixed upper tail quantile, in this paper we tail-trim $y_t$ from below using negative observations $y_t^{(-)} := y_tI(y_t < 0)$, their sample order statistics $y_{(1)}^{(-)} \leq y_{(2)}^{(-)} \cdots \leq y_{(n)}^{(-)} \leq 0$, and an intermediate order sequence $\{k_n\}$, where $k_n \to \infty$ and $k_n/n \to 0$, to determine the tail-trimming threshold. Our first tail-trimmed ES estimator is then

$$\overline{ES}_n^{(*)} = -\frac{1}{\alpha n} \sum_{t=1}^{n} y_t I \left( y_{(k_n)}^{(-)} \leq y_t \leq \hat{\varphi}_{n,\alpha} \right).$$

Thus $k_n$ is the number of trimmed left tail extremes, representing an asymptotically vanishing and therefore negligible sample tail portion $k_n/n \to 0$. Under a geometric $\alpha$-mixing condition on $\{y_t\}$ and few additional regularity conditions, negligibility ensures a consistent and asymptotically normal estimator $\overline{ES}_n^{(*)}$. In terms of inference we propose a HAC estimator for the variance of $\overline{ES}_n^{(*)}$ that allows for standard inference asymptotically.

A fixed quantile lower bound $\hat{\varphi}_{n,\beta} \leq y_t \leq \hat{\varphi}_{n,\alpha}$ for $0 < \beta < \alpha$ implies a non-negligible amount of trimming. This does not permit identification a la $\overline{ES}_n^{(*)} \overset{p}{\to} ES_\alpha$ since non-negligibility forces the probability limit of $|\overline{ES}_n^{(*)}|$ to be strictly less than $|ES_\alpha|$. Only negligible trimming $k_n/n \to 0$ ensures both asymptotic normality and identification of the true ES. Tail-trimming appears in the literature predominantly for mean estimation (cf. Csörgo et al 1986, Hahn et al 1990), including average treatment effects (e.g. Chaudhuri and Hill 2013), and has been used for heavy tail robust moment condition tests in Hill (2012a), Hill and Aguilar (2012) and for robust estimation of conditional mean and random volatility models in Hill (2012b, 2013) and Hill and Renault (2012).
We cover the general tail index case $\kappa > 1$ and show we can always achieve a rate of convergence higher than in the untrimmed case when variance is infinite $\kappa \in (1, 2)$. Further, we require fewer assumptions than Linton and Xiao (2012) since the limit theory here is guided by normal domain of attraction asymptotics for tail-trimmed mixing arrays derived from a general limit theory in Peligrad (1996). Moreover, our results apply to any tail index value $\kappa > 1$, covering the hairline infinite variance case $\kappa = 2$ and finite variance case $\kappa > 2$. Nevertheless, in order to simplify characterizing trimmed moments we impose a stronger tail decay property than the general form (1) used in Linton and Xiao (2012).

One-sided tail-trimming, however, necessarily introduces small sample and possibly asymptotic bias in the limit distribution when $E[y_t^2] = \infty$ since $\hat{E}S_n^{(*)}$ neglects a tail region $(-\infty, y_{(k_n)}^{-})$ that may vanish too slowly as $n \to \infty$ (Csörgö et al 1986, Peng 2001). We therefore exploit ideas in Peng (2001, 2004) that leads to new estimators that augment $\hat{E}S_n^{(*)}$ based on two sample versions $\hat{R}_n^{(1)}$ and $\hat{R}_n^{(2)}$ of the tail mean $E[y_t I(y_t < y_{(k_n)}^{-})]$ that is removed by tail-trimming. Peng (2001) estimates the tail mean with the same fractile $k_n$ in all components. Our first tail mean estimator $\hat{R}_n^{(1)}$ uses a flexible set of fractiles such that $\hat{E}S_n^{(1)} = \hat{E}S_n^{(*)} + \hat{R}_n^{(1)}$ resolves small sample bias and leads to the same easily estimated variance as $\hat{E}S_n^{(*)}$. Our remaining estimator $\hat{R}_n^{(2)}$ selects a fractile associated with $\hat{R}_n^{(1)}$ that forces $\hat{E}S_n^{(2)} = \hat{E}S_n^{(*)} + \hat{R}_n^{(2)}$ to be arbitrarily close to the untrimmed $\hat{E}S_n$ as $n \to \infty$ without undermining asymptotic normality, and again having the same variance as $\hat{E}S_n^{(*)}$. The estimator $\hat{E}S_n^{(2)}$ has exceptional sampling properties with essentially the same empirical bias as the untrimmed estimator $\hat{E}S_n$, and contrary to the latter our estimator is normally distributed asymptotically, and approximately normal in small samples.4

Peng (2004) solves a similar problem also for iid data by exploiting Qin and Wong’s (1996) semi-parametric Empirical Likelihood Method [ELM] to generate a chi-squared based asymptotic confidence region for the trimmed mean with tail mean estimators. The semi-parametric ELM should offer yet another means to gain standard inference for the ES since it side-steps estimation of the standard error (cf. Peng 2004). Under dependence, however, block smoothing is necessary in order to promote a classic EL limit theory (Owen 1991, Kitamura 1997), hence inference still requires use of a bandwidth. Thus, whether ELM has any advantage over our method is unknown and left for future research.

The rate of convergence of our ES estimators are all identical and a function of the unknown tail index $\kappa$, although the rate is $n^{1/2}$ when $\kappa > 2$. Nevertheless, once standardized our estimator is asymptotically standard normal, and if $\kappa \in (1, 2]$ then the rate can always be assured to be $n^{1/2}/L(n)$ for some slowly varying $L(n) \to \infty$. In a general context Ibragimov and Müller (2010) treat a scalar parametric estimator with an asymptotically normal distribution and provide a t-test and confidence band method robust to unknown correlation, heterogeneity and therefore rate of

---

4The use of an optimized bias corrected tail-trimmed variance estimator was developed simultaneously in Hill and Renault (2012) for Gaussian inference in heavy tailed GARCH models based on variance targeting.
convergence properties. Although their goal is improved inference when the dependence structure is unknown, it is applicable here and indeed in any context of scalar inference with Gaussian asymptotics under tail-trimming in view of the unknown rate of convergence (e.g. Hill 2012a,b, 2013, Hill and Aguilar 2012).

Finally, our estimation method carries over to a large variety of related financial risk measures. Examples include those discussed above like the marginal excess shortfall (Brownlees and Engle 2010), other members of the spectral class of risk measures (Acerbi 2002), and expected shortfall under netting agreements (Fermanian and Scaillet 2005) to name a few. Consider, for example, the spectral class presented in Acerbi (2002). Define the Dirac Delta function

\[ \delta(x - c) = f(c) \forall c \in [a, b] \]

for all continuous differentiable functions \( f \) with compact support, and let \( \delta'' \) satisfy \( \int_a^b f(x)\delta''(x - c)dx = -f'(c) \forall c \in [a, b] \). It is easily seen than the ES is identically

\[ ES_\alpha = \int_0^1 \xi \frac{d\mu_\alpha}{d\xi} \]

where \( d\mu_\alpha = \alpha^{-1} \delta(\alpha - \xi)d\xi \), and the VaR is identically \( \int_0^1 \xi \frac{d\mu_\alpha}{d\xi} \) with \( d\mu_\alpha \) = \( \delta'(\alpha - \xi)d\xi \). This suggests a broad spectrum of related risk measures \( M_\mu := \int_0^1 \xi \frac{d\mu_\alpha}{d\xi} \) where \( d\mu_\alpha \) is any measure on \([0, 1]\) that satisfies \( \int_0^1 \xi d\mu_\alpha = 1 \). Our estimators of \( ES_\alpha \), which allow for heavy tails and standard asymptotic inference, instantly apply to this general class of risk measures.

The estimators \( \hat{ES}_n^{(\alpha)} \) are developed in Sections 2 and 3. In Section 3 we present a consistent estimator of the scale used for inference. Sections 4 and 5 contain a simulation study and an empirical application. The appendices contain proofs, and all tables and figures are placed at the end.

We use the following notation. \( \iota > 0 \) is a tiny constant and \( K > 0 \) is a constant, the values of which may change from line to line. \( L(n) \to \infty \) is a slowly varying function, the value and rate of which may change from line to line. \( a_n \sim b_n \) implies \( a_n/b_n \to 1 \). \([z]\) is the integer part of variable \( z \). \( a_n \sim b_n \) implies \( a_n/b_n \to 1 \) as \( n \to \infty \).

2 TAIL-TRIMMED ESTIMATOR

All random variables in this paper exist on the same probability measure space \((\Omega, \mathcal{F}, P)\), and all \( \sigma \)-fields are subsets of \( \mathcal{F} \). Define \( \mathcal{Y}_t := \sigma(y_t : \tau \leq t) \) and \( \mathcal{Y}_s^t := \sigma(y_s : s \leq \tau \leq t) \), and define sample order statistics \( y(i) \) of \( y_t \): \( y(1) \geq y(2) \geq \cdots \geq y(n) \). Similarly, define the left-tail process \( \{y_t^{(-)}\} \) and its order statistics:

\[ y_t^{(-)} := y_t I(y_t < 0) \quad \text{and} \quad y(1)^{(\cdot)} \leq y(2)^{(\cdot)} \leq \cdots \leq y(n)^{(\cdot)} \leq 0. \]

In the ES construction we estimate the VaR \(-q_\alpha\) with a central order statistic for brevity (e.g. Chen 2008),

\[ \hat{q}_{n,\alpha} = y([\alpha n]). \]
although a smoothed estimator can be employed as in Scaillet (2004) and Linton and Xiao (2012).

Let \( \{k_n\} \) be an intermediate order sequence, hence \( k_n \to \infty \) and \( k_n = o(n) \), and let \( \{l_n\} \) be the sequence of positive numbers that satisfies (e.g. Leadbetter et al 1983)

\[
P(y_t < -l_n) = \frac{k_n}{n}.
\]

Distribution continuity ensures such thresholds \( \{l_n\} \) exist for any choice of \( \{k_n\} \), and \(-l_n\) is the left-tail \( k_n/n \) quantile of \( y_t \). By construction \( y_{(k_n)} \) estimates \(-l_n\). The associated ES estimator is an intermediate order left-tail and central order right-tail trimmed mean

\[
\widetilde{ES}_n^{(*)} = -\frac{1}{n} \alpha \sum_{t=1}^n y_t I\left(y_{(k_n)} \leq y_t \leq y_{([\alpha n])}\right).
\]

Asymptotic theory requires a version of the trimmed variable \( y_t I(y_{(k_n)} \leq y_t \leq y_{([\alpha n])}) \) based on the non-random thresholds \( \{l_n\} \), and its long-run variance, so define

\[
y_{n,t} := y_t I(-l_n \leq y_t \leq q_\alpha) \quad \text{and} \quad S_n^2 := \frac{1}{n} E\left(\sum_{t=1}^n \left(y_{n,t}^* - E[y_{n,t}^*]\right)^2\right).
\]

Although \( \widetilde{ES}_n^{(*)} \), \( y_{n,t}^* \) and \( S_n^2 \) implicitly depend on the level of risk \( \alpha \), we suppress such dependence for notational economy.

In order to obtain Gaussian asymptotics we impose the following mixing, identification and tail decay conditions.

**ASSUMPTION D (dgp).** \( y_t \) is governed by a strictly stationary, non-degenerate and absolutely continuous distribution on \((\infty, \infty)\), and is geometrically \( \alpha \)-mixing: \( \sup_{A \in \mathcal{A}^t, B \in \mathcal{A}^\infty} |P(A \cap B) - P(A)P(B)| = O(\rho^h) \) where \( \rho \in (0, 1) \).

**ASSUMPTION I (identification).** \( n^{1/2}S_n^{-1}E[y_t I(y_t < -l_n)] \to 0. \)

We show in the appendix that \( \widetilde{ES}_n^{(*)} \) obtains the expansion

\[
\frac{n^{1/2}}{S_n/\alpha} \left(\widetilde{ES}_n^{(*)} - ES_\alpha\right) = -\frac{1}{n^{1/2}S_n} \sum_{t=1}^n \left(y_{n,t}^* - E[y_{n,t}^*]ight) - \frac{n^{1/2}}{S_n} \left(E[y_{n,t}^*] + \alpha \times ES_\alpha\right)
\]

\[
= -\frac{1}{n^{1/2}S_n} \sum_{t=1}^n \left(y_{n,t}^* - E[y_{n,t}^*]\right) - \frac{n^{1/2}}{S_n} \left(E[y_t I(y_t \leq q_\alpha)] - E[y_t I(-l_n \leq y_t \leq q_\alpha)]\right)
\]

\[
= \mathfrak{R}_n - \frac{n^{1/2}}{S_n} \left(y_t I(y_t < -l_n)\right) = \mathfrak{R}_n + \mathfrak{R}_n
\]
say, where the equality holds asymptotically with probability approaching one. The first term \( \mathcal{Z}_n \) satisfies a Gaussian central limit theorem, thus Assumption I ensures asymptotic unbiasedness by \( \mathcal{R}_n \to 0 \). By arguments in Section 3 we have \( n^{1/2}|E[y_t I(y_t < -l_n)]| \sim Kn^{1/2}(k_n/n)|l_n = Kn^{1/2}(n/k_n)^{1/\kappa - 1} = Kk_n^{1-1/\kappa}/n^{1/2-1/\kappa} \), and if the variance is finite \( \kappa > 2 \) then \( 0 < \lim_{n \to \infty} \mathcal{S}_n^2 < \infty \) under geometric \( \alpha \)-mixing (Ibragimov 1962). In this case Assumption I implies we cannot trim too much, in particular \( k_n \) must not increase too fast

\[
k_{n}/n^{(\kappa/2-1)/(\kappa-1)} \to 0.
\]

This is assured by letting \( k_n \) be slowly varying, for example \( k_n = [\ln(n)] \).

Assumption I otherwise may rule out heavy tails \( \kappa \in (1, 2) \) since \( |\mathcal{R}_n| \to \infty \) is possible. This follows by noting if \( \kappa \in (1, 2) \) then \( \mathcal{S}_n^2 \sim Kr_n(n/k_n)^{2/\kappa - 1} \) by Lemma 2.1, below, for some sequence of non-random positive numbers \( \{r_n\} \) where \( r_n = O(\ln(n)) \), and \( r_n = O(1) \) if \( y_t \) is \( m \)-dependent.

Now use \( |E[y_t I(y_t < -l_n)]| \sim K(n/k_n)^{1/\kappa - 1} \) when \( \kappa \in (1, 2) \) to deduce asymptotic bias \( |\mathcal{R}_n| = n^{1/2}\mathcal{S}_n^{-1}|E[y_t I(y_t < -l_n)]| \sim Kk_n^{1/2}/r_n \to (0, \infty) \) when \( y_t \) is heavy tailed and \( m \)-dependent, or \( y_t \) is heavy tailed and \( k_n \to \infty \) at least as fast as \( (\ln(n))^2 \). See Section 2.2 for an ES estimator based on \( \hat{ES}_n^{(k)} \) with an asymptotic bias correction.

**ASSUMPTION T (tails).** \( P(y_t < -c) = dc^{-\kappa}(1 + o(1)) \) as \( c \to \infty \) where \( d > 0 \) and \( \kappa > 1 \).

**Remark:** Assumption T aligns with the stable domain of attraction for a sample mean of the unbounded left tail variable \( y_t I(y_t \leq q_\alpha) \). See Leadbetter et al (1983) and Resnick (1987).

Recall the tail index \( \kappa \) in Assumption T is identically the left-tail moment supremum:

\[
\kappa = \arg \sup \left\{ \alpha > 0 : E \left| y_t I(y_t < 0) \right|^\alpha < \infty \right\} \text{ hence } E \left| y_t I(y_t < 0) \right|^p < \infty \text{ iff } p \in (0, \kappa).
\]

The assumption is in some sense more restrictive than the general regularly varying form (1) used in Linton and Xiao (2012). We only restrict the left tail since \( y_t I(y_t \leq q_\alpha) \) is bounded from above, but the restriction is of the Pareto class \( dc^{-\kappa}(1 + o(1)) \). We can allow a general left-tail \( P(y_t < -c) = c^{-\kappa}L(c) \), but the Pareto class \( L(c) = d(1 + o(1)) \) greatly simplifies deducing the long-run variance rate \( \mathcal{S}_n^2 \to \infty \) in the infinite variance case in view of Karamata’s Theorem. See Lemma 2.1, below.

As a regularity condition we require that \( \mathcal{S}_n^2 \) be non-degenerate as \( n \to \infty \).

**ASSUMPTION N (non-degeneracy).** \( \liminf_{n \to \infty} \{\mathcal{S}_n^2/E[y_{n,t}^2]\} > 0 \).

**Remark 1:** A non-degenerate distribution under Assumption D and the negligibility of trimming ensure \( \liminf_{n \to \infty} E[y_{n,t}^2] > 0 \). Assumption N, however, rules out perverse cases for the long-run variance \( \mathcal{S}_n^2 \) due to negative correlation wherein \( \mathcal{S}_n^2 \to 0 \). In order to understand the nu-
ance, define tail-trimmed covariances \( \gamma^*_n(h) := E[y^*_n y^*_{n,t-h}] - (E[y^*_n])^2 \) and correlations \( \rho^*_n(h) := \gamma^*_n(h)/\gamma^*_n(0) \), and note \( S^2_n = \gamma^*_n(0) \times (1 + 2 \sum_{h=1}^{n-1} (1 - h/n) \rho^*_n(h)) \geq 0 \). Now, by dominated convergence \( E[y^*_n] \to E[y] \) hence \( \gamma^*_n(0)/E[y^*_n] \sim 1 - (E[y])^2/E[y]_n^2 \). In the infinite variance case \( \gamma^*_n(0)/E[y^*_n^2] \sim 1 \) and otherwise by dominated convergence \( \gamma^*_n(0)/E[y^*_n^2] \sim 1 - (E[y])^2/E[y]^2 \), hence in general \( \gamma^*_n(0)/E[y^*_n^2] \sim K > 0 \). Assumption N therefore implies \( \lim \inf_{n \to \infty} \{S^2_n/E[y^*_n^2]\} = K(1 + 2 \lim \inf_{n \to \infty} \{\sum_{h=1}^{n-1} \rho^*_n(h)\}) > 0 \). The only way for \( \lim \inf_{n \to \infty} \{S^2_n/E[y^*_n^2]\} = 0 \) to hold is for \( \rho^*_n(h) < 0 \) to be large for some lag(s) \( h \geq 1 \). Conditions like Assumption N are standard in the central limit theory literature for dependent data (e.g. Dehling et al 1986) and used here to ensure a central limit theorem for tail-trimmed arrays.

Remark 2: If \( y_t \) is \( m \)-dependent then Assumption N automatically holds.

**Lemma 2.1.** Under Assumptions D and T:

a. \( S^2_n = r_n E[(y^*_n - E[y^*_n])^2] = o(n) \) where \( \{r_n\} \) is a sequence of positive numbers that does not depend on \( k_n \), \( \lim \inf_{n \to \infty} r_n > 0 \) and \( r_n = O(\ln(n)) \), and \( r_n = O(1) \) if \( y_t \) is \( m \)-dependent or \( E[y^2] < \infty \).

b. If \( \kappa = 2 \) then \( E[y^*_n^2] = K \ln(n) \), and if \( \kappa \in (1, 2) \) then \( E[y^*_n^2] \sim \kappa(2 - \kappa)^{-1} d^{2/\kappa}(n/k_n)^{2/\kappa-1} \).

Remark: The term \( r_n = O(\ln(n)) \) arises due to non-\( m \)-dependence and heavy tails by exploiting a covariance bound in Rio (1993).

The tail-trimmed ES estimator is consistent and asymptotically normal.

**Theorem 2.2.** Under Assumptions D, N and T \( \hat{ES}^n^{\alpha} \Rightarrow ES_\alpha \) and \( n^{1/2}S_n^{-1}(\hat{ES}^n^{\alpha} - ES_\alpha + \alpha^{-1}E[yI(y < -l_n)]) \overset{d}{\to} N(0, \alpha^{-2}) \). If additionally Assumption I holds then \( n^{1/2}S_n^{-1} \{\hat{ES}_n^{\alpha} - ES_\alpha\} \overset{d}{\to} N(0, \alpha^{-2}) \). If \( \kappa \leq 2 \) then the rate of convergence is \( n^{1/2}S_n^{-1} = o(n^{1/2}) \).

As discussed above, if \( y_t \) has a finite variance and \( k_n \to \infty \) at a slowly varying rate then identification Assumption I holds. In this case and in view of geometric \( a \)-mixing it follows \( \lim_{n \to \infty} S^2_n \) exists and is positive (cf. Ibragimov 1962).

**Corollary 2.3.** Let \( E[y^2] < \infty \) and \( k_n \to \infty \) at a slowly varying rate (e.g. \( k_n \sim \ln(n) \)). Under Assumptions D, N and T \( n^{1/2}(\hat{ES}^n^{\alpha} - ES_\alpha) \Rightarrow N(0, \alpha^{-2}S^2) \) where \( S^2 := \lim_{n \to \infty} S^2_n \) satisfies \( 0 < S^2 < \infty \).

\(^5\) Notice having some or many large \( \rho^*_n(h) \) is fundamentally different from the long memory property that correlations are not absolutely summable. In the finite variance case long memory is ruled out by geometric \( a \)-mixing since by dominated convergence and Lemma 1.3 in Ibragimov (1962) it follows \( \lim_{n \to \infty} \sum_{h=1}^{n-1} \rho^*_n(h) = \sum_{h=1}^{\infty} \rho(h) < \infty \) where \( \rho(h) \) is the untrimmed correlation. In the infinite variance case \( \sum_{h=1}^{n-1} \rho^*_n(h) \to \infty \) is possible under non-\( m \)-dependence and geometric \( a \)-mixing since \( |\gamma^*_n(h)| \to \infty \) and therefore \( |\rho^*_n(h)| \to 1 \) are possible for infinitely many \( h \in \mathbb{N} \).
The iid case is instructive since if $\kappa \in (1,2)$ then by Lemma 2.1 $S_n^2 \sim E[y_n^2] \sim \kappa(2 - \kappa)^{-1}d^2/\kappa(n/k_n)^{2/\kappa-1} \to \infty$. In this case the asymptotic variance has a simple form that depends on tail parameters.

**COROLLARY 2.4.** Let $y_t$ be iid. Under Assumptions N and T with index $\kappa \in (1,2)$

$$\frac{n^{1/2}}{(n/k_n)^{1/\kappa-1/2}} (ES_n^{(s)} - ES_n + E[y_t I(y_t < -l_n)]) \xrightarrow{d} N \left( 0, \alpha^{-2}d^{-2/\kappa}\left(\frac{2 - \kappa}{\kappa}\right) \right).$$

In general for geometrically $\alpha$-mixing data and in the case of heavy tails, Lemma 2.1 shows $S_n^2 \to \infty$ hence the rate of convergence is $n^{1/2}/S_n = o(n^{1/2})$, similar to the case of untrimmed ES estimation (Linton and Xiao 2012). However, by construction trimming removes the most damaging extremes which can be exploited to elevate the rate. By Lemma 2.1 if $\kappa \in (1,2)$ then $S_n^2 \sim Kr_n(n/k_n)^{2/\kappa-1}$, hence the rate is

$$\frac{n^{1/2}}{S_n} = \frac{n^{1/2}}{r_n^{1/2} \times (n/k_n)^{1/\kappa-1/2}} = n^{1-1/\kappa} \times \frac{k_n^{1/\kappa-1/2}}{r_n^{1/2}} \text{ where } 0 < r_n \leq K \ln(n).$$

A rapid rate of trimming $k_n \to \infty$ therefore optimizes the rate of convergence since the presence of extremes diminishes estimator sharpness, a well known result for a sample tail-trimmed mean (e.g. Csörgo et al 1986, Hill 2012a, Hill and Aguilar 2012). Linton and Xiao (2012: Theorem 2) show the rate for the untrimmed estimator $\hat{ES}_n$ is exactly $n^{1-1/\kappa}$ when $\kappa \in (1,2)$. Thus $\hat{ES}_n^{(s)}$ converges faster than $\hat{ES}_n$ sufficiently if the rate of trimming $k_n \to \infty$ is regularly varying since $r_n \leq K \ln(n)$, for example $k_n \sim \lambda n^\delta$ or $k_n \sim \lambda n/\ln(n)$ for $\lambda \in (0,1]$ and $\delta \in (0,1)$. Notice a regularly varying rate does not necessarily equate to trimming many observations: if $n \in \{100,500,1000\}$ then $k_n \sim .25n^{2/3}$, for example, implies only $\{5.4\%, 3.1\%, 2.5\%\}$ of observations are trimmed. See our simulation study, below. Of course, the fastest rate of convergence $n^{1/2}$ is achieved when $k_n = \lambda n$, a central order sequence, since then by the mixing property $S_n^2 = O(1)$. This implies a fixed portion of the sample is trimmed, a case not considered here since only non-negligible left-tailed trimming induces bias both in small samples and asymptotically.

As an example of a trimming policy that optimizes the convergence rate when $\kappa \in (1,2)$, consider $k_n = [n/g_n]$ for a sequence of positive numbers $\{g_n\}$ where $g_n \to \infty$ as slowly as we choose and not faster than a slowly varying function, e.g. $g_n = K \ln(n)$ or $g_n = K \ln(\ln(n))$. Then $k_n$ is close to the central order $\lambda n$ for $\lambda \in (0,1)$, and if $\kappa \in (1,2)$ then in view of Lemma 2.1 the variance $S_n^2 = r_n^2g_n^{2/\kappa-1} \to \infty$ no faster than a slowly varying rate, and $\hat{ES}_n^{(s)}$ achieves a rate of convergence $n^{1/2} / (r_n^{1/2}g_n^{1/\kappa-1/2})$ which is close to $n^{1/2}$. In the hairline infinite variance case $\kappa = 2$ we have by Lemma 2.1 $S_n^2 = K r_n \ln(n) = O((\ln(n))^2)$ for any intermediate order sequence $\{k_n\}$, hence the rate of convergence is $n^{1/2} / (r_n \ln(n))^{1/2}$ which is again close to $n^{1/2}$.

**COROLLARY 2.5.** Let Assumptions D, N and T hold. If $\kappa \in (1,2)$ and $k_n = [n/g_n]$ for
$g_n \to \infty$ not faster than a slowly varying function, or $\kappa = 2$ and $\{k_n\}$ is any intermediate order sequence, then $S_n^2 \to \infty$ not faster than a slowly varying function. Further, for some $0 < v^2 < \infty$ that may be different in different places,

$$
\kappa \in (1, 2): \frac{n^{1/2}}{r_n^{1/2}g_n^{1/\kappa-1/2}}(\overline{ES}_n^{(s)} - ES_\alpha + \frac{1}{\alpha}E[y_t I(y_t < -l_n)]) \xrightarrow{d} N(0, v^2)
$$

$$
\kappa = 2: \left(\frac{n}{r_n \ln(n)}\right)^{1/2}(\overline{ES}_n^{(s)} - ES_\alpha + \frac{1}{\alpha}E[y_t I(y_t < -l_n)]) \xrightarrow{d} N(0, v^2).
$$

**Remark:** The asymptotic variance $v^2$ is not unique since it depends on the level of risk $\alpha$, the tail index $\kappa$, and the chosen fractile $k_n$. This is irrelevant for inference since we estimate $S_n^2$ non-parametrically. See Section 3.3. See also Ibragimov and Müller (2010) for a method of robust inference when the rate of convergence or asymptotic variance are not known.

### 3 BIAS CORRECTED ES ESTIMATORS

By construction $ES_\alpha = -\alpha^{-1}E[y_t I(-l_n \leq y_t \leq q_\alpha)] + \alpha^{-1}E[y_t I(-l_n < y_t)]$. As discussed above, the statistic $\overline{ES}_n^{(s)}$ is possibly biased asymptotically if $\kappa \in (1, 2)$ since the tail mean $-E[y_t I(y_t < -l_n)]$ is removed and may vanish too slowly asymptotically. Define the quantile function

$$
Q(u) := \inf \{x \geq 0: P(y_t \leq x) \geq u\} \text{ where } 0 \leq u \leq 1,
$$

and note under Assumption T if $\kappa \in (1, 2)$ then $Q(u) = d^{1/\kappa}u^{-1/\kappa}$ as $u \to 0$. In order to construct an estimator of the tail mean that is asymptotically normal, notice

$$
E[y_t I(y_t < -l_n)] \sim \int_0^{k_n/n} Q(u)du \sim -d^{1/\kappa} \left(\frac{\kappa}{\kappa-1}\right) \left(\frac{k_n}{n}\right)^{-1/\kappa+1} = -\frac{\kappa}{\kappa-1} \frac{k_n}{n} l_n. \quad (4)
$$

Peng (2001, 2004) exploits a similar result to improve on a left- and right-tailed trimmed mean for iid data. We exploit the same trick for improved ES estimators for mixing data under left-tail trimming. We then present consistent estimators of the scales for standard inference. Finally, we show that our bias estimator is negligible asymptotically even if tails are so thin that they decay infinitely faster than the Assumption T power law rate (e.g. $y_t$ is Gaussian, or has bounded support).
3.1 Tail Mean Estimator

By (4) we may estimate the tail mean by estimating \(-\kappa(\kappa - 1)^{-1}(k_n/n)l_n\). We already have \(y_{(k_n)}\) as an estimator of \(-l_n\), and in order to estimate \(\kappa\) we consider Hill’s (1975) seminal estimator:

\[
\hat{\kappa}_{k_n}^{(-)} := \left( \frac{1}{k_n} \sum_{i=1}^{k_n} \ln \left( \frac{y_{(i)}}{y_{(k_n)}} \right) \right)^{-1}.
\]

Hill (2010, 2011a) proves \((-y_{(k_n)})/l_n^{(-)} - 1\) and \(\hat{\kappa}_{k_n}^{(-)} - \kappa\) are \(O_p(k_n^{-1/2})\), and asymptotically normal for a large class of dependent and heterogeneous time series that covers Assumptions D and T. In view of formula (4) the enhanced ES estimator based on Peng’s (2001) idea is therefore

\[
\tilde{ES}_n^{(pg)} := \tilde{ES}_n^{(*)} - \frac{1}{\alpha} \left( \frac{\hat{\kappa}_{k_n}^{(-)}}{\hat{\kappa}_{m_n}^{(-)}} - 1 \right) \frac{k_n}{n} y_{(k_n)} = \tilde{ES}_n^{(*)} + \hat{R}_n,
\]

say.

Peng (2001) uses the same fractile sequence \(\{k_n\}\) for estimating all relevant components in \(\tilde{ES}_n^{(*)}\) and \(\hat{R}_n\). In terms of theory, we show in the appendices that only \(\tilde{ES}_n^{(*)}\) and \((k_n/n)y_{(k_n)}\) need to be based on the same \(k_n\) to ensure asymptotic unbiasedness. If we also use \(k_n\) for estimating \(\kappa\) then the limit theory is complicated by the fact that each \(\tilde{ES}_n^{(*)}\), \(\hat{\kappa}_{k_n}^{(-)}\), and \(y_{(k_n)}\) impact \(\tilde{ES}_n^{(pg)}\) asymptotically: even in the iid case inference requires a complicated parametric asymptotic variance estimator (see Peng 2001: Theorem 1).

As a practical matter trimming very few observations for each \(n\) sufficiently erodes the heavy tailedness of the remaining sample for Gaussian asymptotics. Conversely, using many tail observations for an order statistic based estimator of \(\kappa\) improves its rate of convergence, and allows it to not impact a bias-corrected ES estimator asymptotically as we show below.\(^7\) We therefore use \(\hat{\kappa}_{m_n}^{(-)}\) with a larger fractile \(m_n > k_n\) as \(n \to \infty\), specifically

\[
\frac{k_n}{m_n} \to 0 \text{ where } m_n \to \infty \text{ and } m_n = o(n),
\]

hence our first bias-corrected ES estimator is

\[
\tilde{ES}_n^{(1)} := \tilde{ES}_n^{(*)} - \frac{1}{\alpha} \left( \frac{\hat{\kappa}_{m_n}^{(-)}}{\hat{\kappa}_{m_n}^{(-)}} - 1 \right) \frac{k_n}{n} y_{(k_n)} = \tilde{ES}_n^{(*)} + \hat{R}_n^{(1)}.
\]

\(^6\)Many estimators of the tail index are available. See, e.g., de Haan and Peng (1998) and Hill (2010), amongst others, for references and theory.

\(^7\)An estimator of \(\kappa\) for GARCH processes is treated in Berkes et al (2003). This estimator is \(n^{1/2}\)-convergent since all observations are used for estimation, but the GARCH error must have a finite fourth moment which limits it use for heavy tailed data.
If \( y_t \) is known to have the same left- and right-tail indices, or the left tail is known to be heavier\(^8\), then we use a two tailed tail index estimator since this can include more data points and therefore be more efficient:

\[
\hat{\kappa}_{m_n} := \left( \frac{1}{m_n} \sum_{i=1}^{m_n} \ln \left( \frac{y_{(i)}/y_{(m_n)}}{y_{(a)}/y_{(m_n)}} \right) \right)^{-1} \quad \text{where } y_{(a)} := |y_t|.
\]

Property (5) ensures \( \hat{\kappa}_{m_n} \) \( \mathcal{P} \rightarrow \kappa \) fast enough to eradicate any impact on \( \hat{ES}_n^{(1)} \). We specifically require \( \hat{\kappa}_{m_n} = \kappa + O_p(1/m_{n}^{1/2}) \) which in general holds as long as \( y_t \) satisfies a second order power law condition and observations are taken from sufficiently far out in the tails, hence \( m_n \rightarrow \infty \) cannot be too fast (Hill 2010: Theorem 2, cf. Hall 1982: Theorem 2, Haeusler and Teugels 1985: Section 5). We invoke the following second order extension of Paretian decay Assumption T, although other second order properties are possible. See Hall (1982) and Haeusler and Teugels (1985).

**ASSUMPTION T’ (tails).** Let \( P(y_t < -c) = dc^{-\kappa}(1 + O(c^{-\beta})) \) as \( c \rightarrow \infty \) where \( d, \beta > 0 \) and \( \kappa > 1 \). Further

\[
m_n \rightarrow \infty \quad \text{and} \quad m_n = o \left( n^{2\beta/(2\beta+\kappa)} \right).
\]

**Remark 1:** In the exact Pareto case \( P(y_t < -c) = dc^{-\kappa} \) for \( c > 0 \) the rate bound in (6) becomes simply \( m_n = o(n) \), but as the tail deviates from a true Pareto then observations must be taken from farther out in the tails. See Haeusler and Teugels (1985).

**Remark 2:** An example of sequences \( \{k_n, m_n\} \) that satisfy (5) and (6) is

\[
k_n = \left[ \frac{n^{2\beta/(2\beta+\kappa)}}{(\ln(n))^{2\varepsilon}} \right] \quad \text{and} \quad m_n = \left[ \frac{n^{2\beta/(2\beta+\kappa)}}{(\ln(n))^{4\varepsilon}} \right] \quad \text{for tiny } \varepsilon > 0.
\]

If \( \beta \geq \kappa \) then we may use \( k_n \sim \lambda_k n^{2/3}/(\ln(n))^{2\varepsilon} \) and \( m_n \sim \lambda_m n^{2/3}/(\ln(n))^{4\varepsilon} \) for scale fractions \( \lambda_k, \lambda_m > 0 \). Another example is

\[
k_n = \left[ \lambda_k (\ln(n))^{1-\varepsilon} \right] \quad \text{and} \quad m_n = \left[ \lambda_m \ln(n) \right] \quad \text{for } \lambda_k, \lambda_m > 0 \text{ and tiny } \varepsilon > 0.
\]

Recall \( y^*_{n,t} := y_t I(-l_n \leq y_t \leq q_n) \). Stack non-tail and tail variables

\[
W_{n,t} := \left[ y^*_{n,t} - E[y^*_{n,t}], \left( \frac{n}{k_n} \right)^{1/2} \{ I(y_t < -l_n) - P(y_t < -l_n) \} \right]'
\]

---

\(^8\)Suppose, for example, \( P(y_t < -y) \sim d_1 y^{-\kappa_1} \) and \( P(y_t > y) \sim d_2 y^{-\kappa_2} \) where \( d_1, d_2 > 0 \) and \( \kappa_1 \leq \kappa_2 \). Then \( P(|y_t| > y) \sim d_1 y^{-\kappa_1} + d_2 y^{-\kappa_2} = d_1 y^{-\kappa_1}(1 + (d_2/d_1) y^{\kappa_1 - \kappa_2}) = d_1 y^{-\kappa_1}(1 + o(1)) \): two-tailed and left-tailed indices are identical.
and define

\[ \mathcal{V}_n^2 := B_n' \Sigma_n B_n \]  

where

\[ B_n := \left[ 1, -\frac{1}{\kappa - 1} \left( \frac{k_n}{n} \right)^{1/2} l_n \right]' \]

and

\[ \Sigma_n := \frac{1}{n} \sum_{s,t=1}^n E \left[ \mathcal{W}_{n,s} \mathcal{W}_{n,t}' \right] \in \mathbb{R}^{2 \times 2}. \]

Notice the upper left element of \( \Sigma_n \) is identically \( S_n^2 \).

We must strengthen non-degeneracy Assumption N to cover \( \Sigma_n \). Let \( \lambda \in \mathbb{R}^2 \).

**ASSUMPTION N’ (non-degeneracy).** \( \lim \inf_{n \to \infty} \inf_{\lambda' \lambda = 1} ||\lambda' \Sigma_n \lambda / (E(\lambda' \mathcal{W}_{n,t}))|| > 0. \)

Identification Assumption I is now superfluous.

**THEOREM 3.1.** Under Assumptions D, N’, T’, and (5)-(6) we have \( n^{1/2} \mathcal{W}_n^{-1}(\hat{E}_{n}^{(1)} - ES_n) \overset{d}{\to} N(0, \alpha^{-2}) \). Moreover, \( \mathcal{V}_n^2 = S_n^2(1 + o(1)) \) if \( \kappa > 2 \) and \( \mathcal{V}_n^2 = S_n^2(1 + O(1)) \) if \( \kappa \in (1, 2] \).

**Remark 1:** Since \( \hat{\kappa}_{m_n} \) does not affect the limit distribution of \( \hat{E}_{n}^{(1)} \), only \( y_{(k_n)}^{(-)} \) and \( \hat{E}_{n}^{(s)} \) matter asymptotically. The order statistic \( y_{(k_n)}^{(-)} / (-l_n) \) can be written as a sum of \( I(y_t < -l_n) - P(y_t < -l_n) \), cf. Hsing (1991: p. 1553), and \( \hat{E}_{n}^{(s)} \) is asymptotically a sum of \( y_{n,t}^{(+)} - E[y_{n,t}^{(+)}) \) by Lemma A.2 in Appendix A. The scale \( \mathcal{V}_n^2 = B_n' \Sigma_n B_n \) therefore has a classic sandwich form capturing the covariance between non-tail and tail variables \( y_{n,t}^{(+)} \) and \( I(y_t < -l_n) \) which form the asymptotic basis of \( \hat{E}_{n}^{(s)} \) and \( \hat{R}_{n}^{(1)} \).

**Remark 2:** The relative magnitude of \( \mathcal{V}_n^2 \) and \( S_n^2 \) is interesting. By construction and non-degeneracy Assumption N’ the long-run variance \( S_n^2 \) of \( y_{n,t}^{(+)} \) diverges if \( E[y_t^2] = \infty \), but the long-run variance of \( (n/k_n)^{1/2} \{ I(y_t < -l_n) - P(y_t < -l_n) \} \) is bounded (cf. Hill 2009, 2010). Moreover, \( (k_n/n)^{1/2} S_n^2 = O(1) \) since \( \kappa > 1 \), and if the variance is finite \( \kappa > 2 \) then \( S_n^2 = O(1) \) and \( (k_n/n)^{1/2} l_n = K(k_n/n)^{1/2-1/\kappa} \to 0 \). By combining these properties it must be the case that \( \mathcal{V}_n^2 = S_n^2(1 + o(1)) \) when \( \kappa > 2 \), and otherwise \( \mathcal{V}_n^2 = S_n^2(1 + O(1)) \). The general relationship \( \mathcal{V}_n^2 \sim KS_n^2 \) means the rate of convergence details from Section 3.1 carry over here verbatim. See Lemma A.1 in Appendix A for a compendium of tail and non-tail variance properties.

### 3.2 Optimal Bias Correction

A limitation of the tail mean estimators \( \hat{R}_n \) and \( \hat{R}_{n}^{(1)} \) is they rely on point estimates for their components. In practice a better approach is to choose \( \hat{R}_n^{(1)} \) that best approximates the tail mean \( E[y_t I(y_t < -l_n)] \), in particular such that \( \hat{E}_{n}^{(s)} + \hat{R}_n^{(1)} \) is closest to the untrimmed estimator \( \hat{E}_{n} \) := \( -1/(\alpha n) \sum_{t=1}^n y_t I(y_t \leq \hat{q}_n) \). This will eradicate asymptotic bias, and minimize incidental small sample bias that arises from trimming. Further, by assumption \( E|y_t| < \infty \) hence \( \kappa > 1 \), which means that we only use a point tail index estimator if \( \hat{\kappa}_{m_n} > 1 \). As it turns out, we can minimize \( |\hat{E}_{n}^{(s)} + \hat{R}_n^{(1)} - \hat{E}_{n}| \) over the fractile \( m_n \) used to compute \( \hat{\kappa}_{m_n} \) in the tail mean estimator.
Further, let we trim asymptotically normal ES estimator. As long as we impose \( k_n/m_n \) can be computed as a function of second order tail parameters (see Hall 1982, Huisman et al. 2001, Segers 2002). In our context evidently \( \{m_n, k_n\} \) can be chosen to minimize an asymptotic parametric mean-squared-error of \( \hat{ES}_n^{(s)} + \hat{R}_n^{(1)} \), but in the general mixing case little is known about tail estimator variances (see Hill 2010, 2011b). Our optimization method, however, has an obvious advantage since it minimizes bias and eradicates it asymptotically, and still results in an asymptotically normal ES estimator.

Define \( m_n(\lambda) = [\lambda m_n] \) where \( \{m_n\} \) is an intermediate order sequence and \( \lambda \) is taken from a compact set \( [\underline{\lambda}, \bar{\lambda}] \), \( 0 < \underline{\lambda} < \bar{\lambda} \). The new tail mean estimator

\[
\hat{R}_{n}^{(2)} = \hat{R}_{n}^{(2)}(\hat{\lambda}_n) := -\frac{1}{\alpha} \times \left( \frac{\hat{k}_{m_n(\lambda_n)}^{(-)} - k_n^{(-)}}{\hat{k}_{m_n(\lambda_n)}^{(-)} - 1/n y(k_n^{(-)})} \right)
\]

is a function of any \( \hat{\lambda}_n \) that solves the optimization problem

\[
\hat{\lambda}_n = \arg\min_{\lambda \in [\underline{\lambda}, \bar{\lambda}]} \left| \hat{ES}_{n}^{(s)} + \hat{R}_{n}^{(2)}(\lambda) - \hat{ES}_{n} \right|.
\]

Our second ES estimator with optimal bias correction is then

\[
\hat{ES}_{n}^{(2)} := \hat{ES}_{n}^{(s)} + \hat{R}_{n}^{(2)}.
\]

As long as we impose \( k_n/m_n \to 0 \) as in (5) the estimator \( \hat{ES}_{n}^{(2)} \) has the same limit distribution as \( \hat{ES}_{n}^{(s)} \), hence the correct scale is again \( V_n^2 \) in (7). Other fractile functions \( m_n(\lambda) \) can be considered, for example cadlag \( m_n(\lambda) \) that is non-decreasing in \( \lambda \) (see Hill 2009).

Notice in principle \( \hat{\lambda}_n \) may not be unique for finite \( n \) or even asymptotically. This is irrelevant since we show in the proof of the next claim that the Hill (1975) estimator process \( \{\hat{k}_{m_n(\lambda)} : \lambda \in [\underline{\lambda}, \bar{\lambda}]\} \) satisfies sup \( \lambda \in [\underline{\lambda}, \bar{\lambda}] \left| \hat{k}_{m_n(\lambda)} - \kappa \right| = o_p(1/k_n^{1/2}) \) as long we use more tail observations \( m_n \) than we trim \( k_n \) a la (5), hence \( \hat{k}_{m_n(\lambda_n^*)} = \kappa + o_p(1/k_n^{1/2}) \) for any sequence \( \{\lambda_n^*\} \) of possibly stochastic \( \lambda_n^* \) on \( [\underline{\lambda}, \bar{\lambda}] \). This is all that is required for Theorem 3.2 to go through.

**THEOREM 3.2.** Let Assumptions D, D' and T' hold, and assume \( k_n \) and \( m_n \) satisfy (5). Further, let \( \{\hat{\lambda}_n\} \) be any sequence on \( [\underline{\lambda}, \bar{\lambda}] \) that solves (8) for each \( n \). Then \( n^{1/2} V_n^{-1}(\hat{ES}_{n}^{(2)} - ES_{\alpha}) \overset{d}{\to} N(0, \alpha^{-2}) \).
Uniform asymptotics remain valid on any subset of $[\lambda, \bar{\lambda}]$. Consider the subset such that the tail index estimator is above unity:

$$\Lambda_n^* := \left\{ \lambda \in [\lambda, \bar{\lambda}] : \hat{\kappa}_{m_n(\lambda)}^{(-)} > 1 \right\}.$$ 

Then the proposed optimal fractile parameter is

$$\hat{\lambda}_n^* = \arg\min_{\lambda \in \Lambda_n^*} \left| \frac{\hat{E}_{n,\alpha}^{(*)}}{ES_n} + \hat{\mathcal{R}}_n^{(2)}(\lambda) - ES_n \right|. \quad (8)$$

It is easy to show by the argument used to prove Theorem 3.2 that $n^{1/2}V_n^{-1}(\hat{E}_{n,\alpha}^{(2)} - ES_n) \overset{d}{\rightarrow} \mathcal{N}(0, \alpha^{-2})$ when $\hat{R}_n^{(2)}(\hat{\lambda}_n^*)$ is the computed bias estimator.

### 3.3 Bias Correction - Thin Tails

The tail mean formula (4) crucially exploits the Assumption T power law tail form. If the left tail decays faster than any power law such that $P(y_t < -c)/c^{-\xi} \to 0 \ \forall \xi > 0$, for example if $y_t$ has exponentially decaying tails or a bounded distribution support, then (4) does not have meaning and is in general incorrect. Thus, the important questions are what exactly do $\hat{R}_n^{(1)}$ and $\hat{R}_n^{(2)}(\lambda)$ estimate? and what is their limiting behavior when probability tails do not satisfy a power law? Consider $\hat{R}_n^{(1)}$, while a similar argument applies to $\hat{R}_n^{(2)}(\lambda)$.

Assume for simplicity that $y_t$ has the following exponential tail

$$P(|y_t| \geq c) = \vartheta \exp\left\{-\varpi c^\delta\right\} \text{ where } \vartheta, \varpi, \delta \in (0, \infty). \quad (9)$$

The intermediate order statistic $y_{(k_n)}^{(-)}$ is well behaved for a larger variety of distributions (cf. David and Nagaraja 2003), while intuitively the moment supremum under (9) is $\arg\sup\{\alpha > 0 : E|y_t|^\alpha < \infty\} = \infty$ hence $\hat{\kappa}_{m_n}^{(-)} \overset{p}{\rightarrow} \infty$ may reasonably be expected. Thus the limiting properties of $\hat{R}_n^{(1)}$ should be characterizable. Indeed, in the thin tail case the scale is bounded $0 < \lim_{n \to \infty} V_n < \infty$ hence the conclusions of Theorem 3.1 and 3.2 remain valid as long as $n^{1/2} \hat{R}_n^{(1)} \overset{p}{\rightarrow} 0$. The next result shows this occurs as long as trimming is very light.

**THEOREM 3.3.** Let Assumptions D and N’ hold, assume (9), and let $k_n \to \infty$, $k_n = o(\ln(n))$, $m_n \to \infty$ and $m_n = o(n)$. Then $n^{1/2}V_n^{-1}(\hat{E}_{n,\alpha}^{(1)} - ES_n) \overset{d}{\rightarrow} \mathcal{N}(0, \alpha^{-2})$.

**Remark 1:** It is imperative that the trimming fractile $k_n$ increases very slowly such that $k_n = o(\ln(n))$. This ensures the tail-trimmed mean estimator $\hat{E}_{n,\alpha}^{(1)}$ dominates asymptotics by sufficiently elevating its rate of convergence relative to the sample tail exponent in the bias estimator $\hat{R}_n^{(1)}$. In view of Theorems 3.1 and 3.2 it is always valid to use $k_n = [\lambda_k(\ln(n))]^{1-\epsilon}$ and $m_n = [\lambda_m \ln(n)]$ for $\lambda_k, \lambda_m > 0$ and tiny $\epsilon > 0$: in the heavy tail case this satisfies Theorems 3.1 and
3.2 and in the light tail case it satisfies Theorem 3.3.

Remark 2: Notice Theorem 3.3 does not require a restriction on the tail fractile $m_n$ used for tail exponent estimation since power law Assumption T or T' no longer hold.

Remark 3: A similar proof can be constructed for the case where $y_t$ has bounded support.

3.4 Nonparametric Inference

Estimation of the scale $\mathcal{V}_n^2 = \mathcal{B}_n\mathcal{S}_n\mathcal{B}_n$ is straightforward since each component is based on a long-run variance and/or a tail quantile. Write compactly

\[ \hat{y}_{n,t}^* = y_t I\left(y_{(k_n)}^(-) \leq y_t \leq \hat{q}_n\right) \quad \text{and} \quad \hat{y}_n^* := \frac{1}{n} \sum_{t=1}^n \hat{y}_{n,t}^* \quad \text{and} \quad \hat{B}_n := \left[ 1, -\frac{1}{\hat{\kappa}_{m_n}(-)}\left(\frac{k_n}{n}\right)^{1/2}\hat{y}_{(k_n)}^(-) \right] \]

\[ \hat{W}_{n,t} := \left[ \hat{y}_{n,t}^* - \hat{y}_n^*, \left(\frac{n}{k_n}\right)^{1/2} I\left(y_t < \hat{y}_{(k_n)}^(-) - \frac{k_n}{n}\right) \right] \]

and define kernel estimators

\[ \hat{S}_n^2 := \frac{1}{n} \sum_{s,t=1}^n K((s-t)/\gamma_n) \{ \hat{y}_{n,s}^* - \hat{y}_n^* \} \{ \hat{y}_{n,t}^* - \hat{y}_n^* \} \quad \text{and} \quad \hat{\Sigma}_n := \frac{1}{n} \sum_{s,t=1}^n K((s-t)/\gamma_n) \hat{W}_{n,s} \hat{W}_{n,t} \]

where $K(x)$ is a kernel function with $K(0) = 1$, and the bandwidth $\gamma_n$ satisfies $\gamma_n \to \infty$ and $\gamma_n = o(n)$. Note $\hat{S}_n^2$ estimates the scale $\mathcal{S}_n^2$ for the core estimator $\hat{ES}_n^{(s)}$. Now build a scale estimator for the bias-corrected ES estimators:

\[ \hat{V}_n^2 := \hat{B}_n^t \hat{\Sigma}_n \hat{B}_n. \]

The following properties cover Tuckey-Hanning, Parzen and Bartlett kernels, to name a few. See Hill (2010, 2012a,b) for references.

ASSUMPTION K (kernel and bandwidth). Assume $K(\cdot)$ is continuous at 0 and all but a finite number of points, $K: \mathbb{R} \to [-1, 1], K(0) = 1, K(x) = K(-x) \forall x \in \mathbb{R}, \int_{-\infty}^{\infty} |K(x)|dx < \infty$, and $\int_{-\infty}^{\infty} |\varphi(\xi)|d\xi < \infty$. Let $\sum_{s,t=1}^n |K((s-t)/\gamma_n)| = o(n^2), \max_{1 \leq s \leq n} \sum_{t=1}^n K((s-t)/\gamma_n) = o(n)$ and bandwidth $\gamma_n = o(n)$.

THEOREM 3.4. Under Assumptions D, K, N' and T' $\hat{S}_n^2/\mathcal{S}_n^2 \stackrel{p}{\to} 1$ and $\hat{V}_n^2/\mathcal{V}_n^2 \stackrel{p}{\to} 1$.

4 SIMULATION STUDY

We now perform a Monte Carlo experiment to investigate the merits of our estimators. We study both the core estimator $\hat{ES}_n^{(s)} = -\alpha^{-1}n^{-1} \sum_{t=1}^n y_t I(y_{(k_n)}^(-) \leq y_t \leq \hat{q}_n)$ and the optimal bias corrected estimator $\hat{ES}_n^{(2)} = \hat{ES}_n^{(s)} + \hat{R}_n^{(2)}$. 
Let $N_{0,1}$ denote a standard normal distribution, and let $P_\kappa$ denote a symmetrized Pareto distribution: if $\epsilon_t \sim P_\kappa$ then $P(\epsilon_t > c) = P(\epsilon_t < -c) = 0.5(1 + c)^{-\kappa}$ for $\kappa > 0$. The models are

$$\text{AR(2)}: y_t = 0.8y_{t-1} - 0.3y_{t-2} + \epsilon_t, \text{ where } \epsilon_t \sim P_\kappa \text{ with } \kappa_t \in \{1.5, 2.5\}, \text{ or } \epsilon_t \sim N_{0,1}$$

$$\text{GARCH(1,1)}: y_t = \sigma_t \epsilon_t, \text{ where } \sigma_t^2 = 1 + 0.3y_{t-1}^2 + 0.6\sigma_{t-1}^2, \text{ and } \epsilon_t \sim P_{2.5} \text{ or } \epsilon_t \sim N_{0,1}$$

$$\text{IGARCH(1,1)}: y_t = \sigma_t \epsilon_t, \text{ where } \sigma_t^2 = 1 + 0.4y_{t-1}^2 + 0.6\sigma_{t-1}^2 \text{ and } \epsilon_t \sim N_{0,1}$$

$$\text{RiskMetric IGARCH}: y_t = \sigma_t \epsilon_t, \text{ where } \sigma_t^2 = 1 + 0.6y_{t-1}^2 + 0.94\sigma_{t-1}^2 \text{ and } \epsilon_t \sim N_{0,1}.$$  

In all GARCH cases we standardize Pareto errors such that $E[\epsilon_t^2] = 1$. The RiskMetric volatility model follows from JP Morgan’s parametric convention that reduces to an IGARCH with $\sigma_t^2 = \omega + 0.6y_{t-1}^2 + 0.94\sigma_{t-1}^2$ for some $\omega > 0$. See Christoffersen (2012: §2). Each process $\{y_t\}$ is stationary, ergodic, geometrically $\alpha$-mixing, and has a Pareto tail $P(|y_t| > c) = dc^{-\kappa_t}(1 + o(1))$, cf. Brockwell and Cline (1985), Pham and Tran (1985), and Mikosch and Stårică (2000). We simulate 10,000 samples of sizes $n \in \{100, 500, 1000\}$, and estimate the ES of $y_t$ evaluated at a levels of risk $\alpha \in \{0.05, 0.10\}$. All estimators work qualitatively similarly for other small levels of risk. e.g. $\alpha \in \{0.01, 0.025\}$. Since all model parameters are known we compute the tail index $\kappa_y$ of $y_t$ and the corresponding true ES, which we use for simulation bias computation.$^{10,11}$

We compute the untrimmed $\hat{ES}_n$, the core trimmed $\hat{ES}_n^{(*)}$ with fractile $k_n = \min\{1, \lfloor 25n^{2/3}/(\ln(n))^{2/3} \rfloor\}$ and $\ell = 10^{-10}$, and the bias corrected $\tilde{ES}_n^{(2)}$, where the tail index estimator $\hat{k}_{m_n}(\lambda)$ is based on the fractile function $m_n(\lambda) = \min\{1, \lfloor \lambda k_n(\ln(n))^{4/3} \rfloor\}$ and $\lambda \in [\Delta, \bar{\Delta}] = [0.05, 10]$. The interval $[0.05, 10]$ is discretized with increments $1/n$.

In a few cases from an entire sample draw $\{y_t\}_{t=1}^n$ all tail index estimates $\hat{k}_{m_n}(\lambda) \in (0, 1]$ which is too small for our bias estimator to be valid: we simply throw out such samples since $\kappa \leq 1$ is ruled out. This is reasonable since in practice if there is compelling evidence that $\kappa \leq 1$ then the ES may not exist.

Let $\hat{\Delta}_{t,n}$ denote any particular ES estimator for the $t^{th}$ simulated sample out of $N$ samples. We report the simulation bias $1/N\sum_{i=1}^N \hat{\Delta}_{t,n} - ES_\alpha$, root mean-squared-error $s_N := (1/N\sum_{i=1}^N (\hat{\Delta}_{t,n} - ES_\alpha)^2)^{1/2}$ for AR(2), and $\sigma_t^2 = 1$ for GARCH(1,1).

$^9$We draw $2n$ observations of $y_t$ and retain the last $n$ observations. Start values are $y_0 = y_{-1} = 0$ for AR(2), and $\sigma_t^2 = 1$ for GARCH(1,1).

$^{10}$The true $ES_\alpha$ is computed by drawing $N = 100,000$ observations of $y_t$ and computing $1/N\sum_{t=1}^N I(y_t < y_{(n,N)})$. We repeat this 10,000 times and use the median value for bias computation.

$^{11}$The tail index for a stationary AR is the same as the iid error (Brockwell and Cline 1985). In the GARCH case $P(|y_t| > y) = dy^{-\kappa_t}(1 + o(1))$ where $E[|\epsilon_t|^k] = 1/(3\kappa_t + 6)^{k/2}$ (Basrak et al 2002). We draw $N = 100,000$ iid $\epsilon_t$ from $P_{2.5}$ or $N(0,1)$ and compute $\kappa = \arg\min_{\kappa \in K} \{1/N\sum_{t=1}^N |\epsilon_t|^k - 1\}$ where $K \equiv \{0.001, 0.002, ..., 10\}$. We repeat this 10,000 times and report the median value under $\kappa_y$ in Tables 1-3. In the Gaussian IGARCH case $\kappa_y = 2$. See Mikosch and Stårică (2000).
and perform Kolmogorov-Smirnov and Anderson-Darling tests of standard normality on the standardized variable \((\hat{E}_{i,n} - 1/N \sum_{i=1}^N \hat{E}_{i,n})/\hat{s}_N\) where \(\hat{s}_N^2 := (1/N \sum_{i=1}^N (\hat{E}_{i,n} - 1/N \sum_{i=1}^N \hat{E}_{i,n})^2)^{1/2}\). Finally, we use the simulation ratio \(1/N^{1/2} \sum_{i=1}^N (\hat{E}_{i,n} - e_0)/\hat{s}_N\) to perform asymptotic tests of whether the true ES is \(e_0 \in \{ES_0, ES_0/2, 0\}\).

The simulation results are presented in Tables 1 and 2 for \(\alpha = .05\) and .10 respectively. The core estimator \(\widehat{ES}_n^{(1)}\) exhibits substantial bias as predicted, with essentially no change as \(n\) increases. In fact, there must be asymptotic bias for any \(\kappa > 1\) since \(k_n\) is regularly varying: see the discussions in Sections 2 and 3. Bias is essentially eradicated with \(\widehat{ES}_n^{(2)}\) since it exploits an optimally fitted tail mean estimator that renders \(\widehat{ES}_n^{(2)}\) close to the untrimmed \(\widehat{ES}_n\). The untrimmed estimator \(\widehat{ES}_n\) exhibits the greatest deviations from normality and the largest empirical size distortions relative to the nominal size when the variance is infinite. Of all estimators \(\widehat{ES}_n^{(2)}\) has the smallest RMSE, it is closest to normal by either normality test, and when used for inference its t-ratio exhibits the sharpest empirical size and exceptional power. Finally, our bias corrected \(\widehat{ES}_n^{(2)}\) compares well with the untrimmed estimator \(\widehat{ES}_n\) when the variance is finite. This suggests trimming does not impose a detectable penalty in terms of small sample mean-squared-error.

5 Empirical Application

Finally, we estimate ES for several financial returns series. We study the stock market Hang Seng Index [HSI] and the Russian Ruble - U.S. Dollar exchange rate considering evidence of infinite variance returns over different periods (Ling 2005, Ibragimov et al 2010).\(^{12}\) The period for the Ruble is Jan. 1, 1999 - Oct. 29, 2008 as in Ibragimov et al (2010), spanning a period between major financial crises in Russia and globally.\(^{13}\) The period for HSI is June 3, 1996 - May 31, 1998 as in Ling (2005). We take each series \(\{x_t\}\) and compute the daily log returns \(y_t = \ln(x_t) - \ln(x_{t-1})\), resulting in 2449 and 489 returns for the Ruble and HSI respectively.

In Figure 1 we plot the returns series \(\{y_t\}_{t=1}^n\) and Hill (1975) estimator 95% confidence bands \(\kappa_{ma} \pm 1.96\hat{\sigma}_{ma}/m^{1/2}\), where \(\hat{\sigma}_{ma}^2\) is a kernel estimator of the mean-squared-error of the inverted estimator \(E(m_n^{1/2}(\kappa_{ma}^{-1} - \kappa)^{-1}))\). In particular \(\hat{\sigma}_{ma}^2 = 1/m_n \sum_{s,t=1}^n K((s-t)/b_n)\{\ln(y_s^{(a)}/y_{(m_n+1)}^{(a)}) + (m_n/n)\kappa_{ma}^{-1}\} \times \{(\ln(y_t^{(a)}/y_{(m_n+1)}^{(a)}) + (m_n/n)\kappa_{ma}^{-1}\} with Bartlett kernel \(K()\) and bandwidth \(b_n = n^{-25}\). See Hill (2010: Theorem 3) for a proof of consistency under mixing Assumption D and tail decay Assumption T'. Each series displays evidence of clustered volatility and extreme spikes. Activity in the Ruble corresponds to the recovery following the 1998 crisis and subsequent Ruble devaluation, and the mounting crisis in late 2008. The tail index estimates for both series are predominantly near or below 2 and the confidence bands contain 2 for all \(m_n \in \{5, \ldots, 200\}\).

\(^{12}\)We thank Rustam Ibragimov for providing us the Ruble-Dollar exchange rate data.

\(^{13}\)In 1998 Russia experienced the so-called Ruble Crisis, with a major devaluation occurring in August 1998. In 2008 a global financial crisis emerged from U.S. banking and investment practices.
We now turn to value-at-risk and expected shortfall computation for the risk levels \( \alpha \in \{0.05, 0.10\} \). We follow Linton and Xiao (2012) and compute the VaR \( \hat{q}_{n, \alpha} \) and bias-corrected ES estimator \( \hat{ES}_{n}^{(2)} \) over backward looking rolling windows of size 250 days or about one year. There are therefore 2200 and 240 windows respectively for the Ruble and HSI. See Figures 2 and 3. We also include a rolling window surface plot of the Hill (1975) estimator over the number of tail observations \( m_{n} \in \{5, \ldots, 200\} \) and window. In this way we can see how risk dynamics are associated with heaviness of returns tails. See Figure 4.

Each estimator is computed as in Section 4. We include 95% confidence bands of the VaR \(-\hat{q}_{n, \alpha}\) and \(\hat{ES}_{n}^{(2)} \). In the case of \(\hat{ES}_{n}^{(2)} \) we use the kernel estimator \(\hat{V}_{n}^{2}\) from Section 3.4 to compute \(\hat{ES}_{n}^{(2)} = 1.96 \alpha^{-1} \hat{V}_{n} / n^{1/2}, \) with Bartlett kernel and bandwidth \(\gamma_{n} = n^{25}\).

In the case of the intermediate order statistic \(\hat{q}_{n, \alpha} = y_{\lfloor \alpha n \rfloor} \) and subsequent VaR \(-\hat{q}_{n, \alpha} > 0\) the computed confidence band is \(-\hat{q}_{n, \alpha} \pm 1.96 \hat{V}_{n} / n^{1/2}\) where \(\hat{V}_{n}^{2}\) is constructed by first deriving the limit distribution of \(\hat{q}_{n, \alpha}\). Let \(f(y)\) be the pdf of \(y\), assume \(f(y)\) is differentiable, and define \(I_{t}(\xi) := I(y_{t} > q_{a} \xi)\). By replicating the argument in Hsing (1991: p. 1553) for an intermediate order statistic, it is easily verified that \(n^{1/2}(y_{\lfloor \alpha n \rfloor} - q_{a}) \leq \xi\) for any \(\xi \in \mathbb{R}\) if and only if \(I_{n}(\xi/n^{1/2}) := 1/n^{1/2} \sum_{t=1}^{n} \{I_{t}(\xi/n^{1/2}) - E[I_{t}(\xi/n^{1/2})] \} / f(q_{a}) \leq \xi + o(1)\). If we define \(\hat{V}_{n}^{2}(\xi) := E[I_{n}^{2}(\xi)]\) then by mixing Assumption D \(\mathbb{I}_{n}(\xi/n^{1/2})/\mathbb{V}_{n}(\xi/n^{1/2}) \xrightarrow{d} N(0, 1)\), hence by dominated convergence \(n^{1/2}(y_{\lfloor \alpha n \rfloor} - q_{a})/\hat{V}_{n}(0) \xrightarrow{d} N(0, 1)\). We therefore estimate \(\mathbb{V}_{n}(0)\) with the kernel estimator \(\hat{V}_{n}\) of the form \(\hat{V}_{n} = 1/n \sum_{s,t=1}^{n} K(s-t)/\gamma n \{I(y_{t} \leq \hat{q}_{n, \alpha}) - \alpha\} / f_{n}(\hat{q}_{n, \alpha})\) where \(f_{n}(y)\) is a Gaussian kernel estimator of the pdf \(f(y)\) with bandwidth \(\gamma\), cf. Silverman (1986).

Both VaR and ES estimates capture roughly the same patterns of increasing and decreasing risk. The Ruble displays decreasing risk following the 1998 crisis as market volatility decreases, and increasing risk as the 2008 crisis emerged. This is matched by smaller tail index values over most fractiles in the months near 1999 and 2008, showing higher risk coincides with heavier tails. Similarly, the HSI tail index estimates are smaller in windows with higher risk.

6 CONCLUSION

We develop Expected Shortfall estimators that can be used for standard inference even if the data are heavy tailed. Although bootstrap methods in the heavy tailed case exist (Linton and Xiao 2012), our estimators are aimed at computational simplicity since they lead to classic inference irrespective of heavy tails. We combine classic tail-trimming with an improved bias-correction method based on Peng’s (2001, 2004) work for power law distributions to deliver an estimator that is asymptotically unbiased and normal even if the probability tails do not satisfy a power law. Further, we exploit an empirical process method to best approximate the removed tail mean. This leads to a new estimator that can be made arbitrarily close to the untrimmed estimator as the sample grows, and yet be asymptotically normal. A simulation study shows the new bias-corrected
APPENDIX A: Proofs of Main Results

We require several preliminary results proved in Appendix B. Define tail and non-tail indicator variables

\[ \hat{I}_{n,t} := I(y_t \geq y_{(k_n)}^{(c)}) \quad \text{and} \quad I_{n,t} := I(y_t \geq -l_n) \]

\[ \mathcal{I}_{n,t} := \left( \frac{n}{k_n} \right)^{1/2} \{ I(y_t < -l_n) - P(y_t < -l_n) \}. \]

Stack non-tail and tail variables and define their long-run covariance matrix

\[ W_{n,t} := \left[ y_{n,t}^* - E[y_{n,t}^*], \mathcal{I}_{n,t} \right]' \quad \text{and} \quad \Sigma_n := \frac{1}{n} \sum_{s,t=1}^{n} E[W_{n,s}W_{n,t}']. \]

Recall \( S_n^2 := E(n^{-1/2} \sum_{t=1}^{n} \{ y_{n,t}^* - E[y_{n,t}^*] \})^2 \).

**LEMMA A.1 (variance rate).** Let Assumptions D and T hold.

a. \( S_n^2 = r_n E[y_{n,t}^2] = O(n) \) where \( \{ r_n \} \) is a sequence of positive numbers \( \lim_{n \to \infty} r_n > 0, r_n = O(\ln(n)) \), \( r_n \) does not depend on \( k_n \), and \( r_n \sim K \) if \( y_t \) is \( m \)-dependent or \( E[y_t^2] < \infty \).

b. If \( \kappa = 2 \) then \( E[y_{n,t}^2] \sim K \ln(n) \), and if \( \kappa \in (1, 2) \) then \( E[y_{n,t}^2] \sim \kappa(2 - \kappa)^{-1} d^2/\kappa(n/k_n)^{2/\kappa-1} \).

c. \( E(n^{-1/2} \sum_{t=1}^{n} \mathcal{I}_{n,t})^2 = r_n E[\mathcal{I}_{n,t}^2] \) where \( \lim \inf_{n \to \infty} r_n > 0 \) and \( r_n = O(1) \).

**LEMMA A.2 (intermediate order statistic approximation).** Let \( \{ a_{n,t} : 1 \leq t \leq n \}_{n \geq 1} \) be any stochastic triangular array where \( a_{n,t} \) is \( \mathcal{F}_t \)-measurable and \( \max_{1 \leq t \leq n} \{ |a_{n,t}| \} = O_p(1) \). Under Assumptions D, N and T \( n^{-1/2} S_n^{-1} \sum_{t=1}^{n} a_{n,t} y_{t} \{ \hat{I}_{n,t} - I_{n,t} \} = o_p(1) \).

**LEMMA A.3 (central order statistic approximation).** Under Assumptions D, N and T

\[ \frac{1}{n^{1/2}S_n} \sum_{t=1}^{n} y_t I_{n,t} \{ I(y_t \leq y_{(|t|n)}) \} - I(y_t \leq q_n) \} = o_p(1). \]

**LEMMA A.4 (intermediate order statistic expansion).** Let Assumptions D and T hold. For any positive sequences \( \{ l_n, k_n \} \) that satisfy \( k_n \to \infty, k_n = o(n) \) and \( (n/k_n) P(y_t < -l_n) \to 1 \) we have \( k_n^{1/2}(y_{(k_n)}^{(c)}(-l_n) - 1) = \kappa^{-1} n^{-1/2} \sum_{t=1}^{n} \mathcal{I}_{n,t} \times (1 + o_p(1)) = O_p(1) \).

**LEMMA A.5 (central limit theorem).** Under Assumptions D, N' and T \( n^{-1/2}(\lambda'\Sigma_n\lambda)^{-1/2} \times \sum_{t=1}^{n} \lambda'W_{n,t} \overset{d}{\to} N(0,1) \) for any \( \lambda \in \mathbb{R}^2 \) such that \( \lambda' \lambda = 1 \).
Remark 1: If $\lambda = [1, 0]$ then $\lambda'\Sigma_n\lambda' = S_n^2$ hence $n^{-1/2}S_n^{-1} \sum_{t=1}^n \{y_{n,t}^* - E[y_{n,t}^*]\} \xrightarrow{d} N(0, 1)$.

Remark 2: By the Cramér-Wold Theorem $n^{-1/2} \sum_{t=1}^n W_{n,t} \xrightarrow{d} N(0, I_2)$.

We are now ready to prove the main results.

PROOF OF LEMMA 2.1. Each claim follows from Lemma A.1.a,b. \(\Box\)

PROOF OF THEOREM 2.2. Recall $y_{n,t}^* := y_t I(-l_n \leq y_t \leq q_\alpha)$. By the construction of $\widehat{ES}_n^{(s)}$ use intermediate order statistic approximation Lemma A.2 with $a_{n,t} = I(y_t \leq y_{\lfloor an \rfloor})$ to deduce

$$
\frac{n^{1/2}}{S_n/\alpha} \left\{ \widehat{ES}_n^{(s)} - E[y_{n,t}^*] \right\} = -\frac{1}{n^{1/2}S_n} \sum_{t=1}^n \{y_{n,t}^* - E[y_{n,t}^*]\} - \frac{1}{n^{1/2}S_n} \sum_{t=1}^n y_{n,t} I_{n,t} \left\{ I \left(y_t \leq y_{\lfloor an \rfloor}\right) - I \left(y_t \leq q_\alpha\right) \right\} + o_p(1)
$$

$$
= A_n + B_n + o_p(1),
$$

say. By central order statistic approximation Lemma A.3 $B_n = o_p(1)$. Now use $S_n = o(n^{1/2})$ by Lemma A.1.a to obtain

$$
\widehat{ES}_n^{(s)} - ES_\alpha = -\frac{1}{\alpha} E[y_{n,t}^*] - ES_\alpha - \frac{1}{\alpha} \times \frac{1}{n} \sum_{t=1}^n \{y_{n,t}^* - E[y_{n,t}^*]\} + o_p(1).
$$

The first claim $\widehat{ES}_n^{(s)} \xrightarrow{p} ES_\alpha$ now follows from $-\alpha^{-1} E[y_{n,t}^*] \rightarrow ES_\alpha$ by dominated convergence, and $1/n \sum_{t=1}^n \{y_{n,t}^* - E[y_{n,t}^*]\} \xrightarrow{p} 0$ by Chebyshev’s inequality in view of $S_n^2 = o(n)$ by Lemma A.1.a.

Next, apply Remark 1 of central limit theorem Lemma A.5 to $A_n$ to conclude

$$
\frac{n^{1/2}}{S_n/\alpha} \left( \widehat{ES}_n^{(s)} + \frac{1}{\alpha} E[y_{n,t}^*] \right) = A_n + o_p(1) \xrightarrow{d} N(0, 1).
$$

Since $\alpha^{-1} E[y_{n,t}^*] = -ES_\alpha + \alpha^{-1} E[y_t I(y_t < -l_n)]$ the second claim is proved. The third claim follows from (11) and identification Assumption I. \(\Box\)

PROOF OF THEOREM 3.1. Define

$$
\gamma_{n,t}^* := y_{n,t}^* - E[y_{n,t}^*] \quad \text{and} \quad \mathcal{I}_{n,t} := \left( \frac{n}{kn} \right)^{1/2} \left\{ I(y_t < -l_n) - P(y_t < -l_n) \right\},
$$
We first relate $\mathcal{V}_n^2$ to $\mathcal{S}_n^2$ and then derive the distribution limit of $\hat{ES}_{n}^{(1)}$.

**Step 1 ($\mathcal{V}_n^2$):** Note

$$\mathcal{V}_n^2 = \mathcal{B}'_n \Sigma \mathcal{B}_n = \mathcal{S}_n^2 - 2 \frac{1}{\kappa - 1} \left( \frac{k_n}{n} \right)^{1/2} I_n \mathcal{H}_n + \frac{1}{(\kappa - 1)^2} \frac{k_n}{n} Y_n \sigma_n^2.$$  

By Lemma A.1.a,c $\lim \inf_{n \to \infty} \mathcal{S}_n^2 > 0$ and $\sigma_n^2 = O(1)$, hence $\sigma_n^2 = O(\mathcal{S}_n^2)$ and therefore by the Cauchy-Schwartz inequality $\mathcal{H}_n = O(\mathcal{S}_n^2)$. Moreover, by power law tail decay Assumption D and the threshold construction (2) it follows $l_n/(n/k_n)^{1/\kappa} = K > 0$ hence the $\mathcal{B}_n$ term $(k_n/n)^{1/2} l_n \sim K (k_n/n)^{1/2-1/\kappa}$. Therefore $(k_n/n)^{1/2} l_n \to 0$ if $\kappa > 2$, $(k_n/n)^{1/2} l_n \to K > 0$ if $\kappa = 2$, and $(k_n/n)^{1/2} l_n \to \infty$ if $\kappa > 2$. Finally, if $\kappa \in (1,2]$ then by Lemma A.1.a,b $\mathcal{S}_n^2 \sim r_n (n/k_n)^{2/\kappa-1}$ where $r_n = O(\ln(n))$ and $\lim \inf_{n \to \infty} r_n > 0$, hence $(k_n/n)^2/\mathcal{S}_n^2 = O(1)$.

It now follows by case $\mathcal{V}_n^2 = \mathcal{S}_n^2(1 + o(1))$ if $\kappa > 2$ and $\mathcal{V}_n^2 = \mathcal{S}_n^2(1 + O(1))$ if $\kappa \in (1,2]$.

**Step 2 ($\hat{ES}_{n}^{(1)}$):** By the proof of Theorem 2.1 $\hat{ES}_{n}^{(s)}$ obtains the expansion

$$
\frac{n^{1/2}}{\mathcal{S}_n/\alpha} \left( \hat{ES}_{n}^{(s)} - \mathcal{E}\mathcal{S}_n \right) = -\frac{1}{n^{1/2} \mathcal{S}_n} \sum_{t=1}^{n} \mathcal{Y}_{n,t}^* + \frac{n^{1/2}}{\mathcal{S}_n} \left( -E \left[ \mathcal{Y}_{n,t}^* \right] - \mathcal{E}\mathcal{S}_n \right) + o_p(1). \tag{12}
$$

Under Assumptions D and T' it follows $\mathcal{k}_{mn}(-) = \kappa + O_p(1/m_n^{1/2})$ by Theorem 3 in Hill (2010), hence

$$
\frac{\mathcal{k}_{mn}(-)}{\mathcal{k}_{mn}(-) - 1} \frac{\mathcal{Y}_{n}(-)}{n^{1/2}} = -l_n \frac{\kappa}{\kappa - 1} \frac{k_n}{n} \frac{1}{\kappa - 1} \frac{k_n^{1/2}}{n} l_n \times k_n^{1/2} \left( \frac{\mathcal{Y}_{n}(-)}{l_n} / l_n + 1 \right) + \left( \frac{\mathcal{k}_{mn}(-)}{\mathcal{k}_{mn}(-) - 1} - \frac{\kappa}{\kappa - 1} \right) \frac{k_n^{1/2}}{n} l_n \times k_n^{1/2} \left( \frac{\mathcal{Y}_{n}(-)}{l_n} / l_n + 1 \right) - \left( \frac{\mathcal{k}_{mn}(-)}{\mathcal{k}_{mn}(-) - 1} - \frac{\kappa}{\kappa - 1} \right) \frac{k_n}{n} l_n
$$

$$
= -\frac{\kappa}{\kappa - 1} \frac{k_n}{n} l_n + \frac{\kappa}{\kappa - 1} \frac{k_n^{1/2}}{n} l_n \times k_n^{1/2} \left( \frac{\mathcal{Y}_{n}(-)}{l_n} / l_n + 1 \right) + O_p \left( m_n^{-1/2} k_n / n \right).
$$
Further, by intermediate order statistic expansion Lemma A.4

$$k_n^{1/2} \left( y_{(k_n)} / l_n + 1 \right) = -\kappa^{-1} \frac{1}{n^{1/2}} \sum_{t=1}^{n} I_{n,t} \times (1 + o_p(1)).$$

Finally, by tail mean formula (4)

$$E \left[ y_t I (y_t < -l_n) \right] = -\frac{\kappa}{\kappa - 1} \frac{k_n}{n} l_n \times (1 + o(1))$$

and by construction

$$E [y_{n,t}] - \frac{\kappa}{\kappa - 1} \frac{k_n}{n} l_n - ES_\alpha = -\frac{\kappa}{\kappa - 1} \frac{k_n}{n} l_n - E [y_t I (y_t < -l_n)].$$

Combine (12)-(15) and simplify to obtain

$$\frac{n^{1/2}}{V_n/\alpha} (\bar{E}S_n^{(1)} - ES_\alpha) = \frac{n^{1/2}}{V_n/\alpha} \left( \bar{E}S_n + \frac{1}{\alpha} \left( \frac{\kappa}{\kappa m_n} \frac{k_n}{n} y_{(k_n)} \right) - ES_\alpha \right)$$

$$= \left\{ \frac{1}{n^{1/2} V_n} \sum_{t=1}^{n} y_{n,t} - \frac{1}{\kappa - 1} \frac{l_n}{V_n} \left( \frac{k_n}{n} \right)^{1/2} l_n - \frac{n^{1/2}}{V_n} \left[ E [y_t I (y_t < -l_n)] + \frac{\kappa}{\kappa - 1} \frac{k_n}{n} l_n \right] + O_p \left( \frac{1}{V_n} \left( \frac{n}{m_n} \right)^{1/2} \frac{k_n}{n} l_n \right) + o_p(1) \right\}$$

$$= \mathcal{Z}_n + A_{1,n} + A_{2,n} + o_p(1),$$

say. We will show $A_{1,n} = o(1)$, $A_{2,n} = o_p(1)$ and $\mathcal{Z}_n \overset{d}{\to} N(0,1)$.

Consider $A_{1,n}$. Peng’s (2001: p. 259-264) arguments in conjunction with non-degeneracy Assumption N’ and tail decay Assumption T’ prove

$$\left( \frac{n}{E [y_{n,t}^{*2}]} \right)^{1/2} \left( E [y_t I (y_t < -l_n)] + \frac{\kappa}{\kappa - 1} \frac{k_n}{n} l_n \right) \rightarrow 0.$$
Step 1. Therefore
\[
\frac{1}{V_n} \left( \frac{n}{m_n} \right)^{1/2} \frac{\kappa_n}{n} l_n \leq K \left( \frac{n}{m_n} \right)^{1/2} \left( \frac{n}{\kappa_n} \right)^{1/\kappa - 1} = K \frac{\kappa_n^{1-1/\kappa}}{m_n^{1/2}} \frac{1}{n^{1/2 - 1/\kappa}} \leq K \left( \frac{\kappa_n}{m_n} \right)^{1-1/\kappa}.
\]

Since by assumption \( \kappa > 1 \) and \( \kappa_n/m_n = o(1) \) it follows \( A_{2,n} = o_p(1) \).

Finally, by construction and Remark 2 of Lemma A.5

\[
Z_n = \frac{1}{n^{1/2} V_n} \sum_{i=1}^{n} \gamma_{n,i}^* - \frac{1}{\kappa - 1} \left( \frac{\kappa_n}{n} \right)^{1/2} l_n \sum_{i=1}^{n} f_{n,i} = \frac{1}{n^{1/2} V_n} \sum_{i=1}^{n} W_{n,i} \Rightarrow \mathcal{N}(0, 1). \quad \text{QED}
\]

**PROOF OF THEOREM 3.2.** Recall \( 0 < \underline{\lambda} < \bar{\lambda} \) and \( m_n(\lambda) = [\lambda m_n] \). Since \( \hat{\lambda}_n = \text{arginf}_{\lambda \in [\underline{\lambda}, \bar{\lambda}]} [\tilde{E}_{S_n}^{(\kappa)} + \tilde{K}_n^{(2)}(\lambda) - \tilde{E}_S] \) operates only on \( \hat{\kappa}_{m_n}^{(\lambda)} \) in \( \tilde{K}_n^{(2)}(\lambda) \), consider the weak limit of \( \{ m_n^{1/2}(\hat{\kappa}_{m_n}^{(-1)} - \kappa^{-1}) : \lambda \in [\underline{\lambda}, \bar{\lambda}] \} \). Define \( \sigma_n^2(\lambda) := E[(m_n^{1/2}(\lambda)(\hat{\kappa}_{m_n}^{(-1)} - \kappa^{-1}))^2] \). Under the maintained assumptions and Theorem 5.1 in Hill (2009) there exists a zero mean Gaussian process \( \{ Z(\lambda) : \lambda \in [\underline{\lambda}, \bar{\lambda}] \} \) on \( C[\underline{\lambda}, \bar{\lambda}] \) such that \( \{ m_n^{1/2}(\hat{\kappa}_{m_n}^{(-1)} - \kappa^{-1}) : \lambda \in [\underline{\lambda}, \bar{\lambda}] \} \Rightarrow \{ Z(\lambda) : \lambda \in [\underline{\lambda}, \bar{\lambda}] \} \) where \( \Rightarrow \) denotes weak convergence on \( C[\underline{\lambda}, \bar{\lambda}] \) the space of continuous functions. Hence \( \sup_{\lambda \in [\underline{\lambda}, \bar{\lambda}]} | \hat{\kappa}_{m_n}^{(-1)} - \kappa | = O_p(1/\inf_{\lambda \in [\underline{\lambda}, \bar{\lambda}]} m_n^{1/2}(\lambda)) \) by the continuous mapping theorem, which is \( o_p(1/k_n^{1/2}) \) given \( k_n/m_n \to 0 \). The argument used to prove Theorem 2.4 now carries over verbatim. QED.

**PROOF OF THEOREM 3.3.** None of the following arguments depend on the sign of \( y_t \), so assume \( y_t > 0 \) a.s. for notational convenience. Therefore all components of the bias corrected estimator are positive, in particular \( \{ l_n \} \) satisfies \( P(y_t > l_n) = k_n/n \).

Write \( \hat{\kappa}_{kn} = \kappa_k^{(-)} \) and \( y(\kappa_n) = y(\kappa_n^{'}) \), and observe for any tiny \( \epsilon > 0 \),
\[
n^{1/2} \hat{\kappa}_{mn}^{(1)} = \frac{1}{\alpha} \left( \hat{\kappa}_{mn}^{(-)} \frac{k_n}{l_n y(\kappa_n^{'})} \right).
\]

It suffices to show \( l_n/n^\epsilon \to 0 \), \( y(\kappa_n)/l_n \to P_1 \), and \( 1/\hat{\kappa}_{mn} \to 0 \). By supposition \( k_n/n^{1/2 - \epsilon} \to 0 \) hence the proof is complete.

**Step 1 (\( l_n \)).** By exponential tail decay \( P(y_t \geq c) = \vartheta \exp\{-\vartheta c^\delta\} \) in (9), and the definition of \( l_n \), it follows
\[
l_n = \left( \frac{1}{\alpha} \ln \vartheta + \frac{1}{\alpha} \ln \frac{n}{\kappa_n} \right)^{1/\delta} \sim \frac{1}{\alpha^{1/\delta}} \left( \ln (n) \right)^{1/\delta} = o(n^\epsilon).
\]

(16)
Step 2 \( (y_{(k_n)}) \). We will prove \( k_n^{1/2} \ln(y_{(k_n)}/l_n) = O_p(1) \) hence \( y_{(k_n)}/l_n = 1 + O_p(1/k_n^{1/2}) \). Define \( I_{n,u}(u) := (n/k_n)I(y_t > l_ne^u) \) for any \( u \in \mathbb{R} \), and \( I_{n}(u) := 1/n \sum_{t=1}^{n} I_{n,t}(u) \). By construction \( k_n^{1/2} \ln(y_{(k_n)}/l_n) \leq u \) for \( u \in \mathbb{R} \) if and only if \( I_n(u/k_n^{1/2}) \leq 1 \), if and only if

\[
I_n(u/k_n^{1/2}) - E[I_n(u/k_n^{1/2})] \leq 1 - \frac{n}{k_n} P \left( y_t > l_ne^{u/k_n^{1/2}} \right) = 1 - \frac{P \left( y_t > l_ne^{u/k_n^{1/2}} \right)}{P(y_t > l_n)}.
\]

Expand \( P(y_t > l_ne^{u/k_n^{1/2}}) \) around \( u = 0 \). Use \( l_n \sim \varpi^{-1/\delta}(\ln(n))^{1/\delta} \) and \( (\partial/\partial c)P(y_t > c) = -\varpi \vartheta \delta c^{\delta-1} \exp\{-\varpi c\} \), and the mean-value-theorem, to deduce \( k_n^{1/2} \ln(y_{(k_n)}/l_n) \leq u \) if and only if for some \( |u^*| \leq |u| \)

\[
k_n^{1/2} \left( I_n(u/k_n^{1/2}) - E[I_n(u/k_n^{1/2})] \right) \leq -\frac{1}{P(|y_t| > l_n)} \frac{\partial}{\partial c} P(y_t > c) \bigg|_{c=l_ne^{u*/k_n^{1/2}}} \times l_ne^{u*/k_n^{1/2}} \times u = \delta \vartheta \varpi \frac{n}{k_n} \exp \left\{ -\varpi \left( l_ne^{u*/k_n^{1/2}} \right)^{\delta-1} \times l_ne^{u*/k_n^{1/2}} \right\} \frac{\delta \vartheta \ln(n)}{k_n} u \times (1 + o(1)).
\]

Therefore \( k_n^{1/2} \ln(y_{(k_n)}/l_n) \) and \( (k_n^{3/2}/(\delta \vartheta \ln(n))) \times (I_n(u/k_n^{1/2}) - E[I_n(u/k_n^{1/2})]) \) have the same limit distribution. But in view of the Assumption D mixing property and the zero mean \( L_2 \)-boundedness of \( (n/k_n)^{1/2} \{I(y_t > l_ne^u) - E[I(y_t > l_ne^u)]\} \) we have by Theorem 1.6 and Lemma 2.1 in McLeish (1975)

\[
k_n^{1/2} \left( I_n(u/k_n^{1/2}) - E[I_n(u/k_n^{1/2})] \right) = \frac{1}{n^{1/2}} \sum_{t=1}^{n} \left( \frac{n}{k_n} \right)^{1/2} \left\{ I \left( y_t > l_ne^{u/k_n^{1/2}} \right) - E \left[ I \left( y_t > l_ne^{u/k_n^{1/2}} \right) \right] \right\} = O_p(1).
\]

By the supposition \( k_n = O(\ln(n)) \) it therefore follows

\[
\frac{k_n^{3/2}}{\ln(n)} \left( I_n(u/k_n^{1/2}) - E[I_n(u/k_n^{1/2})] \right) = \frac{k_n}{\ln(n)} k_n^{1/2} \left( I_n(u/k_n^{1/2}) - E[I_n(u/k_n^{1/2})] \right) = O_p(1).
\]

Therefore \( k_n^{1/2} \ln(y_{(k_n)}/l_n) = O_p(1) \).

Step 3 \( (\bar{c}_n) \). Let \( \{\bar{c}_n\} \) be the sequence of positive numbers that satisfies \( P(y_t > \bar{c}_n) = m_n/n \).
We have

\[ \hat{\kappa}_{m_n}^{-1} = \frac{1}{m_n} \sum_{j=1}^{m_n} \ln \left( \frac{y(j)}{y(m_n+1)} \right) \]

(17)

\[ = \frac{1}{m_n} \sum_{i=1}^{n} \ln \left( \frac{y(i)}{y(m_n+1)} \right) \times I \left( y_i > y(m_n+1) \right) \]

\[ = \frac{1}{m_n} \sum_{i=1}^{n} \ln \left( \frac{y(i)}{\hat{c}_n} \right) \times I \left( y_i > \hat{c}_n \right) + o_p(1) \]

\[ = \frac{1}{m_n} \sum_{i=1}^{n} \ln \left( \frac{y(i)}{\hat{c}_n} \right) \times I \left( y_i > \hat{c}_n \right) + o_p(1) = \frac{1}{n} \sum_{i=1}^{n} x_{n,i} + o_p(1). \]

say. The third equality follows from \( y_{(m_n+1)}/\hat{c}_n = 1 + O_p(1/m_n^{1/2}) \) by an application of Step 2 since

\[ \frac{1}{m_n} \sum_{i=1}^{n} \ln \left( \frac{y(i)}{y(m_n+1)} \right) \times I \left( y_i > y(m_n+1) \right) - \frac{1}{m_n} \sum_{i=1}^{n} \ln \left( \frac{y(i)}{\hat{c}_n} \right) \times I \left( y_i > y(m_n+1) \right) \]

\[ = (\ln (y(m_n+1)) - \ln (\hat{c}_n)) \times \frac{1}{m_n} \sum_{i=1}^{n} I \left( y_i > y(m_n+1) \right) \]

\[ = O_p \left( \frac{1}{m_n^2} \frac{1}{m_n} \sum_{i=1}^{n} I \left( y_i > y(m_n+1) \right) \right) = O_p \left( 1/m_n^{1/2} \right). \]

Notice this exploits the construction of \( y_{(m_n+1)} \) and distribution continuity: \( 1/m_n \sum_{i=1}^{n} I(y_i > y_{(m_n+1)}) = 1 \) a.s. The fourth equality in (17) can be verified under very general conditions in view of the stationary mixing property and the fact that \( y_{(m_n+1)} \) is the sample \( m_n/n \) tail quantile. The method of proof is identical to Lemmas A.2 and A.3 in Hill (2013).

In order to prove \( \hat{\kappa}_{m_n}^{-1} \overset{p}{\to} 0 \) it remains to verify \( 1/n \sum_{i=1}^{n} x_{n,i} \overset{p}{\to} 0 \). The variable \( x_{n,i} = (n/m_n) \ln(y_i/\hat{c}_n)I(y_i > \hat{c}_n) \) has a positive finite mean that decays to zero. This follows by invoking exponential tail bound (9) and \( P(y_i > \hat{c}_n) = m_n/n \) to obtain

\[ E[x_{n,i}] = \frac{n}{m_n - 1} \int_{0}^{\infty} P(\ln(y_i/\hat{c}_n) > u) \, du = \frac{n}{m_n - 1} \int_{0}^{\infty} P(y_i > \hat{c}_n e^u) \, du \]

\[ = \frac{n}{m_n - 1} \int_{0}^{\infty} P(y_i > \hat{c}_n) \int_{0}^{\infty} P(y_i > \hat{c}_n e^u) \, du \]

\[ = \frac{m_n}{m_n - 1} \int_{0}^{\infty} \exp \left\{ -\varpi \hat{c}_n \left( e^{\delta u} - 1 \right) \right\} \, du > 0. \]

Seeing that \( \varpi > 0 \), and (16) implies \( \hat{c}_n \to \infty \), it follows by dominated convergence \( E[x_{n,i}] \downarrow 0 \). Hence the Cesáro sum \( 1/n \sum_{i=1}^{n} E[x_{n,i}] \downarrow 0 \). Now use Markov’s inequality and \( 1/n \sum_{i=1}^{n} E[x_{n,i}] \)
\( \sum_{i=1}^{n} x_{n,t} \xrightarrow{p} 0. \) \( \text{QED} \).

**PROOF OF THEOREM 3.4.** We have \(-y_{(k_n)}^{(\cdot)}/l_n = 1 + O_p(1/k_n^{1/2})\) by Lemma B.3, while Hill (2012b: Lemma A.9) proves \( \mathcal{S}_n^2/\mathcal{S}_n^2 \xrightarrow{p} 1 \) under conditions that cover Assumptions D, K, N and T'. Further, the proofs of tail array HAC consistency Theorem 3 in Hill (2010) and tail-trimmed HAC consistency Lemma A.9 in Hill (2012b) can be easily extended to prove \( ||\Sigma_n\Sigma_n^{-1} - I_2|| \xrightarrow{p} 0. \) The claim \( \mathcal{V}_n^2/\mathcal{V}_n^2 \xrightarrow{p} 1 \) now follows from the Slutsky Theorems. \( \text{QED} \).

**APPENDIX B: Proofs of Lemmas A.1-A.5**

**PROOF OF LEMMA A.1.**

**Claim (a):** We simplify notation by proving \( E(1/n^{1/2}\sum_{t=1}^{n} y_{n,t} - E[y_{n,t}])^2 = r_n E[y_{n,t}^2] \) for the two-tailed intermediate order trimmed \( y_{n,t} := y_t I(|y_t| \leq l_n) \). Simply replace \( y_t I(|y_t| \leq l_n) \) with \( y_t(-l_n \leq y_t \leq q_n) \) to prove the claim.

Note \( \mathcal{S}_n \sim E[y_{n,t}^2] + 2\sum_{i=1}^{n-1}(1-i/n)E[y_{n,1y_{n,i+1}}] \). If \( E[y_{n,t}^2] < \infty \) then \( \mathcal{S}_n \sim K \) by the Assumption D geometric \( \alpha \)-mixing property (Ibragimov 1962) hence the claim \( \mathcal{S}_n = r_n E[y_{n,t}^2] \) holds for some \( r_n = O(1) \). Now assume \( E[y_{n,t}^2] = \infty \) such that the tail index \( \kappa \in (1,2] \).

Define quantile functions \( Q_n(u) = \inf\{ a \geq 0 : P(|y_{n,t}| > a) \leq u \} \) and \( Q(u) = \inf\{ a \geq 0 : P(|y_t| > a) \leq u \} \) for \( u \in [0,1] \), and recall under geometric \( \alpha \)-mixing the mixing coefficients satisfy \( \alpha_h \leq K\rho^h \) for \( \rho \in (0,1) \). By Theorem 1.1 of Rio (1993)

\[
\sum_{i=1}^{n-1} |E[y_{n,1y_{n,i+1}}]| \leq 2\sum_{i=1}^{n-1} \int_{0}^{\alpha_h} Q_n^2(u)du \leq 2\sum_{i=1}^{n-1} \int_{0}^{\rho^h} Q_n^2(u)du.
\]

By tail-trimming and distribution continuity \( P(y_{n,t} = 0) = k_n/n \). Hence by the threshold construction (2) we have \( Q_n(u) = 0 \) for \( u \in [0,k_n/n] \) and \( Q_n(u) = Q(u) \) for \( u \in (k_n/n,1] \), and by the Assumption T power law property \( Q(u) = Ku^{-1/\kappa}(1 + o(1)) \) as \( u \to 0 \). Now exploit \( Q(u) \leq K u^{-1/\kappa} \) and dominated convergence to deduce

\[
\sum_{i=1}^{n-1} |E[y_{n,1y_{n,i+1}}]| \leq K \sum_{i=1}^{n-1} \int_{k_n/n}^{\rho^h} u^{-2/\kappa}du = K \sum_{i=1}^{n-1} \max\left\{ 0, \left(\frac{n}{k_n}\right)^{2/\kappa-1} - \rho^{-i(2/\kappa-1)} \right\}
\]

\[
= K \sum_{i=1}^{\mathcal{A}\ln(n/k_n)} \left\{ \left(\frac{n}{k_n}\right)^{2/\kappa-1} - \rho^{-i(2/\kappa-1)} \right\},
\]

where \( \mathcal{A} := (2/\kappa - 1)/\ln(\rho) \). The final equality follows from \( (n/k_n)^{2/\kappa-1} < \rho^{-i(2/\kappa-1)} \) when \( i > \mathcal{A}\ln(n/k_n) \). Further \( \sum_{i=1}^{\mathcal{A}\ln(n/k_n)} \left\{ (n/k_n)^{2/\kappa-1} - \rho^{-i(2/\kappa-1)} \right\} = K \ln(n/k_n) \times (n/k_n)^{2/\kappa-1}(1 +
Claim (b): Under Assumption T and (2) the thresholds are by construction \( l_n = d^{1/\kappa}(n/k_n)^{1/\kappa} \). The claims follow from power law decay Assumption T and Karamata’s Theorem (Resnick 1987: Theorem 0.6): if \( \kappa = 2 \) then \( E[y_{n,t}^2] \sim K \ln(n) \), and if \( \kappa \in (1, 2) \) then \( E[y_{n,t}^2] \sim \kappa(2 - \kappa)^{-1/2}(n/k_n) = \kappa(2 - \kappa)^{-1}d^{2/\kappa}(n/k_n)^{2/\kappa-1} \).

Claim (c): Observe \( \{I(y_t < -l_n)/k_n^{1/2}, \mathbb{F}_t\} \) forms a geometric \( L_2 \)-mixingale array with mixingale constants \( Kn^{-1/2} \) under geometric \( \alpha \)-mixing Assumption D (Hill 2010: Lemma 2). Therefore \( E(1/n^{1/2} \sum_{t=1}^n I_{n,t})^2 \leq K \sum_{t=1}^n (Kn^{-1/2})^2 \leq K \) by Theorem 1.6 in McLeish (1975). Since \( E[T_{n,t}^2] \) = 1 + \( o(1) \) by construction we have shown \( E(1/n^{1/2} \sum_{t=1}^n I_{n,t})^2 \leq K E[T_{n,t}^2] \). QED.

PROOF OF LEMMA A.2. The indicator function \( I(u) := I(u \leq 0) \) can be approximated by a smooth regular sequence \( \{J_n(u)\}_{n \geq 1} \), cf. Lighthill (1958). Let \( \{N_n\} \) be a sequence of finite positive numbers, \( N_n \rightarrow \infty \), the rate to be chosen below. Now define \( J_n(u) := \int_{-\infty}^u I(\tau)S(N_n(\tau - u))N_n e^{-\tau^2/2N_n^2}d\tau \), where \( S(\xi) = e^{-1/(1-\xi^2)}/\int_{-1}^1 e^{-1/(1-w^2)}dw \) if \( |\xi| < 1 \) and \( S(\xi) = 0 \) if \( |\xi| \geq 1 \). The function \( S(N_n(\tau - u)) \) blots out \( I(\tau) \) when \( \tau \) is outside the open interval \( (u - 1/N_n, u + 1/N_n) \). The function \( I(u) \) is differentiable with derivative \( \delta(u) \) the Dirac delta function. Note \( \delta(u) \) has a regular sequence approximation \( \mathcal{D}_n(u) := (N_n/\pi)^{1/2}\exp\{-N_nu^2\} \). See Lighthill (1958: p. 22).

Define \( Y_t(a) := y_t - a \). We have

\[
\frac{1}{n^{1/2}S_n} \sum_{t=1}^n y_t I_{n,t} \left\{ I(y_t \leq y_{t,(\gamma,n)}) \right\} - I(y_t \leq q_n) \right\}
= \frac{1}{n^{1/2}S_n} \sum_{t=1}^n y_t I_{n,t} \left\{ J_n(Y_t(y_{t,(\gamma,n)})) - J_n(y_{t,(\gamma,n)}) \right\}
+ \frac{1}{n^{1/2}S_n} \sum_{t=1}^n y_t I_{n,t} \left\{ I(Y_t(y_{t,(\gamma,n)})) - J_n(Y_t(y_{t,(\gamma,n)})) \right\}
- \frac{1}{n^{1/2}S_n} \sum_{t=1}^n y_t I_{n,t} \left\{ I(Y_t(q_n)) - J_n(Y_t(q_n)) \right\}.
\]
The second and third terms can be forced to be as close to zero as we like for each \(n\) by choosing \(N_n \to \infty\) sufficiently fast, hence they are \(o_p(1)\). Cf. Phillips (1995).

The first term on the right-hand-side can be expanded by the mean-value-theorem. Use \(y_{([\alpha n])} = q_\alpha + O_p(1/n^{1/2})\) as in Čížek (2008: proof of Lemma A.3), and note for some \(q_n^*\), \(|y_{([\alpha n])} - q_\alpha| \leq |y_{([\alpha n])}| + q_\alpha|\),

\[
A_n := \left| \frac{1}{n^{1/2}S_n} \sum_{t=1}^{n} y_t I_{n,t} \{J_n(Y_t(y_{([\alpha n])})) - J_n(Y_t(q_\alpha))\} \right|
\]

\[
= \left| \frac{1}{n^{1/2}S_n} \sum_{t=1}^{n} y_t I_{n,t} \mathcal{D}_n(Y_t(q_\alpha^*)) \times (y_{([\alpha n])} + q_\alpha) \right|
\]

\[
= O_p \left( \frac{E|y_t I_{n,t}|}{S_n} \frac{1}{n} \sum_{t=1}^{n} \frac{|y_t I_{n,t}|}{E|y_t I_{n,t}|} \mathcal{D}_n(Y_t(q_\alpha^*)) \right).
\]

By Lyapunov's inequality and variance bound Lemma A.1.a \(E|y_t I_{n,t}|/S_n \leq (E|y_t I_{n,t}|)^{1/2}/S_n = O(1)\); by stationarity and ergodicity of \(y_t\) it follows \(1/n \sum_{t=1}^{n} |y_t I_{n,t}|/E|y_t I_{n,t}| \to 1\) since \(|y_t I_{n,t}|/|y_t I_{n,t}|\) is integrable; and since by distribution continuity \(Y_t(q_\alpha^*) = y_t - q_\alpha^* \neq 0\ a.s.\) it follows \(\max_{1 \leq t \leq n} \{\mathcal{D}_n(Y_t(q_\alpha^*))\} \to 0\) as fast as we choose by setting \(N_n \to \infty\) sufficiently fast. Therefore \(A_n = o_p(1)\), which completes the proof. \(\Box\).

**PROOF OF Lemma A.4.** Define \(I_{n,t}(u) := (n/k_n)^{1/2} \{I(y_t < -l_ne^{u}) - P(y_t < -l_ne^{u})\}\) for \(u \in \mathbb{R}\), hence by the notation of Appendix A \(I_{n,t} = I_{n,t}(0)\). The dependence and tail conditions of Lemma 3 in Hill (2010) are satisfied under Assumptions D and T, hence \(n^{-1/2} \sum_{t=1}^{n} I_{n,t}(u) \to N(0,v^2(u))\) where \(v^2(u) > 0\) for each \(u\). This implies \(n^{-1/2} \sum_{t=1}^{n} I_{n,t}(0) = O_p(1)\) as claimed. Further, it implies by dominated convergence \(n^{-1/2} \sum_{t=1}^{n} I_{n,t}(u/k_n^{1/2}) \to N(0,v^2(0))\) for each \(u\) hence by arguments in Hsing (1991: p. 1553) \(k_n^{1/2} (y_{(k_n)} /(-l_n) - 1) = k_n^{-1/2} \sum_{t=1}^{n} I_{n,t}(u/k_n^{1/2}) \times (1 + o_p(1))\) for each \(u\). Pick \(u = 0\) to deduce as claimed \(k_n^{1/2} (y_{(k_n)} /(-l_n) - 1) = k_n^{-1/2} \sum_{t=1}^{n} I_{n,t}(0) \times (1 + o_p(1))\). \(\Box\).

**PROOF OF Lemma A.5.** We first require some notation. Recall \(I_{n,t} = (n/k_n)^{1/2} \{I(y_t < -l_n) - P(y_t < -l_n)\}\) and \(W_{n,t} := [y_{n,t}^* - E[y_{n,t}^*], I_{n,t}]\), and define \(Z_{n,t}^* := \lambda W_{n,t}\) and \(\mathcal{G}_n := E(\sum_{j=1}^{n} Z_{n,t}^*)^2\) where we hide dependence on \(\lambda\). Let \(S\) and \(T\) be families of integers, and define the sigma-field \(\mathcal{G}_n, T := \sigma(Z_{n,t}^*/\mathcal{G}_n : t \in T)\). Let \(L_2(\mathcal{A})\) denote the space of zero mean and finite variance \(\mathcal{A}\)-measurable random variables. Define \(\rho_k := \sup_{n \geq 1} \sup_{\text{dist}(T-S) \geq k} \max_{f \in L_2(\mathcal{G}_n, T), g \in L_2(\mathcal{G}_n, S)} |E(y_{n,t}^{*\prime})| (\text{see Bradley} 1996\text{ for references on the latticed or interlaced maximal correlation coefficient})\).

We must show \(\mathcal{G}_n^{-1} \sum_{t=1}^{n} Z_{n,t}^* \to N(0,1)\). It suffices to verify the conditions of Theorem 2.1 in Peligrad (1996). By \(\mathcal{G}_t\)-measurability, Assumptions D and \(N\) and variance properties Lemma A.1, it follows \(Z_{n,t}^*/\mathcal{G}_n\) is geometrically \(\alpha\)-mixing, stationary over \(1 \leq t \leq n\), and uniformly \(L_2\)-bounded
\[
\sup_{n \geq 1} E(Z_{n,t}^*/\mathcal{G}_n)^2 \leq K. \]
It remains to verify \((i)\) \(\lim_{k \to \infty} \tilde{\rho}_k^* < 1\), \((ii)\) \(\sup_{n \geq 1} \mathcal{G}_n^{-2} \sum_{t=1}^n E[Z_{n,t}^2] \leq K\), \((iii)\) \(\mathcal{G}_n^{-2} \sum_{t=1}^n E[Z_{n,t}^2 I(Z_{n,t}^* > \varepsilon \mathcal{G}_n)] \to 0 \forall \varepsilon > 0\).

Observe \(\tilde{\rho}_k^*\) is defined over finite variance \(\sigma(\cup_{n \geq 1} \cup_T \mathcal{G}_{n,t})\)-measurable random variables, where \(\sigma(\cup_{n \geq 1} \cup_T \mathcal{G}_{n,t}) \subseteq \sigma(\cup_t \sigma(y_t : \tau \leq t))\) while \(y_t\) is stationary and geometrically \(\alpha\)-mixing. Therefore \(\tilde{\rho}_k^* = O(\rho^k)\) for some \(\rho \in (0, 1)\) can be verified as in Theorem 1 of Bradley (1993). Since \(\lim_{k \to \infty} \tilde{\rho}_k^* = 0\) we have shown \((i)\). Assumption N’ and stationarity ensure \(nE[Z_{n,t}^*]/\mathcal{G}_n^2 = O(1)\) hence \((ii)\).

Finally, consider the Lindeberg condition \((iii)\) and assume \(E[y_{n,t}^*] = 0\) to simplify notation. In view of stationarity we need only prove the well known sufficient Lyapunov condition (e.g. Davidson 1994: Theorem 23.11)

\[
\frac{n}{\mathcal{G}_n^{2+\delta}} E[Z_{n,t}^*]^{2+\delta} \to 0 \text{ for some } \delta > 0.
\] (18)

By variance properties Lemma A.1 and the fact that \(E[Z_{n,t}^2] = 1 + o(1)\) by construction, it is straightforward to verify \(\mathcal{G}_n^2 = E(\sum_{t=1}^n Z_{n,t}^*)^2 = E(\sum_{t=1}^n y_{n,t}^2) \times (1 + O(1)).\) See the proof of Theorem 3.1. Further, by the threshold construction \(E[I(y_t < -l_n)] = k_n/n\) and Minkowski’s inequality it follows

\[
E[Z_{n,t}^*]^{2+\delta} \leq K \left( E[y_{n,t}^*]^{2+\delta} \right)^{1/(2+\delta)} + \left( \frac{n}{k_n} \right)^{1/(2+\delta)} \left( \frac{k_n}{n} \right)^{1/(2+\delta)} \right)^{2+\delta}
\]
for some finite \(K > 0\). Therefore since \(nE[y_{n,t}^2]/E(\sum_{t=1}^n y_{n,t}^2) = O(1)\) by Assumption N’, we have

\[
\frac{n}{\mathcal{G}_n^{2+\delta}} E[Z_{n,t}^*]^{2+\delta} \leq K \frac{1}{n^{\delta/2}(E[y_{n,t}^2])^{1+\delta/2}} \left( E[y_{n,t}^*]^{2+\delta} \right)^{1/(2+\delta)} + \left( \frac{n}{k_n} \right)^{1/2} \left( \frac{k_n}{n} \right)^{1/(2+\delta)} \right)^{2+\delta}
= : \mathcal{A}_n.
\]

Write \((z)_+ := \max(0, z)\). Under power law Assumption T and the threshold construction \(P(y_t < -l_n) = (k_n/n)\) it follows \(l_n = K(n/k_n)^{1/\kappa}\). If \(\kappa \in (0, 2)\) then by Karamata’s Theorem \(E[y_{n,t}^*]^{2+\tau}\)
\[ K^{l_{n}+r}P(|y_t| > l_n) \] for any \( r \geq 0 \), hence

\[
A_n \leq K \frac{1}{n^{\delta/2} (l_n^2 (k_n/n))^{1+\delta/2}} \left( l_n^{2+\delta} (k_n/n)^{1/(2+\delta)} \right) + \left( \frac{n}{k_n} \right)^{1/2} (k_n/n))^{1/(2+\delta)} \right)^{2+\delta}
\]

\[
= K \frac{1}{l_n^{2+\delta} k_n^{\delta/2}} \left( l_n + \left( \frac{n}{k_n} \right)^{1/2} \right)^{2+\delta}
= K \frac{1}{k_n^{\delta/2}} \left( 1 + \left( \frac{k_n}{n} \right)^{1/(\kappa-1/2)} \right)^{2+\delta}
= O(k_n^{-\delta/2}) = o(1)
\]

since \( 1/\kappa > 1/2 \).

If \( \kappa \in [2, 2 + \delta) \) then \( E[y_{n,t}^*]^2 \sim K l_{n}^{2+\delta} (k_n/n) \) and \( \lim \inf_{n \to \infty} E[y_{n,t}^2] > 0 \), hence

\[
A_n \leq K \frac{k_n}{n^{\delta/2}} \left( l_n + \left( \frac{n}{k_n} \right)^{1/2} \right)^{2+\delta}
\]

\[
= K \frac{1}{n^{\delta/2}} \left( \frac{n}{k_n} \right)^{\delta/2} \left( \left( \frac{k_n}{n} \right)^{1/2-1/\kappa} + 1 \right)^{2+\delta} = O(k_n^{-\delta/2}) = o(1),
\]

because \( 1/2 \geq 1/\kappa \). If \( \kappa = 2 + \delta \) then \( E[y_{n,t}^*]^2 = K \ln(n) \) by Karamata’s Theorem hence

\[
A_n \leq K \frac{1}{n^{\delta/2}} \left( \ln(n) + \left( \frac{n}{k_n} \right)^{1/2} \left( \frac{k_n}{n} \right)^{1/(2+\delta)} \right)^{2+\delta}
\]

\[
= K \frac{k_n}{n^{\delta/2}} \left( \left( \frac{n}{k_n} \right)^{1/(2+\delta)} \ln(n) + \left( \frac{n}{k_n} \right)^{1/2} \right)^{2+\delta}
\]

\[
\leq K \frac{k_n}{n^{\delta/2}} \left( \left( \frac{n}{k_n} \right)^{1/2} \right)^{2+\delta} = K \frac{1}{k_n^{\delta/2}} = o(1).
\]

Finally, if \( \kappa > 2 + \delta \) then by the same argument \( A_n \sim K/k_n^{\delta/2} \to 0 \). This verifies (18) which completes the proof. \( \Box \).

**References**


FIGURE 1: Daily Returns and Hill-Plots


FIGURE 4: Rolling Window Hill-Plot

Notes: $m_n$ is the number of tail observations.
### TABLE 1.a: \( \alpha = .05 \) and \( n = 100 \)

#### Untrimmed \( \widetilde{ES}_{n} \)

<table>
<thead>
<tr>
<th>Model</th>
<th>( \epsilon_t )</th>
<th>( \kappa_y )</th>
<th>Bias</th>
<th>RMSE</th>
<th>KS</th>
<th>AD</th>
<th>Null</th>
<th>Alt1</th>
<th>Alt2</th>
</tr>
</thead>
<tbody>
<tr>
<td>AR</td>
<td>( P_{1.5} )</td>
<td>1.50</td>
<td>.333</td>
<td>.407</td>
<td>3.31</td>
<td>2.88</td>
<td>.000, .036, .228, .053, .104, .136, .168, .304, .402</td>
<td></td>
<td></td>
</tr>
<tr>
<td>AR</td>
<td>( P_{2.5} )</td>
<td>2.50</td>
<td>.009</td>
<td>.046</td>
<td>.694</td>
<td>2.43</td>
<td>.000, .035, .111, .155, .296, .406, .688, .858, .937</td>
<td></td>
<td></td>
</tr>
<tr>
<td>AR</td>
<td>( N_{0.1} )</td>
<td>( \infty )</td>
<td>.004</td>
<td>.015</td>
<td>.850</td>
<td>2.25</td>
<td>.015, .050, .995, .991</td>
<td>1.00, 1.00, 1.00</td>
<td></td>
</tr>
<tr>
<td>GARCH</td>
<td>( P_{2.5} )</td>
<td>1.50</td>
<td>.017</td>
<td>.065</td>
<td>1.27</td>
<td>4.89</td>
<td>.024, .064, .102, .266, .512, .659, .918, .970, .982</td>
<td></td>
<td></td>
</tr>
<tr>
<td>IGARCH</td>
<td>( N_{0.1} )</td>
<td>2.00</td>
<td>.055</td>
<td>.132</td>
<td>2.47</td>
<td>1.79</td>
<td>.008, .030, .106, .204, .358, .471, .991, 1.00, 1.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>RiskMetric</td>
<td>( N_{0.1} )</td>
<td>2.00</td>
<td>.025</td>
<td>.098</td>
<td>1.87</td>
<td>3.07</td>
<td>.008, .067, .155, .995, 1.00, 1.00, 1.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>GARCH</td>
<td>( N_{0.1} )</td>
<td>4.10</td>
<td>.014</td>
<td>.038</td>
<td>1.67</td>
<td>2.16</td>
<td>.003, .048, .123, .995, 1.00, 1.00, 1.00</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

#### Trimmed \( \widetilde{ES}_{n,a}^{(s)} \)

<table>
<thead>
<tr>
<th>Model</th>
<th>( \epsilon_t )</th>
<th>( \kappa_y )</th>
<th>Bias</th>
<th>RMSE</th>
<th>KS</th>
<th>AD</th>
<th>Null</th>
<th>Alt1</th>
<th>Alt2</th>
</tr>
</thead>
<tbody>
<tr>
<td>AR</td>
<td>( P_{1.5} )</td>
<td>1.50</td>
<td>.560</td>
<td>.250</td>
<td>2.95</td>
<td>2.85</td>
<td>.480, .703, .779, .017, .054, .083, .206, .368, .481</td>
<td></td>
<td></td>
</tr>
<tr>
<td>AR</td>
<td>( P_{2.5} )</td>
<td>2.50</td>
<td>-.025</td>
<td>.063</td>
<td>.821</td>
<td>2.29</td>
<td>.216, .409, .526, .045, .157, .213, .764, .924, .960</td>
<td></td>
<td></td>
</tr>
<tr>
<td>AR</td>
<td>( N_{0.1} )</td>
<td>( \infty )</td>
<td>.035</td>
<td>.011</td>
<td>.777</td>
<td>2.38</td>
<td>.718, .891, .956, .597, .800, .881, 1.00, 1.00, 1.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>GARCH</td>
<td>( P_{2.5} )</td>
<td>1.50</td>
<td>.081</td>
<td>.055</td>
<td>1.32</td>
<td>5.85</td>
<td>.151, .294, .392, .046, .205, .330, .848, .937, .965</td>
<td></td>
<td></td>
</tr>
<tr>
<td>IGARCH</td>
<td>( N_{0.1} )</td>
<td>2.00</td>
<td>.143</td>
<td>.105</td>
<td>2.37</td>
<td>1.86</td>
<td>.057, .320, .481, .148, .359, .526, 1.00, 1.00, 1.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>RiskMetric</td>
<td>( N_{0.1} )</td>
<td>2.00</td>
<td>.165</td>
<td>.083</td>
<td>2.43</td>
<td>3.64</td>
<td>.294, .514, .624, .945, .990, 1.00, 1.00, 1.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>GARCH</td>
<td>( N_{0.1} )</td>
<td>4.10</td>
<td>.067</td>
<td>.021</td>
<td>1.45</td>
<td>2.14</td>
<td>.145, .356, .478, 1.00, 1.00, 1.00, 1.00, 1.00, 1.00</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

#### Optimally Trimmed \( \widetilde{ES}_{n,a}^{(2)} = \widetilde{ES}_{n,a}^{(s)} + \tilde{\alpha}_{n,a}^{(2)} \)

<table>
<thead>
<tr>
<th>Model</th>
<th>( \epsilon_t )</th>
<th>( \kappa_y )</th>
<th>Bias</th>
<th>RMSE</th>
<th>KS</th>
<th>AD</th>
<th>Null</th>
<th>Alt1</th>
<th>Alt2</th>
</tr>
</thead>
<tbody>
<tr>
<td>AR</td>
<td>( P_{1.5} )</td>
<td>1.50</td>
<td>.088</td>
<td>.380</td>
<td>1.69</td>
<td>2.45</td>
<td>.021, .057, .084, .076, .126, .160, .158, .268, .364</td>
<td></td>
<td></td>
</tr>
<tr>
<td>AR</td>
<td>( P_{2.5} )</td>
<td>2.50</td>
<td>-.025</td>
<td>.063</td>
<td>.821</td>
<td>2.29</td>
<td>.020, .062, .122, .223, .394, .518, .640, .835, .920</td>
<td></td>
<td></td>
</tr>
<tr>
<td>AR</td>
<td>( N_{0.1} )</td>
<td>( \infty )</td>
<td>-.003</td>
<td>.017</td>
<td>.892</td>
<td>2.21</td>
<td>.014, .054, .103, .994, 1.00, 1.00, 1.00, 1.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>GARCH</td>
<td>( P_{2.5} )</td>
<td>1.50</td>
<td>-.021</td>
<td>.082</td>
<td>.852</td>
<td>3.88</td>
<td>.021, .065, .112, .539, .769, .832, .904, .960, .978</td>
<td></td>
<td></td>
</tr>
<tr>
<td>IGARCH</td>
<td>( N_{0.1} )</td>
<td>2.00</td>
<td>-.080</td>
<td>.211</td>
<td>1.75</td>
<td>1.83</td>
<td>.021, .064, .126, .334, .592, .772, .926, 1.00, 1.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>RiskMetric</td>
<td>( N_{0.1} )</td>
<td>2.00</td>
<td>.031</td>
<td>.110</td>
<td>1.48</td>
<td>3.11</td>
<td>.017, .061, .121, 1.00, 1.00, 1.00, 1.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>GARCH</td>
<td>( N_{0.1} )</td>
<td>4.10</td>
<td>-.053</td>
<td>.065</td>
<td>1.03</td>
<td>2.05</td>
<td>.015, .064, .123, 1.00, 1.00, 1.00, 1.00</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

a. Root Mean Squared Error.
b. Kolmogorov-Smirnov statistic for a test of standard normality, divided by the 5% critical value.
c. Anderson-Darling statistic for a test of standard normality, divided by the 5% critical value.
d. The hypotheses are Null : \( ES = ES_{a} \); Alt1 : \( ES = ES_{a}/2 \); Alt2 : \( ES = 0 \).
e. The GARCH model is \( y_t = \sigma_t \epsilon_t \); \( \sigma_t^2 = \omega + \alpha y_{t-1}^2 + \beta \sigma_{t-1}^2 \). IGARCH \( \alpha, \beta = .4, .6 \); RiskMetric \( \alpha, \beta = .06, .94 \).
<table>
<thead>
<tr>
<th>Model</th>
<th>$\epsilon_t$</th>
<th>$\kappa_y$</th>
<th>Bias</th>
<th>RMSE(^a)</th>
<th>KS(^b)</th>
<th>AD(^c)</th>
<th>Null(^d)</th>
<th>Alt1</th>
<th>Alt2</th>
</tr>
</thead>
<tbody>
<tr>
<td>AR</td>
<td>$P_{1.5}$</td>
<td>1.50</td>
<td>.203</td>
<td>.301</td>
<td>2.21</td>
<td>2.61</td>
<td>.000,050,189</td>
<td>.908,185,244</td>
<td>.493,772,911</td>
</tr>
<tr>
<td>AR</td>
<td>$P_{2.5}$</td>
<td>2.50</td>
<td>.005</td>
<td>.030</td>
<td>.954</td>
<td>2.29</td>
<td>.023,039,072</td>
<td>.445,682,795</td>
<td>.976,988,994</td>
</tr>
<tr>
<td>AR</td>
<td>$N_{0.1}$</td>
<td>$\infty$</td>
<td>.002</td>
<td>.007</td>
<td>.831</td>
<td>1.63</td>
<td>.100,051,092</td>
<td>1.00,100,100</td>
<td>1.00,100,100</td>
</tr>
<tr>
<td>GARCH(^e)</td>
<td>$P_{2.5}$</td>
<td>1.50</td>
<td>.013</td>
<td>.056</td>
<td>1.29</td>
<td>3.89</td>
<td>.026,070,091</td>
<td>.379,638,743</td>
<td>.957,985,992</td>
</tr>
<tr>
<td>IGARCH</td>
<td>$N_{0.1}$</td>
<td>2.00</td>
<td>.082</td>
<td>.178</td>
<td>2.67</td>
<td>2.21</td>
<td>.008,038,109</td>
<td>.182,326,434</td>
<td>.989,100,100</td>
</tr>
<tr>
<td>RiskMetric</td>
<td>$N_{0.1}$</td>
<td>2.00</td>
<td>.081</td>
<td>.287</td>
<td>1.84</td>
<td>2.02</td>
<td>.004,030,076</td>
<td>.688,920,982</td>
<td>1.00,100,100</td>
</tr>
<tr>
<td>GARCH</td>
<td>$N_{0.1}$</td>
<td>4.10</td>
<td>.008</td>
<td>.026</td>
<td>1.01</td>
<td>1.47</td>
<td>.000,040,129</td>
<td>1.00,100,100</td>
<td>1.00,100,100</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Model</th>
<th>$\epsilon_t$</th>
<th>$\kappa_y$</th>
<th>Bias</th>
<th>RMSE(^a)</th>
<th>KS(^b)</th>
<th>AD(^c)</th>
<th>Null(^d)</th>
<th>Alt1</th>
<th>Alt2</th>
</tr>
</thead>
<tbody>
<tr>
<td>AR</td>
<td>$P_{1.5}$</td>
<td>1.50</td>
<td>.436</td>
<td>.172</td>
<td>1.85</td>
<td>2.51</td>
<td>.562,753,805</td>
<td>.043,108,150</td>
<td>.827,991,999</td>
</tr>
<tr>
<td>AR</td>
<td>$P_{2.5}$</td>
<td>2.50</td>
<td>.034</td>
<td>.023</td>
<td>1.188</td>
<td>2.59</td>
<td>.128,295,403</td>
<td>.232,481,630</td>
<td>.977,990,994</td>
</tr>
<tr>
<td>AR</td>
<td>$N_{0.1}$</td>
<td>$\infty$</td>
<td>.017</td>
<td>.006</td>
<td>1.018</td>
<td>1.67</td>
<td>.491,708,794</td>
<td>1.00,100,100</td>
<td>1.00,100,100</td>
</tr>
<tr>
<td>GARCH</td>
<td>$P_{2.5}$</td>
<td>1.50</td>
<td>.056</td>
<td>.050</td>
<td>1.76</td>
<td>4.91</td>
<td>.096,170,254</td>
<td>.132,397,535</td>
<td>.922,965,976</td>
</tr>
<tr>
<td>IGARCH</td>
<td>$N_{0.1}$</td>
<td>2.00</td>
<td>.158</td>
<td>.147</td>
<td>2.46</td>
<td>2.25</td>
<td>.004,182,368</td>
<td>.158,296,400</td>
<td>1.00,100,100</td>
</tr>
<tr>
<td>RiskMetric</td>
<td>$N_{0.1}$</td>
<td>2.00</td>
<td>.213</td>
<td>.263</td>
<td>2.04</td>
<td>2.02</td>
<td>.003,108,232</td>
<td>.500,857,947</td>
<td>1.00,100,100</td>
</tr>
<tr>
<td>GARCH</td>
<td>$N_{0.1}$</td>
<td>4.10</td>
<td>.035</td>
<td>.021</td>
<td>1.65</td>
<td>1.62</td>
<td>.200,422,529</td>
<td>1.00,100,100</td>
<td>1.00,100,100</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Model</th>
<th>$\epsilon_t$</th>
<th>$\kappa_y$</th>
<th>Bias</th>
<th>RMSE(^a)</th>
<th>KS(^b)</th>
<th>AD(^c)</th>
<th>Null(^d)</th>
<th>Alt1</th>
<th>Alt2</th>
</tr>
</thead>
<tbody>
<tr>
<td>AR</td>
<td>$P_{1.5}$</td>
<td>1.50</td>
<td>.120</td>
<td>.273</td>
<td>1.53</td>
<td>2.27</td>
<td>.000,038,130</td>
<td>.112,203,269</td>
<td>.527,827,946</td>
</tr>
<tr>
<td>AR</td>
<td>$P_{2.5}$</td>
<td>2.50</td>
<td>-.002</td>
<td>.031</td>
<td>.708</td>
<td>2.18</td>
<td>.016,039,088</td>
<td>.488,738,812</td>
<td>.975,988,994</td>
</tr>
<tr>
<td>AR</td>
<td>$N_{0.1}$</td>
<td>$\infty$</td>
<td>-.004</td>
<td>.008</td>
<td>.798</td>
<td>1.49</td>
<td>.004,105,087</td>
<td>1.00,100,100</td>
<td>1.00,100,100</td>
</tr>
<tr>
<td>GARCH</td>
<td>$P_{2.5}$</td>
<td>1.50</td>
<td>-.021</td>
<td>.069</td>
<td>.949</td>
<td>3.45</td>
<td>.003,064,124</td>
<td>.507,745,828</td>
<td>.955,981,990</td>
</tr>
<tr>
<td>IGARCH</td>
<td>$N_{0.1}$</td>
<td>2.00</td>
<td>-.027</td>
<td>.225</td>
<td>1.58</td>
<td>2.20</td>
<td>.019,064,131</td>
<td>.218,385,515</td>
<td>.929,100,100</td>
</tr>
<tr>
<td>RiskMetric</td>
<td>$N_{0.1}$</td>
<td>2.00</td>
<td>.047</td>
<td>.330</td>
<td>1.54</td>
<td>1.98</td>
<td>.021,064,117</td>
<td>.830,981,100</td>
<td>1.00,100,100</td>
</tr>
<tr>
<td>GARCH</td>
<td>$N_{0.1}$</td>
<td>4.10</td>
<td>-.012</td>
<td>.033</td>
<td>.932</td>
<td>1.51</td>
<td>.014,057,109</td>
<td>1.00,100,100</td>
<td>1.00,100,100</td>
</tr>
</tbody>
</table>

---

a. Root Mean Squared Error.
b. Kolmogorov-Smirnov statistic for a test of standard normality, divided by the 5% critical value.
c. Anderson-Darling statistic for a test of standard normality, divided by the 5% critical value.
d. The hypotheses are Null : ES = ES\(_0\); Alt1 : ES = ES\(_{0.2}\)/2; Alt2 : ES = 0.
e. The GARCH model is $y_t = \sigma_t \epsilon_t$, $\sigma_t^2 = \omega + \alpha \sigma_{t-1}^2 + \beta \sigma_{t-1}^2$. IGARCH $\alpha, \beta = .4, .6$; RiskMetric $\alpha, \beta = .06, .94$.  

\(^{a,b,c,d,e}\)
### TABLE 1.c : \( \alpha = .05 \) and \( n = 1000 \)

**Untrimmed \( \hat{ES}_n \)**

<table>
<thead>
<tr>
<th>Model</th>
<th>( \epsilon_t )</th>
<th>( \kappa_y )</th>
<th>Bias</th>
<th>RMSE(^a)</th>
<th>KS(^b)</th>
<th>AD(^c)</th>
<th>Null(^d)</th>
<th>Alt1</th>
<th>Alt2</th>
</tr>
</thead>
<tbody>
<tr>
<td>AR</td>
<td>( P_{1.5} )</td>
<td>1.50</td>
<td>.178</td>
<td>.252</td>
<td>2.05</td>
<td>2.59</td>
<td>.000,.082,.191</td>
<td>1.47,.152,.347</td>
<td>.842,.991,1.00</td>
</tr>
<tr>
<td>AR</td>
<td>( P_{2.5} )</td>
<td>2.50</td>
<td>.005</td>
<td>.024</td>
<td>1.17</td>
<td>2.43</td>
<td>.024,.052,.089</td>
<td>.677,.841,.891</td>
<td>.995,.999,.999</td>
</tr>
<tr>
<td>AR</td>
<td>( N_{0.1} )</td>
<td>( \infty )</td>
<td>.001</td>
<td>.005</td>
<td>.722</td>
<td>2.15</td>
<td>.016,.058,.093</td>
<td>1.00,1.00,1.00</td>
<td>1.00,1.00,1.00</td>
</tr>
<tr>
<td>GARCH</td>
<td>( P_{2.5} )</td>
<td>1.50</td>
<td>.010</td>
<td>.050</td>
<td>1.98</td>
<td>4.63</td>
<td>.027,.066,.095</td>
<td>.485,.724,.804</td>
<td>.973,.989,.995</td>
</tr>
<tr>
<td>IGARCH</td>
<td>( N_{0.1} )</td>
<td>2.00</td>
<td>.074</td>
<td>.146</td>
<td>2.23</td>
<td>2.03</td>
<td>.000,.051,.134</td>
<td>.320,.579,.721</td>
<td>1.00,1.00,1.00</td>
</tr>
<tr>
<td>RiskMetric</td>
<td>( N_{0.1} )</td>
<td>2.00</td>
<td>.143</td>
<td>.428</td>
<td>1.83</td>
<td>2.23</td>
<td>.000,.034,.091</td>
<td>.469,.695,.856</td>
<td>1.00,1.00,1.00</td>
</tr>
<tr>
<td>GARCH</td>
<td>( N_{0.1} )</td>
<td>4.10</td>
<td>.005</td>
<td>.020</td>
<td>.951</td>
<td>1.28</td>
<td>.004,.044,.116</td>
<td>1.00,1.00,1.00</td>
<td>1.00,1.00,1.00</td>
</tr>
</tbody>
</table>

**Trimmed \( \hat{ES}_{n,\alpha}^{(s)} \)**

<table>
<thead>
<tr>
<th>Model</th>
<th>( \epsilon_t )</th>
<th>( \kappa_y )</th>
<th>Bias</th>
<th>RMSE(^a)</th>
<th>KS</th>
<th>AD</th>
<th>Null(^d)</th>
<th>Alt1</th>
<th>Alt2</th>
</tr>
</thead>
<tbody>
<tr>
<td>AR</td>
<td>( P_{1.5} )</td>
<td>1.50</td>
<td>.360</td>
<td>.145</td>
<td>1.82</td>
<td>2.21</td>
<td>.518,.712,.801</td>
<td>.102,.211,.300</td>
<td>1.00,1.00,1.00</td>
</tr>
<tr>
<td>AR</td>
<td>( P_{2.5} )</td>
<td>2.50</td>
<td>-.002</td>
<td>.024</td>
<td>1.51</td>
<td>2.87</td>
<td>.085,.188,.271</td>
<td>.479,.717,.809</td>
<td>.995,.999,.999</td>
</tr>
<tr>
<td>AR</td>
<td>( N_{0.1} )</td>
<td>( \infty )</td>
<td>.009</td>
<td>.005</td>
<td>.768</td>
<td>2.22</td>
<td>.310,.515,.635</td>
<td>1.00,1.00,1.00</td>
<td>1.00,1.00,1.00</td>
</tr>
<tr>
<td>GARCH</td>
<td>( P_{2.5} )</td>
<td>1.50</td>
<td>.039</td>
<td>.048</td>
<td>1.95</td>
<td>4.91</td>
<td>.069,.130,.181</td>
<td>.268,.535,.664</td>
<td>.947,.979,.989</td>
</tr>
<tr>
<td>IGARCH</td>
<td>( N_{0.1} )</td>
<td>2.00</td>
<td>.126</td>
<td>.128</td>
<td>2.16</td>
<td>1.99</td>
<td>.017,.152,.302</td>
<td>.309,.548,.690</td>
<td>1.00,1.00,1.00</td>
</tr>
<tr>
<td>RiskMetric</td>
<td>( N_{0.1} )</td>
<td>2.00</td>
<td>.240</td>
<td>.406</td>
<td>1.88</td>
<td>2.40</td>
<td>.000,.049,.160</td>
<td>.436,.667,.808</td>
<td>1.00,1.00,1.00</td>
</tr>
<tr>
<td>GARCH</td>
<td>( N_{0.1} )</td>
<td>4.10</td>
<td>.021</td>
<td>.018</td>
<td>.847</td>
<td>1.39</td>
<td>.006,.054,.109</td>
<td>1.00,1.00,1.00</td>
<td>1.00,1.00,1.00</td>
</tr>
</tbody>
</table>

**Optimally Trimmed \( \hat{ES}_{n,\alpha}^{(2)} = \hat{ES}_{n,\alpha}^{(s)} + \hat{r}_{n,\alpha}^{(2)} \)**

<table>
<thead>
<tr>
<th>Model</th>
<th>( \epsilon_t )</th>
<th>( \kappa_y )</th>
<th>Bias</th>
<th>RMSE(^a)</th>
<th>KS</th>
<th>AD</th>
<th>Null(^d)</th>
<th>Alt1</th>
<th>Alt2</th>
</tr>
</thead>
<tbody>
<tr>
<td>AR</td>
<td>( P_{1.5} )</td>
<td>1.50</td>
<td>1.04</td>
<td>.251</td>
<td>1.33</td>
<td>1.65</td>
<td>.000,.060,.134</td>
<td>.166,.283,.398</td>
<td>.889,.997,1.00</td>
</tr>
<tr>
<td>AR</td>
<td>( P_{2.5} )</td>
<td>2.50</td>
<td>-.002</td>
<td>.025</td>
<td>1.05</td>
<td>2.40</td>
<td>.015,.041,.087</td>
<td>.714,.864,.906</td>
<td>.994,.999,.999</td>
</tr>
<tr>
<td>AR</td>
<td>( N_{0.1} )</td>
<td>( \infty )</td>
<td>-.003</td>
<td>.005</td>
<td>.678</td>
<td>1.98</td>
<td>.012,.063,.122</td>
<td>1.00,1.00,1.00</td>
<td>1.00,1.00,1.00</td>
</tr>
<tr>
<td>GARCH</td>
<td>( P_{2.5} )</td>
<td>1.50</td>
<td>-.022</td>
<td>.061</td>
<td>1.63</td>
<td>2.78</td>
<td>.009,.065,.130</td>
<td>.614,.790,.851</td>
<td>.968,.988,.992</td>
</tr>
<tr>
<td>IGARCH</td>
<td>( N_{0.1} )</td>
<td>2.00</td>
<td>-.027</td>
<td>.182</td>
<td>1.73</td>
<td>2.03</td>
<td>.020,.063,.120</td>
<td>.350,.618,.758</td>
<td>1.00,1.00,1.00</td>
</tr>
<tr>
<td>RiskMetric</td>
<td>( N_{0.1} )</td>
<td>2.00</td>
<td>-.046</td>
<td>.478</td>
<td>1.46</td>
<td>2.06</td>
<td>.018,.057,.094</td>
<td>.526,.775,.916</td>
<td>1.00,1.00,1.00</td>
</tr>
<tr>
<td>GARCH</td>
<td>( N_{0.1} )</td>
<td>4.10</td>
<td>-.001</td>
<td>.026</td>
<td>.784</td>
<td>1.28</td>
<td>.016,.058,.119</td>
<td>1.00,1.00,1.00</td>
<td>1.00,1.00,1.00</td>
</tr>
</tbody>
</table>

a. Root Mean Squared Error.
b. Kolmogorov-Smirnov statistic for a test of standard normality, divided by the 5% critical value.
c. Anderson-Darling statistic for a test of standard normality, divided by the 5% critical value.
d. The hypotheses are Null : ES = \( \hat{ES}_n \); Alt1 : ES = \( \hat{ES}_{n,\alpha}^{(s)} / 2 \); Alt2 : ES = 0.
e. The GARCH model is \( y_t = \sigma_t \epsilon_t, \sigma_t^2 = \omega + \alpha \sigma_{t-1}^2 + \beta \sigma_{t-1}^2, \) IGARCH \( \alpha, \beta = 4, .6 \); RiskMetric \( \alpha, \beta = .06, .94 \).
TABLE 1.d: $\alpha = .05$ and $n = 10,000$

### Untrimmed $\overline{ES}_n$

<table>
<thead>
<tr>
<th>Model</th>
<th>$\epsilon_t$</th>
<th>$\kappa_y$</th>
<th>Bias</th>
<th>RMSE$^a$</th>
<th>KS$^b$</th>
<th>AD$^c$</th>
<th>Null$^d$</th>
<th>Alt1</th>
<th>Alt2</th>
</tr>
</thead>
<tbody>
<tr>
<td>AR</td>
<td>$P_{1.5}$</td>
<td>1.50</td>
<td>.093</td>
<td>.139</td>
<td>1.95</td>
<td>2.29</td>
<td>.001,.0788,.186</td>
<td>.692,.910,.975</td>
<td>1.00,1.00,1.00</td>
</tr>
<tr>
<td>AR</td>
<td>$P_{2.5}$</td>
<td>2.50</td>
<td>.008</td>
<td>.014</td>
<td>2.68</td>
<td>2.11</td>
<td>.046,.092,.118</td>
<td>.959,.976,.983</td>
<td>1.00,1.00,1.00</td>
</tr>
<tr>
<td>AR</td>
<td>$N_{0.1}$</td>
<td>$\infty$</td>
<td>.000</td>
<td>.002</td>
<td>.985</td>
<td>1.64</td>
<td>.004,.025,.093</td>
<td>1.00,1.00,1.00</td>
<td>1.00,1.00,1.00</td>
</tr>
<tr>
<td>GARCH$^e$</td>
<td>$P_{2.5}$</td>
<td>1.50</td>
<td>.024</td>
<td>.032</td>
<td>2.26</td>
<td>8.21</td>
<td>.059,.099,.144</td>
<td>.815,.910,.932</td>
<td>1.00,1.00,1.00</td>
</tr>
<tr>
<td>IGARCH</td>
<td>$N_{0.1}$</td>
<td>2.00</td>
<td>-.014</td>
<td>.089</td>
<td>1.21</td>
<td>1.65</td>
<td>.019,.067,.098</td>
<td>1.00,1.00,1.00</td>
<td>1.00,1.00,1.00</td>
</tr>
<tr>
<td>RiskMetric</td>
<td>$N_{0.1}$</td>
<td>2.00</td>
<td>.096</td>
<td>.121</td>
<td>1.93</td>
<td>2.38</td>
<td>.003,.038,.089</td>
<td>1.00,1.00,1.00</td>
<td>1.00,1.00,1.00</td>
</tr>
<tr>
<td>GARCH</td>
<td>$N_{0.1}$</td>
<td>4.10</td>
<td>-.002</td>
<td>.008</td>
<td>.729</td>
<td>1.47</td>
<td>.001,.046,.121</td>
<td>1.00,1.00,1.00</td>
<td>1.00,1.00,1.00</td>
</tr>
</tbody>
</table>

### Trimmed $\overline{ES}_{n,\alpha}^{(e)}$

<table>
<thead>
<tr>
<th>Model</th>
<th>$\epsilon_t$</th>
<th>$\kappa_y$</th>
<th>Bias</th>
<th>RMSE</th>
<th>KS</th>
<th>AD</th>
<th>Null</th>
<th>Alt1</th>
<th>Alt2</th>
</tr>
</thead>
<tbody>
<tr>
<td>AR</td>
<td>$P_{1.5}$</td>
<td>1.50</td>
<td>.208</td>
<td>.082</td>
<td>.191</td>
<td>2.18</td>
<td>.517,.709,.794</td>
<td>.911,.988,.997</td>
<td>1.00,1.00,1.00</td>
</tr>
<tr>
<td>AR</td>
<td>$P_{2.5}$</td>
<td>2.50</td>
<td>.015</td>
<td>.014</td>
<td>2.91</td>
<td>2.43</td>
<td>.991,.149,.214</td>
<td>.922,.963,.970</td>
<td>1.00,1.00,1.00</td>
</tr>
<tr>
<td>AR</td>
<td>$N_{0.1}$</td>
<td>$\infty$</td>
<td>.001</td>
<td>.002</td>
<td>1.01</td>
<td>1.63</td>
<td>.054,.160,.243</td>
<td>1.00,1.00,1.00</td>
<td>1.00,1.00,1.00</td>
</tr>
<tr>
<td>GARCH</td>
<td>$P_{2.5}$</td>
<td>1.50</td>
<td>.035</td>
<td>.031</td>
<td>2.42</td>
<td>8.44</td>
<td>.082,.158,.224</td>
<td>.755,.881,.917</td>
<td>1.00,1.00,1.00</td>
</tr>
<tr>
<td>IGARCH</td>
<td>$N_{0.1}$</td>
<td>2.00</td>
<td>.010</td>
<td>.080</td>
<td>1.15</td>
<td>1.58</td>
<td>.004,.044,.112</td>
<td>1.00,1.00,1.00</td>
<td>1.00,1.00,1.00</td>
</tr>
<tr>
<td>RiskMetric</td>
<td>$N_{0.1}$</td>
<td>2.00</td>
<td>.285</td>
<td>.219</td>
<td>2.13</td>
<td>2.45</td>
<td>.005,.067,.211</td>
<td>1.00,1.00,1.00</td>
<td>1.00,1.00,1.00</td>
</tr>
<tr>
<td>GARCH</td>
<td>$N_{0.1}$</td>
<td>4.10</td>
<td>.003</td>
<td>.007</td>
<td>.727</td>
<td>1.48</td>
<td>.014,.059,.115</td>
<td>1.00,1.00,1.00</td>
<td>1.00,1.00,1.00</td>
</tr>
</tbody>
</table>

### Optimally Trimmed $\overline{ES}_{n,\alpha}^{(2)} = \overline{ES}_{n,\alpha}^{(e)} + \tilde{R}_{n,\alpha}^{(2)}$

<table>
<thead>
<tr>
<th>Model</th>
<th>$\epsilon_t$</th>
<th>$\kappa_y$</th>
<th>Bias</th>
<th>RMSE</th>
<th>KS</th>
<th>AD</th>
<th>Null</th>
<th>Alt1</th>
<th>Alt2</th>
</tr>
</thead>
<tbody>
<tr>
<td>AR</td>
<td>$P_{1.5}$</td>
<td>1.50</td>
<td>.081</td>
<td>.141</td>
<td>.853</td>
<td>1.81</td>
<td>.000,.064,.136</td>
<td>.703,.918,.981</td>
<td>1.00,1.00,1.00</td>
</tr>
<tr>
<td>AR</td>
<td>$P_{2.5}$</td>
<td>2.50</td>
<td>.006</td>
<td>.014</td>
<td>1.43</td>
<td>2.18</td>
<td>.014,.058,.119</td>
<td>.960,.976,.983</td>
<td>1.00,1.00,1.00</td>
</tr>
<tr>
<td>AR</td>
<td>$N_{0.1}$</td>
<td>$\infty$</td>
<td>-.001</td>
<td>.02</td>
<td>.892</td>
<td>1.63</td>
<td>.004,.061,.119</td>
<td>1.00,1.00,1.00</td>
<td>1.00,1.00,1.00</td>
</tr>
<tr>
<td>GARCH</td>
<td>$P_{2.5}$</td>
<td>1.50</td>
<td>.011</td>
<td>.034</td>
<td>1.45</td>
<td>2.88</td>
<td>.020,.065,.120</td>
<td>.857,.922,.944</td>
<td>1.00,1.00,1.00</td>
</tr>
<tr>
<td>IGARCH</td>
<td>$N_{0.1}$</td>
<td>2.00</td>
<td>-.023</td>
<td>.103</td>
<td>1.11</td>
<td>1.77</td>
<td>.012,.058,.109</td>
<td>1.00,1.00,1.00</td>
<td>1.00,1.00,1.00</td>
</tr>
<tr>
<td>RiskMetric</td>
<td>$N_{0.1}$</td>
<td>2.00</td>
<td>-.013</td>
<td>.226</td>
<td>1.39</td>
<td>2.29</td>
<td>.013,.056,.111</td>
<td>1.00,1.00,1.00</td>
<td>1.00,1.00,1.00</td>
</tr>
<tr>
<td>GARCH</td>
<td>$N_{0.1}$</td>
<td>4.10</td>
<td>-.009</td>
<td>.009</td>
<td>.645</td>
<td>1.46</td>
<td>.012,.055,.112</td>
<td>1.00,1.00,1.00</td>
<td>1.00,1.00,1.00</td>
</tr>
</tbody>
</table>

---

a. Root Mean Squared Error.

b. Kolmogorov-Smirnov statistic for a test of standard normality, divided by the 5% critical value.

c. Anderson-Darling statistic for a test of standard normality, divided by the 5% critical value.

d. The hypotheses are Null : $ES = ES_{\alpha}$; Alt1 : $ES = ES_{\alpha}/2$; Alt2 : $ES = 0$.

e. The GARCH model is $y_t = \sigma_t \epsilon_t$, $\sigma_t^2 = \omega + \alpha y_{t-1}^2 + \beta \sigma_{t-1}^2$. IGARCH $\alpha, \beta = .4, .6$; RiskMetric $\alpha, \beta = .06, .94$. 

### TABLE 2.a: $\alpha = .10$ and $n = 100$

<table>
<thead>
<tr>
<th>Model</th>
<th>$\epsilon_t$</th>
<th>$\kappa_y$</th>
<th>Bias</th>
<th>RMSE$^a$</th>
<th>KS$^b$</th>
<th>AD$^c$</th>
<th>Null$^d$</th>
<th>Alt1</th>
<th>Alt2</th>
</tr>
</thead>
<tbody>
<tr>
<td>AR</td>
<td>$P_{1.5}$</td>
<td>1.50</td>
<td>.197</td>
<td>.962</td>
<td>5.63</td>
<td>2.34</td>
<td>.008,.012,.018</td>
<td>.017,.045,.069</td>
<td>.064,.153,.246</td>
</tr>
<tr>
<td>AR</td>
<td>$P_{2.5}$</td>
<td>2.50</td>
<td>.001</td>
<td>.065</td>
<td>.707</td>
<td>3.06</td>
<td>.007,.048,.097</td>
<td>.218,.395,.511</td>
<td>.815,.945,.973</td>
</tr>
<tr>
<td>AR</td>
<td>$N_{0.1}$</td>
<td>$\infty$</td>
<td>.001</td>
<td>.027</td>
<td>.678</td>
<td>1.94</td>
<td>.007,.047,.097</td>
<td>.953,.995,1.000</td>
<td>1.00,1.00,1.000</td>
</tr>
<tr>
<td>GARCH$^e$</td>
<td>$P_{2.5}$</td>
<td>1.50</td>
<td>-.004</td>
<td>.113</td>
<td>1.57</td>
<td>4.91</td>
<td>.023,.052,.086</td>
<td>.15,.371,.515</td>
<td>.830,.943,.970</td>
</tr>
<tr>
<td>IGARCH</td>
<td>$N_{0.1}$</td>
<td>2.00</td>
<td>.067</td>
<td>.188</td>
<td>3.12</td>
<td>1.62</td>
<td>.023,.036,.075</td>
<td>.254,.443,.587</td>
<td>1.00,1.00,1.000</td>
</tr>
<tr>
<td>RiskMetric</td>
<td>$N_{0.1}$</td>
<td>2.00</td>
<td>.122</td>
<td>.126</td>
<td>1.49</td>
<td>1.61</td>
<td>.059,.163,.245</td>
<td>1.00,.100,.100</td>
<td>1.00,1.00,1.000</td>
</tr>
<tr>
<td>GARCH</td>
<td>$N_{0.1}$</td>
<td>4.10</td>
<td>-.002</td>
<td>.074</td>
<td>2.81</td>
<td>1.70</td>
<td>.026,.045,.064</td>
<td>.930,1.00,.100</td>
<td>1.00,1.00,1.000</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Model</th>
<th>$\epsilon_t$</th>
<th>$\kappa_y$</th>
<th>Bias</th>
<th>RMSE$^a$</th>
<th>KS$^b$</th>
<th>AD$^c$</th>
<th>Null$^d$</th>
<th>Alt1</th>
<th>Alt2</th>
</tr>
</thead>
<tbody>
<tr>
<td>AR</td>
<td>$P_{1.5}$</td>
<td>1.50</td>
<td>.899</td>
<td>.119</td>
<td>3.16</td>
<td>2.38</td>
<td>.999,.100,.100</td>
<td>.500,.778,.844</td>
<td>.492,.784,.926</td>
</tr>
<tr>
<td>AR</td>
<td>$P_{2.5}$</td>
<td>2.50</td>
<td>.132</td>
<td>.026</td>
<td>.770</td>
<td>2.97</td>
<td>.992,.996,.998</td>
<td>.039,.093,.160</td>
<td>.865,.949,.970</td>
</tr>
<tr>
<td>AR</td>
<td>$N_{0.1}$</td>
<td>$\infty$</td>
<td>.110</td>
<td>.015</td>
<td>.771</td>
<td>2.16</td>
<td>.100,.100,.100</td>
<td>.013,.066,.110</td>
<td>1.00,1.00,1.000</td>
</tr>
<tr>
<td>GARCH$^e$</td>
<td>$P_{2.5}$</td>
<td>1.50</td>
<td>.179</td>
<td>.062</td>
<td>1.81</td>
<td>5.31</td>
<td>.605,.823,.901</td>
<td>.005,.055,.114</td>
<td>.777,.903,.939</td>
</tr>
<tr>
<td>IGARCH</td>
<td>$N_{0.1}$</td>
<td>2.00</td>
<td>.383</td>
<td>.069</td>
<td>1.89</td>
<td>1.88</td>
<td>.994,.999,.100</td>
<td>.078,.176,.260</td>
<td>1.00,1.00,1.000</td>
</tr>
<tr>
<td>RiskMetric</td>
<td>$N_{0.1}$</td>
<td>2.00</td>
<td>.571</td>
<td>.093</td>
<td>1.69</td>
<td>1.82</td>
<td>.100,.100,.100</td>
<td>.658,.835,.907</td>
<td>1.00,1.00,1.000</td>
</tr>
<tr>
<td>GARCH</td>
<td>$N_{0.1}$</td>
<td>4.10</td>
<td>.164</td>
<td>.034</td>
<td>1.74</td>
<td>1.80</td>
<td>.968,.986,.992</td>
<td>.574,.827,.924</td>
<td>1.00,1.00,1.000</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Model</th>
<th>$\epsilon_t$</th>
<th>$\kappa_y$</th>
<th>Bias</th>
<th>RMSE$^a$</th>
<th>KS$^b$</th>
<th>AD$^c$</th>
<th>Null$^d$</th>
<th>Alt1</th>
<th>Alt2</th>
</tr>
</thead>
<tbody>
<tr>
<td>AR</td>
<td>$P_{1.5}$</td>
<td>1.50</td>
<td>.174</td>
<td>.477</td>
<td>2.13</td>
<td>2.28</td>
<td>.006,.043,.099</td>
<td>.089,.159,.223</td>
<td>.330,.540,.708</td>
</tr>
<tr>
<td>AR</td>
<td>$P_{2.5}$</td>
<td>2.50</td>
<td>-.021</td>
<td>.067</td>
<td>.526</td>
<td>2.45</td>
<td>.006,.054,.111</td>
<td>.319,.536,.654</td>
<td>.877,.962,.977</td>
</tr>
<tr>
<td>AR</td>
<td>$N_{0.1}$</td>
<td>$\infty$</td>
<td>-.012</td>
<td>.049</td>
<td>.676</td>
<td>1.87</td>
<td>.014,.061,.110</td>
<td>.767,.931,.973</td>
<td>1.00,1.00,1.000</td>
</tr>
<tr>
<td>GARCH$^e$</td>
<td>$P_{2.5}$</td>
<td>1.50</td>
<td>-.009</td>
<td>.163</td>
<td>.937</td>
<td>4.20</td>
<td>.019,.045,.090</td>
<td>.407,.664,.784</td>
<td>.805,.921,.955</td>
</tr>
<tr>
<td>IGARCH</td>
<td>$N_{0.1}$</td>
<td>2.00</td>
<td>-.043</td>
<td>.202</td>
<td>1.96</td>
<td>1.60</td>
<td>.019,.068,.112</td>
<td>.215,.387,.504</td>
<td>.707,.968,1.000</td>
</tr>
<tr>
<td>RiskMetric</td>
<td>$N_{0.1}$</td>
<td>2.00</td>
<td>-.142</td>
<td>.165</td>
<td>1.72</td>
<td>1.62</td>
<td>.032,.075,.145</td>
<td>.685,.899,.100</td>
<td>1.00,1.00,1.000</td>
</tr>
<tr>
<td>GARCH</td>
<td>$N_{0.1}$</td>
<td>4.10</td>
<td>-.012</td>
<td>.112</td>
<td>1.91</td>
<td>1.71</td>
<td>.021,.071,.115</td>
<td>.560,.878,.977</td>
<td>1.00,1.00,1.000</td>
</tr>
</tbody>
</table>

---

a. Root Mean Squared Error.
b. Kolmogorov-Smirnov statistic for a test of standard normality, divided by the 5% critical value.
c. Anderson-Darling statistic for a test of standard normality, divided by the 5% critical value.
d. The hypotheses are Null : ES = $ES_\alpha$; Alt1 : ES = $ES_\alpha/2$; Alt2 : ES = 0.
e. The GARCH model is $y_t = \sigma_t \epsilon_t$, $\sigma_t^2 = \omega + \alpha y_{t-1}^2 + \beta \sigma_{t-1}^2$. IGARCH $\alpha, \beta = .4, .6$; RiskMetric $\alpha, \beta = .06, .94$. 

---
TABLE 2.b : $\alpha = .10$ and $n = 500$

<table>
<thead>
<tr>
<th>Model</th>
<th>$\epsilon_t$</th>
<th>$\kappa_y$</th>
<th>Bias</th>
<th>RMSE$^a$</th>
<th>KS$^b$</th>
<th>AD$^c$</th>
<th>Null$^d$</th>
<th>Alt1</th>
<th>Alt2</th>
</tr>
</thead>
<tbody>
<tr>
<td>AR</td>
<td>$P_{1.5}$</td>
<td>1.50</td>
<td>.031</td>
<td>1.09</td>
<td>6.41</td>
<td>1.75</td>
<td>.014,.016,.016</td>
<td>.016,.021,.030</td>
<td>.019,.045,.087</td>
</tr>
<tr>
<td>AR</td>
<td>$P_{2.5}$</td>
<td>2.50</td>
<td>.001</td>
<td>.042</td>
<td>1.28</td>
<td>3.02</td>
<td>.025,.054,.098</td>
<td>.538,.781,.852</td>
<td>.983,.994,.997</td>
</tr>
<tr>
<td>AR</td>
<td>$N_{0.1} \infty$</td>
<td></td>
<td>.015</td>
<td>.042</td>
<td>.858</td>
<td>2.03</td>
<td>.028,.062,.106</td>
<td>.444,.698,.783</td>
<td>.985,.994,.998</td>
</tr>
<tr>
<td>GARCH$^e$</td>
<td>$P_{2.5}$</td>
<td>1.50</td>
<td>.001</td>
<td>.084</td>
<td>2.15</td>
<td>3.83</td>
<td>.018,.056,.098</td>
<td>.308,.613,.727</td>
<td>.941,.978,.992</td>
</tr>
<tr>
<td>IGARCH</td>
<td>$N_{0.1}$</td>
<td>2.00</td>
<td>.031</td>
<td>.203</td>
<td>2.87</td>
<td>2.37</td>
<td>.015,.039,.089</td>
<td>.333,.587,.751</td>
<td>1.00,1.00,1.00</td>
</tr>
<tr>
<td>RiskMetric</td>
<td>$N_{0.1}$</td>
<td>2.00</td>
<td>.037</td>
<td>.289</td>
<td>2.13</td>
<td>1.83</td>
<td>.070,.274,.431</td>
<td>.991,.999,1.00</td>
<td>1.00,1.00,1.00</td>
</tr>
<tr>
<td>GARCH</td>
<td>$N_{0.1}$</td>
<td>4.10</td>
<td>.001</td>
<td>.044</td>
<td>2.35</td>
<td>1.76</td>
<td>.022,.030,.048</td>
<td>1.00,1.00,1.00</td>
<td>1.00,1.00,1.00</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Model</th>
<th>$\epsilon_t$</th>
<th>$\kappa_y$</th>
<th>Bias</th>
<th>RMSE$^a$</th>
<th>KS</th>
<th>AD$^c$</th>
<th>Null$^d$</th>
<th>Alt1</th>
<th>Alt2</th>
</tr>
</thead>
<tbody>
<tr>
<td>AR</td>
<td>$P_{1.5}$</td>
<td>1.50</td>
<td>.836</td>
<td>.108</td>
<td>3.75</td>
<td>1.82</td>
<td>1.00,1.00,1.00</td>
<td>.159,.502,.678</td>
<td>.990,.998,.999</td>
</tr>
<tr>
<td>AR</td>
<td>$P_{2.5}$</td>
<td>2.50</td>
<td>.109</td>
<td>.021</td>
<td>2.14</td>
<td>2.87</td>
<td>1.00,1.00,1.00</td>
<td>.016,.055,.095</td>
<td>.986,.995,.997</td>
</tr>
<tr>
<td>AR</td>
<td>$N_{0.1} \infty$</td>
<td></td>
<td>.086</td>
<td>.008</td>
<td>.364</td>
<td>2.10</td>
<td>1.00,1.00,1.00</td>
<td>.902,.971,.989</td>
<td>1.00,1.00,1.00</td>
</tr>
<tr>
<td>GARCH</td>
<td>$P_{2.5}$</td>
<td>1.50</td>
<td>.144</td>
<td>.051</td>
<td>1.71</td>
<td>4.62</td>
<td>.553,.792,.890</td>
<td>.015,.084,.185</td>
<td>.926,.970,.983</td>
</tr>
<tr>
<td>IGARCH</td>
<td>$N_{0.1}$</td>
<td>2.00</td>
<td>.415</td>
<td>.078</td>
<td>1.54</td>
<td>2.63</td>
<td>.990,.998,1.00</td>
<td>.217,.428,.545</td>
<td>1.00,1.00,1.00</td>
</tr>
<tr>
<td>RiskMetric</td>
<td>$N_{0.1}$</td>
<td>2.00</td>
<td>.139</td>
<td>.212</td>
<td>1.79</td>
<td>1.85</td>
<td>.964,.993,.997</td>
<td>.565,.831,.919</td>
<td>1.00,1.00,1.00</td>
</tr>
<tr>
<td>GARCH</td>
<td>$N_{0.1}$</td>
<td>4.10</td>
<td>.137</td>
<td>.022</td>
<td>1.41</td>
<td>1.93</td>
<td>.997,1.00,1.00</td>
<td>1.00,1.00,1.00</td>
<td>1.00,1.00,1.00</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Model</th>
<th>$\epsilon_t$</th>
<th>$\kappa_y$</th>
<th>Bias</th>
<th>RMSE$^a$</th>
<th>KS</th>
<th>AD$^c$</th>
<th>Null$^d$</th>
<th>Alt1</th>
<th>Alt2</th>
</tr>
</thead>
<tbody>
<tr>
<td>AR</td>
<td>$P_{1.5}$</td>
<td>1.50</td>
<td>.034</td>
<td>.398</td>
<td>2.19</td>
<td>1.34</td>
<td>.021,.053,.089</td>
<td>.140,.280,.368</td>
<td>.663,.929,.982</td>
</tr>
<tr>
<td>AR</td>
<td>$P_{2.5}$</td>
<td>2.50</td>
<td>-.007</td>
<td>.049</td>
<td>1.10</td>
<td>2.53</td>
<td>.013,.061,.112</td>
<td>.597,.797,.868</td>
<td>.976,.991,.996</td>
</tr>
<tr>
<td>AR</td>
<td>$N_{0.1} \infty$</td>
<td></td>
<td>-.009</td>
<td>.050</td>
<td>.901</td>
<td>2.03</td>
<td>.011,.048,.099</td>
<td>.507,.718,.808</td>
<td>.974,.992,.995</td>
</tr>
<tr>
<td>GARCH</td>
<td>$P_{2.5}$</td>
<td>1.50</td>
<td>-.014</td>
<td>.094</td>
<td>1.34</td>
<td>3.24</td>
<td>.016,.054,.094</td>
<td>.207,.454,.646</td>
<td>.869,957,.975</td>
</tr>
<tr>
<td>IGARCH</td>
<td>$N_{0.1}$</td>
<td>2.00</td>
<td>-.038</td>
<td>.214</td>
<td>1.58</td>
<td>2.06</td>
<td>.017,.062,.097</td>
<td>.193,.363,.509</td>
<td>1.00,1.00,1.00</td>
</tr>
<tr>
<td>RiskMetric</td>
<td>$N_{0.1}$</td>
<td>2.00</td>
<td>.031</td>
<td>.346</td>
<td>1.41</td>
<td>1.77</td>
<td>.008,.041,.085</td>
<td>.973,.996,.998</td>
<td>1.00,1.00,1.00</td>
</tr>
<tr>
<td>GARCH</td>
<td>$N_{0.1}$</td>
<td>4.10</td>
<td>-.008</td>
<td>.053</td>
<td>1.48</td>
<td>1.50</td>
<td>.012,.055,.095</td>
<td>1.00,1.00,1.00</td>
<td>1.00,1.00,1.00</td>
</tr>
</tbody>
</table>

---

a. Root Mean Squared Error.

b. Kolmogorov-Smirnov statistic for a test of standard normality, divided by the 5% critical value.

c. Anderson-Darling statistic for a test of standard normality, divided by the 5% critical value.

d. The hypotheses are Null : ES = $ES_0$; Alt1 : ES = $ES_0/2$; Alt2 : ES = 0.

e. The GARCH model is $y_t = \sigma_t \epsilon_t$, $\sigma_t^2 = \omega + \alpha \sigma_{t-1}^2 + \beta \sigma_{t-1}^2$. IGARCH $\alpha, \beta = .4, .6$; RiskMetric $\alpha, \beta = .06, .94.$
### TABLE 2.c : $\alpha = .10$ and $n = 1000$

#### Untrimmed $\hat{ES}_n$

<table>
<thead>
<tr>
<th>Model</th>
<th>$\epsilon_t$</th>
<th>$\kappa_y$</th>
<th>Bias</th>
<th>RMSE$^a$</th>
<th>KS$^b$</th>
<th>AD$^c$</th>
<th>Null$^d$</th>
<th>Alt1</th>
<th>Alt2</th>
</tr>
</thead>
<tbody>
<tr>
<td>AR</td>
<td>$P_{1.5}$</td>
<td>1.50</td>
<td>.002</td>
<td>.111</td>
<td>7.24</td>
<td>1.92</td>
<td>.006,.090,.012</td>
<td>.009,.015,.024</td>
<td>.014,.039,.084</td>
</tr>
<tr>
<td>AR</td>
<td>$P_{2.5}$</td>
<td>2.50</td>
<td>.001</td>
<td>.035</td>
<td>1.28</td>
<td>2.76</td>
<td>.015,.042,.079</td>
<td>.737,.878,.922</td>
<td>.994,.995,.998</td>
</tr>
<tr>
<td>AR</td>
<td>$N_{0.1}$</td>
<td>$\infty$</td>
<td>.000</td>
<td>.009</td>
<td>.548</td>
<td>1.67</td>
<td>.009,.053,.100</td>
<td>1.00,1.00,1.00</td>
<td>1.00,1.00,1.00</td>
</tr>
<tr>
<td>GARCH$^e$</td>
<td>$P_{2.5}$</td>
<td>1.50</td>
<td>.005</td>
<td>.074</td>
<td>2.34</td>
<td>3.89</td>
<td>.036,.062,.097</td>
<td>.383,.674,.783</td>
<td>.965,.981,.987</td>
</tr>
<tr>
<td>IGARCH</td>
<td>$N_{0.1}$</td>
<td>2.00</td>
<td>.024</td>
<td>.153</td>
<td>2.15</td>
<td>2.37</td>
<td>.008,.059,.161</td>
<td>.724,.927,.984</td>
<td>1.00,1.00,1.00</td>
</tr>
<tr>
<td>RiskMetric</td>
<td>$N_{0.1}$</td>
<td>2.00</td>
<td>.057</td>
<td>.338</td>
<td>1.48</td>
<td>2.02</td>
<td>.182,.438,.570</td>
<td>.987,.999,1.00</td>
<td>1.00,1.00,1.00</td>
</tr>
<tr>
<td>GARCH</td>
<td>$N_{0.1}$</td>
<td>4.10</td>
<td>.002</td>
<td>.033</td>
<td>1.94</td>
<td>1.79</td>
<td>.022,.050,.077</td>
<td>1.00,1.00,1.00</td>
<td>1.00,1.00,1.00</td>
</tr>
</tbody>
</table>

#### Trimmed $\hat{ES}_{n,\alpha}$

<table>
<thead>
<tr>
<th>Model</th>
<th>$\epsilon_t$</th>
<th>$\kappa_y$</th>
<th>Bias</th>
<th>RMSE$^a$</th>
<th>KS</th>
<th>AD</th>
<th>Null</th>
<th>Alt1</th>
<th>Alt2</th>
</tr>
</thead>
<tbody>
<tr>
<td>AR</td>
<td>$P_{1.5}$</td>
<td>1.50</td>
<td>.815</td>
<td>.083</td>
<td>3.60</td>
<td>1.88</td>
<td>1.00,1.00,1.00</td>
<td>.269,.594,.746</td>
<td>.996,.997,.999</td>
</tr>
<tr>
<td>AR</td>
<td>$P_{2.5}$</td>
<td>2.50</td>
<td>.101</td>
<td>.029</td>
<td>2.11</td>
<td>2.43</td>
<td>1.00,1.00,1.00</td>
<td>.012,.060,.012</td>
<td>.997,.998,.999</td>
</tr>
<tr>
<td>AR</td>
<td>$N_{0.1}$</td>
<td>$\infty$</td>
<td>.075</td>
<td>.006</td>
<td>1.23</td>
<td>1.67</td>
<td>1.00,1.00,1.00</td>
<td>1.00,1.00,1.00</td>
<td>1.00,1.00,1.00</td>
</tr>
<tr>
<td>GARCH</td>
<td>$P_{2.5}$</td>
<td>1.50</td>
<td>.131</td>
<td>.048</td>
<td>1.69</td>
<td>4.14</td>
<td>.509,.799,.891</td>
<td>.024,.124,.221</td>
<td>.948,.977,.983</td>
</tr>
<tr>
<td>IGARCH</td>
<td>$N_{0.1}$</td>
<td>2.00</td>
<td>.413</td>
<td>.073</td>
<td>1.56</td>
<td>2.49</td>
<td>.997,1.00,1.00</td>
<td>.375,.606,.746</td>
<td>1.00,1.00,1.00</td>
</tr>
<tr>
<td>RiskMetric</td>
<td>$N_{0.1}$</td>
<td>2.00</td>
<td>.129</td>
<td>.240</td>
<td>1.22</td>
<td>2.34</td>
<td>.988,.998,1.00</td>
<td>.677,.874,.948</td>
<td>1.00,1.00,1.00</td>
</tr>
<tr>
<td>GARCH</td>
<td>$N_{0.1}$</td>
<td>4.10</td>
<td>.126</td>
<td>.016</td>
<td>1.31</td>
<td>1.71</td>
<td>1.00,1.00,1.00</td>
<td>1.00,1.00,1.00</td>
<td>1.00,1.00,1.00</td>
</tr>
</tbody>
</table>

#### Optimally Trimmed $\hat{ES}_{n,\alpha}^{(2)} = \hat{ES}_{n,\alpha}^{(s)} + \hat{r}_{n,\alpha}^{(2)}$

<table>
<thead>
<tr>
<th>Model</th>
<th>$\epsilon_t$</th>
<th>$\kappa_y$</th>
<th>Bias</th>
<th>RMSE$^a$</th>
<th>KS</th>
<th>AD</th>
<th>Null</th>
<th>Alt1</th>
<th>Alt2</th>
</tr>
</thead>
<tbody>
<tr>
<td>AR</td>
<td>$P_{1.5}$</td>
<td>1.50</td>
<td>.006</td>
<td>.377</td>
<td>1.75</td>
<td>1.48</td>
<td>.009,.058,.096</td>
<td>.159,.315,.447</td>
<td>.835,.983,.994</td>
</tr>
<tr>
<td>AR</td>
<td>$P_{2.5}$</td>
<td>2.50</td>
<td>-.01</td>
<td>.041</td>
<td>.712</td>
<td>2.15</td>
<td>.015,.064,.113</td>
<td>.624,.821,.895</td>
<td>.991,.994,.996</td>
</tr>
<tr>
<td>AR</td>
<td>$N_{0.1}$</td>
<td>$\infty$</td>
<td>-.003</td>
<td>.009</td>
<td>.402</td>
<td>1.58</td>
<td>.008,.054,.102</td>
<td>1.00,1.00,1.00</td>
<td>1.00,1.00,1.00</td>
</tr>
<tr>
<td>GARCH</td>
<td>$P_{2.5}$</td>
<td>1.50</td>
<td>-.009</td>
<td>.087</td>
<td>1.38</td>
<td>3.03</td>
<td>.020,.055,.096</td>
<td>.274,.560,.707</td>
<td>.934,.976,.984</td>
</tr>
<tr>
<td>IGARCH</td>
<td>$N_{0.1}$</td>
<td>2.00</td>
<td>-.026</td>
<td>1.83</td>
<td>1.42</td>
<td>2.30</td>
<td>.007,.043,.094</td>
<td>.588,.854,.940</td>
<td>1.00,1.00,1.00</td>
</tr>
<tr>
<td>RiskMetric</td>
<td>$N_{0.1}$</td>
<td>2.00</td>
<td>.042</td>
<td>.240</td>
<td>1.10</td>
<td>1.99</td>
<td>.015,.069,.123</td>
<td>.988,.997,1.00</td>
<td>1.00,1.00,1.00</td>
</tr>
<tr>
<td>GARCH</td>
<td>$N_{0.1}$</td>
<td>4.10</td>
<td>.005</td>
<td>.035</td>
<td>1.17</td>
<td>1.68</td>
<td>.020,.049,.091</td>
<td>1.00,1.00,1.00</td>
<td>1.00,1.00,1.00</td>
</tr>
</tbody>
</table>

---

a. Root Mean Squared Error.
b. Kolmogorov-Smirnov statistic for a test of standard normality, divided by the 5% critical value.
c. Anderson-Darling statistic for a test of standard normality, divided by the 5% critical value.
d. The hypotheses are Null : ES = $\hat{ES}_n$; Alt1 : ES = $\hat{ES}_n / 2$; Alt2 : ES = 0.
e. The GARCH model is $y_t = \sigma_t \epsilon_t$, $\sigma_t^2 = \omega + \alpha \sigma_{t-1}^2 + \beta \sigma_{t-1}^2$. IGARCH $\alpha, \beta = .4, .6$; RiskMetric $\alpha, \beta = .06,.94$. 
TABLE 3.d : $\alpha = .10$ and $n = 10,000$

### Untrimmed $\widehat{ES}_{n}$

<table>
<thead>
<tr>
<th>Model</th>
<th>$\epsilon_t$</th>
<th>$\kappa_y$</th>
<th>Bias</th>
<th>RMSE$^a$</th>
<th>KS$^b$</th>
<th>AD$^c$</th>
<th>Null$^d$</th>
<th>Alt1</th>
<th>Alt2</th>
</tr>
</thead>
<tbody>
<tr>
<td>AR</td>
<td>$P_{1.5}$</td>
<td>1.50</td>
<td>.040</td>
<td>.116</td>
<td>2.01</td>
<td>1.33</td>
<td>.007,062,.152</td>
<td>.942,999,1.00</td>
<td>1.00,1.00,1.00</td>
</tr>
<tr>
<td>AR</td>
<td>$P_{2.5}$</td>
<td>2.50</td>
<td>.012</td>
<td>.022</td>
<td>3.07</td>
<td>2.37</td>
<td>.045,.083,1.144</td>
<td>.940,971,978</td>
<td>1.00,1.00,1.00</td>
</tr>
<tr>
<td>AR</td>
<td>$N_{0.1}$</td>
<td>$\infty$</td>
<td>-.000</td>
<td>.028</td>
<td>.509</td>
<td>1.99</td>
<td>.011,.056,107</td>
<td>1.00,1.00,1.00</td>
<td>1.00,1.00,1.00</td>
</tr>
<tr>
<td>GARCH$^e$</td>
<td>$P_{2.5}$</td>
<td>1.50</td>
<td>.030</td>
<td>.051</td>
<td>2.69</td>
<td>4.17</td>
<td>.050,.080,.115</td>
<td>.672,858,897</td>
<td>.986,992,995</td>
</tr>
<tr>
<td>IGARCH</td>
<td>$N_{0.1}$</td>
<td>2.00</td>
<td>-.023</td>
<td>.109</td>
<td>1.71</td>
<td>1.91</td>
<td>.030,.070,.014</td>
<td>1.00,1.00,1.00</td>
<td>1.00,1.00,1.00</td>
</tr>
<tr>
<td>RiskMetric</td>
<td>$N_{0.1}$</td>
<td>2.00</td>
<td>.119</td>
<td>1.75</td>
<td>4.15</td>
<td>2.15</td>
<td>.024,.039,.055</td>
<td>.151,280,398</td>
<td>.968,1.00,1.00</td>
</tr>
<tr>
<td>GARCH</td>
<td>$N_{0.1}$</td>
<td>4.10</td>
<td>.000</td>
<td>.011</td>
<td>1.77</td>
<td>1.79</td>
<td>.022,.051,.095</td>
<td>1.00,1.00,1.00</td>
<td>1.00,1.00,1.00</td>
</tr>
</tbody>
</table>

### Trimmed $\widehat{ES}_{n,\alpha}$

<table>
<thead>
<tr>
<th>Model</th>
<th>$\epsilon_t$</th>
<th>$\kappa_y$</th>
<th>Bias</th>
<th>RMSE$^a$</th>
<th>KS</th>
<th>AD</th>
<th>Null</th>
<th>Alt1</th>
<th>Alt2</th>
</tr>
</thead>
<tbody>
<tr>
<td>AR</td>
<td>$P_{1.5}$</td>
<td>1.50</td>
<td>.159</td>
<td>.100</td>
<td>1.91</td>
<td>1.48</td>
<td>.154,.395,534</td>
<td>1.00,1.00,1.00</td>
<td>1.00,1.00,1.00</td>
</tr>
<tr>
<td>AR</td>
<td>$P_{2.5}$</td>
<td>2.50</td>
<td>.017</td>
<td>.022</td>
<td>2.98</td>
<td>2.34</td>
<td>.059,.117,.158</td>
<td>.909,949,964</td>
<td>1.00,1.00,1.00</td>
</tr>
<tr>
<td>AR</td>
<td>$N_{0.1}$</td>
<td>$\infty$</td>
<td>.001</td>
<td>.003</td>
<td>.510</td>
<td>1.98</td>
<td>.014,.073,.133</td>
<td>1.00,1.00,1.00</td>
<td>1.00,1.00,1.00</td>
</tr>
<tr>
<td>GARCH</td>
<td>$P_{2.5}$</td>
<td>1.50</td>
<td>.041</td>
<td>.051</td>
<td>2.83</td>
<td>5.26</td>
<td>.037,.069,.092</td>
<td>.614,825,881</td>
<td>.984,991,994</td>
</tr>
<tr>
<td>IGARCH</td>
<td>$N_{0.1}$</td>
<td>2.00</td>
<td>.003</td>
<td>.098</td>
<td>2.11</td>
<td>2.16</td>
<td>.023,.072,.132</td>
<td>1.00,1.00,1.00</td>
<td>1.00,1.00,1.00</td>
</tr>
<tr>
<td>RiskMetric</td>
<td>$N_{0.1}$</td>
<td>2.00</td>
<td>.174</td>
<td>1.72</td>
<td>4.11</td>
<td>2.24</td>
<td>.023,.038,.061</td>
<td>.149,281,399</td>
<td>.975,1.00,1.00</td>
</tr>
<tr>
<td>GARCH</td>
<td>$N_{0.1}$</td>
<td>4.10</td>
<td>.004</td>
<td>.010</td>
<td>1.97</td>
<td>1.71</td>
<td>.037,.084,.125</td>
<td>1.00,1.00,1.00</td>
<td>1.00,1.00,1.00</td>
</tr>
</tbody>
</table>

### Optimally Trimmed $\widehat{ES}_{n,\alpha}^{(2)} = \widehat{ES}_{n,\alpha}^{(s)} + \delta_{n,\alpha}^{(2)}$

<table>
<thead>
<tr>
<th>Model</th>
<th>$\epsilon_t$</th>
<th>$\kappa_y$</th>
<th>Bias</th>
<th>RMSE$^a$</th>
<th>KS</th>
<th>AD</th>
<th>Null</th>
<th>Alt1</th>
<th>Alt2</th>
</tr>
</thead>
<tbody>
<tr>
<td>AR</td>
<td>$P_{1.5}$</td>
<td>1.50</td>
<td>.028</td>
<td>.157</td>
<td>1.32</td>
<td>4.85</td>
<td>.010,.050,.093</td>
<td>.949,999,1.00</td>
<td>1.00,1.00,1.00</td>
</tr>
<tr>
<td>AR</td>
<td>$P_{2.5}$</td>
<td>2.50</td>
<td>.010</td>
<td>.022</td>
<td>1.86</td>
<td>2.18</td>
<td>.022,.062,.108</td>
<td>.939,970,978</td>
<td>1.00,1.00,1.00</td>
</tr>
<tr>
<td>AR</td>
<td>$N_{0.1}$</td>
<td>$\infty$</td>
<td>-.001</td>
<td>.003</td>
<td>.578</td>
<td>1.98</td>
<td>.013,.065,.122</td>
<td>1.00,1.00,1.00</td>
<td>1.00,1.00,1.00</td>
</tr>
<tr>
<td>GARCH</td>
<td>$P_{2.5}$</td>
<td>1.50</td>
<td>.017</td>
<td>.054</td>
<td>1.93</td>
<td>2.88</td>
<td>.028,.064,.094</td>
<td>.713,.875,.907</td>
<td>.985,991,994</td>
</tr>
<tr>
<td>IGARCH</td>
<td>$N_{0.1}$</td>
<td>2.00</td>
<td>-.001</td>
<td>.122</td>
<td>1.65</td>
<td>1.72</td>
<td>.016,.052,.082</td>
<td>1.00,1.00,1.00</td>
<td>1.00,1.00,1.00</td>
</tr>
<tr>
<td>RiskMetric</td>
<td>$N_{0.1}$</td>
<td>2.00</td>
<td>.022</td>
<td>1.82</td>
<td>2.03</td>
<td>2.15</td>
<td>.019,.043,.086</td>
<td>.154,295,398</td>
<td>1.00,1.00,1.00</td>
</tr>
<tr>
<td>GARCH</td>
<td>$N_{0.1}$</td>
<td>4.10</td>
<td>-.007</td>
<td>.012</td>
<td>1.62</td>
<td>1.68</td>
<td>.016,.068,.130</td>
<td>1.00,1.00,1.00</td>
<td>1.00,1.00,1.00</td>
</tr>
</tbody>
</table>

a. Root Mean Squared Error.
b. Kolmogorov-Smirnov statistic for a test of standard normality, divided by the 5% critical value.
c. Anderson-Darling statistic for a test of standard normality, divided by the 5% critical value.
d. The hypotheses are Null : $ES = ES_o$; Alt1 : $ES = ES_o/2$; Alt2 : $ES = 0$.
e. The GARCH model is $y_t = \sigma_t \epsilon_t$, $\sigma_t^2 = \omega + \alpha \sigma_{t-1}^2 + \beta \sigma_{t-1}^2$. IGARCH $\alpha, \beta = .4, .6$; RiskMetric $\alpha, \beta = .06, .94$. 

47