A Parametric Bootstrap for Heavy Tailed Distributions

by

Adriana Cornea

Department of Economics
University of Exeter
Streatham Campus
Exeter EX4 4ST, UK

A.Cornea@exeter.ac.uk

and

Russell Davidson

GREQAM
Centre de la Vieille Charité
2, rue de la Charité
13236 Marseille cedex 02, France

Department of Economics and CIREQ
McGill University
Montreal, Quebec, Canada
H3A 2T7

russell.davidson@mcgill.ca

Abstract

It is known that Efron's resampling bootstrap of the mean of random variables with common distribution in the domain of attraction of the stable laws with infinite variance is not consistent, in the sense that the limiting distribution of the bootstrap mean is not the same as the limiting distribution of the mean from the real sample. Moreover, the limiting distribution of the bootstrap mean is random and unknown. The conventional remedy for this problem, at least asymptotically, is either the m out of n bootstrap or subsampling. However, we show that both these procedures can be quite unreliable in other than very large samples. We introduce a parametric bootstrap that overcomes the failure of Efron's resampling bootstrap and performs better than the m out of n bootstrap and subsampling. The quality of inference based on the parametric bootstrap is examined in a simulation study, and is found to be satisfactory with heavy-tailed distributions unless the tail index is close to 1 and the distribution is heavily skewed.

Keywords: bootstrap inconsistency, stable distribution, domain of attraction, t-statistic, self-normalised sum, parametric bootstrap, m out of n bootstrap, subsampling, infinite variance

JEL codes: C10, C12, C15

The authors would like to thank the co-editor and one of the referees for many suggestions which helped to greatly improve the paper. Valuable comments from Keith Knight on an earlier version are also acknowledged. This research was supported by the British Academy's PDF/2009/370, the Canada Research Chair program (Chair in Economics, McGill University) and by grants from the Social Sciences and Humanities Research Council of Canada, and the Fonds Québécois de Recherche sur la Société et la Culture.

November 2012

1 Introduction

Let F be the cumulative distribution function (CDF) of the independent and identically distributed (IID) random variables $Y_1, \ldots, Y_n$. We are interested in the inference on the parameter μ in the location model

$$Y_j = \mu + U_j, \quad \mathrm{E}(U_j) = 0, \quad j = 1, \ldots, n. \tag{1}$$

It has been known since Bahadur and Savage (1956) that such inference is impossible unless moderately restrictive conditions are imposed on the distribution of the disturbances U_j. Here, we investigate bootstrap inference when the variance of the U's does not exist. Even when it does, there are still further conditions needed for inference to be possible.

The focus of this paper is the set of stable laws, and their domains of attraction. Since we know in advance that complete generality is impossible, we hope that considering laws in the domains of attraction of stable laws will provide at least some generality. Our main requirement is that F is in the domain of attraction of a stable law with a tail index α greater than 1 and smaller than or equal to 2. A distribution F is said to be in the domain of attraction of a stable law with $\alpha \leq 2$, if centered and normalized sums of independent and identically distributed variables with that distribution converge in distribution to that stable law. We write $F \in DA(\alpha)$.

The stable laws, introduced by Lévy (1925), are the only possible limiting laws for suitably centered and normalised sums of independent and identically distributed random variables. They allow for asymmetries and heavy tails, properties frequently encountered with financial data. They are characterized by four parameters: the tail index α $(0 < \alpha \leq 2)$, the asymmetry parameter β $(-1 < \beta < 1)$, the scale parameter σ $(\sigma > 0)$, and the location parameter μ. A stable random variable X can be written as $X = \mu + \sigma Z$, where the location parameter of Z is zero, and its scale parameter unity. We write the distribution of X as $S_{\alpha,\beta,\sigma,\mu}$. The distribution of Z is $S_{\alpha,\beta,1,0}$. All the moments of X of order larger than or equal to α do not exist. When $1 < \alpha \leq 2$, the parameter μ in model (1) can be consistently estimated by the sample mean. When $\alpha < 2$, the variance does not exist. When $\alpha = 2$, the distribution $S_{\alpha,\beta,\sigma,\mu}$ is the normal distribution $N(\mu, 2\sigma^2)$ and σ^2 can be consistently estimated by the sample variance.

It is documented in numerous studies that many series in finance and economics are heavy-tailed. The first study goes back to Mandelbrot (1963). More recent studies are Mittnik and Rachev (2000), Ibragimov (2011) and the references therein.

Since there has been some confusion in the literature occasioned by the existence of more than one parametrization of the stable laws, we specify here that the characteristic function of what we have called the $S_{\alpha,\beta,1,0}$ distribution is

$$\mathrm{E}\big(\exp(\mathrm{i}tY)\big) = \exp\big(-|t|^\alpha[1 - \mathrm{i}\beta\tan(\pi\alpha/2)(\operatorname{sign} t)]\big). \tag{2}$$

In simulation exercises, we generate realizations of this distribution using the algorithm proposed by Chambers, Mallows, and Stuck (1981), their formula modified somewhat to take account of their use of a different parametrization. Specifically, a drawing from the $S_{\alpha,\beta,1,0}$ distribution is given by

$$\big(1 + \beta^2\tan^2(\pi\alpha/2)\big)^{1/2\alpha}\frac{\sin(\alpha(W_1 + b(\alpha,\beta)))}{(\cos W_1)^{1/\alpha}}\Big(\frac{\cos(W_1 - \alpha(W_1 + b(\alpha,\beta)))}{W_2}\Big)^{(1-\alpha)/\alpha} \tag{3}$$

where W_1 is uniformly distributed on $[-\pi/2, \pi/2]$, W_2 is exponentially distributed with expectation 1, and $b(\alpha, \beta) = \tan^{-1}\big(\beta \tan(\pi\alpha/2)\big)/\alpha$. W_1 and W_2 are independently generated.

It was shown by Athreya (1987) that, when the variance does not exist, the conventional resampling bootstrap of Efron (1979) is not valid (the bootstrap distribution of the sample mean does not converge to a deterministic distribution as the sample size n tends to infinity). This is due to the fact that the sample mean is greatly influenced by the extreme observations in the sample, and these are very different for the sample under analysis and the bootstrap samples obtained by resampling, as shown clearly in Knight (1989) and Hall (1990a).

A proposed remedy for the failure of the conventional bootstrap is the m out of n bootstrap; see Arcones and Gine (1989). It is based on the same principle as Efron's bootstrap, but the bootstrap sample size is m, smaller than n. If $m/n \to 0$ as $n \to \infty$, this bootstrap is consistent. Like the m out of n bootstrap, the subsampling method proposed in Romano and Wolf (1999) makes use of samples of size m smaller than n, but the subsamples are obtained without replacement. If m is chosen appropriately, this method too is consistent. However, as we will see in simulation experiments, the m out of n bootstrap fails and subsampling does not always provide reliable inference if the sample size is not very large.

In this paper, we introduce a parametric bootstrap for the parameter μ of model (1), that overcomes the failure of the conventional bootstrap test and performs better than the m out of n bootstrap and subsampling. The parametric bootstrap is based on a central limit argument for self-normalised sums. Section 2 introduces the main theoretical results, followed by a simulation study in Section 3 and a conclusion in Section 4.

2 Main results

2.1 Introduction

Let $\{Y_j\}_{j=1}^n$ be IID random variables from a distribution $F \in DA(\alpha)$, $1 < \alpha < 2$. Restricting α to be greater than 1 ensures that the expectation of F exists. Then, the parameter μ of model (1) can be consistently estimated by the sample mean $\bar{Y}_n = n^{-1}\sum_{j=1}^n Y_j$. Ibragimov (2007) shows that the sample mean is the best linear unbiased estimator of the mean of heavy-tailed populations with $\alpha > 1$, in the sense of peakedness.

Suppose we wish to test the hypothesis

$$H_0 : \mu = \mu_0. \tag{4}$$

A possible test statistic is

$$\tau_n = \frac{\sum_{j=1}^n (Y_j - \mu_0)}{a_n} = \frac{\bar{Y}_n - \mu_0}{n^{-1} a_n}, \tag{5}$$

where a_n is a positive constant and $a_n \to \infty$. For the case when $F = S_{\alpha,\beta,\sigma,\mu}$ we have $a_n = n^{1/\alpha}$ as the stable distributions are in their own domain of attraction (Feller (1971), p.576). This choice of a_n also holds for the Pareto, Burr and t distributions with $1 < \alpha < 2$, but not for the log-gamma distribution with scale $\nu > 1$, for which $a_n = \big(\Gamma(\nu)^{-1}(\log n)^{\nu-1} n\big)^{1/\alpha}$ (see Embrechts, Kluppelberg, and Mikosch (1997), p.133-134). In general, the choice of a_n can be made such that $n\int_{-a_n}^{a_n} y^2 \,\mathrm{d}F(y)/a_n^2 \to 1$. However, the choice of a_n is not unique. Any other sequence b_n such that b_n/a_n tends to a positive limit may be used in place of a_n (Feller (1971), p.314-315; Mittnik, Rachev, and Kim (1998), p.343).

The distribution F from $DA(\alpha)$ has the property that a sum of IID random variables from F, suitably centred and normalised, has an asymptotic stable distribution (Gnedenko and Kolmogorov (1954), Theorem 2 p.227; Feller (1971) Definition 2 p.172, p.312). We call this property the Generalized Central Limit Theorem (GCLT). Hence, by the GCLT, the asymptotic distribution of τ_n under the true null (4) is the stable distribution $S_{\alpha,\beta,\sigma,0}$. If σ, α, and β are known, then we can perform asymptotic inference by comparing the realization of the statistic τ_n with a quantile of the stable distribution $S_{\alpha,\beta,\sigma,0}$. The asymptotic P value for a test that rejects in the left tail of the distribution is

$$P_{n,\sigma,\alpha,\beta} = S_{\alpha,\beta,\sigma,0}(\tau_n). \tag{6}$$

The hypothesis (4) can also be tested using the $t-$statistic

$$T_n = \frac{n^{1/2}(\bar{Y}_n - \mu_0)}{\left((n-1)^{-1}\sum_{j=1}^n (Y_j - \bar{Y}_n)^2\right)^{1/2}}. \tag{7}$$

Efron (1969) shows that T_n has a limiting distribution that coincides with the limiting distribution of the self-normalised sum

$$t_n = \frac{\sum_{j=1}^n (Y_j - \mu_0)}{(\sum_{j=1}^n (Y_j - \mu_0)^2))^{1/2}}. \tag{8}$$

This follows from noticing that

$$T_n = t_n[(n-1)/(n-t_n^2)]^{1/2}, \tag{9}$$

where $(n-1)/(n-t_n^2)$ converges to one in probability.

Logan, Mallows, Rice, and Shepp (1973) (henceforth LMRS) derive the expression of the limiting distribution of t_n for the case when $\{Y_j\}_{j=1}^n$ are IID from $F \in DA(\alpha)$, $1 < \alpha < 2$ and $\mathrm{E}(Y_1) = 0$. We denote the limiting distribution of t_n and T_n under (4) by $G_{\alpha,\beta}$. Hence, the asymptotic P value for a test that rejects in the left tail of $G_{\alpha,\beta}$ is

$$P_{n,\alpha,\beta} = G_{\alpha,\beta}(T_n). \tag{10}$$

One advantage of (10) over (6) is that the parameter σ does not have to be known as t_n and T_n are scale invariant. However, explicit estimation of $G_{\alpha,\beta}$ is a nontrivial task because this distribution is expressed in terms of integrals of parabolic cylinder functions in the complex plane. Moreover, care is needed when evaluating the integrals in the complex plane to make sure that the transition from the rectangular to polar representation of complex numbers is done correctly.

A well-known alternative to the asymptotic tests is the bootstrap. If $F \in DA(\alpha)$, $1 < \alpha < 2$, as assumed in this paper, the ordinary nonparametric bootstrap, based on resampling with replacement, is asymptotically invalid, as shown by Athreya (1987), Knight (1989), and Hall (1990a). The invalidity is due to the fact that the nonparametric bootstrap fails to model the relationship among extreme order statistics in the sample correctly. Asymptotic and finite-sample properties of the bootstrap mean are dictated precisely by the behavior of the extreme order statistics.

The solutions to the nonparametric bootstrap failure proposed so far in the literature are the m out of n bootstrap (Athreya (1985), Arcones and Gine (1989), Bickel, Gotze, and van Zwet (1997), Hall and Jing (1998)) and subsampling (Romano and Wolf (1999), Politis, Romano, and Wolf (1999)). These methods can be based on the non-studentised bootstrap statistic

$$\tau_m^* = \frac{\sum_{j=1}^m (Y_j^* - \bar{Y}_n)}{a_m} \tag{11}$$

or the studentised bootstrap statistics

$$t_m^* = \frac{\sum_{j=1}^m (Y_j^* - \bar{Y}_n)}{\left(\sum_{j=1}^m (Y_j^* - \bar{Y}_n)^2\right)^{1/2}}, \quad T_m^* = \frac{m^{1/2}(\bar{Y}_m^* - \bar{Y}_n)}{\left((m-1)^{-1}\sum_{j=1}^m (Y_j^* - \bar{Y}_m^*)^2\right)^{1/2}} \tag{12}$$

where the Y_j^*'s are drawings from the empirical distribution function (EDF) of $\{Y_j\}_{j=1}^n$ and have mean $\bar{Y}_n$; a_m is a positive constant, $a_m \to \infty$ and $\bar{Y}_m^* = m^{-1}\sum_{j=1}^m Y_j^*$. The bootstrap P value is then given by the proportion of bootstrap statistics more extreme than τ_n, t_n, or T_n, depending on the specific choice of test statistic. The asymptotic validity of the m out of n bootstrap and subsampling is guaranteed by taking a bootstrap sample of size $m < n$ such that $m/n \to 0$ as n and $m \to \infty$. If m fails to satisfy these conditions, the m out of n bootstrap and subsampling distributions are random (as can be concluded from Hall (1990a) and Hall and Yao (2003)). Moreover, Hall and Jing (1998) show that in the case of the statistic τ_n, for optimal choice of m and for a certain class of distributions $F \in DA(\alpha)$, the m out of n bootstrap has an error of order larger than the error of the asymptotic test based on $S_{\alpha,\beta,\sigma,0}$ and $G_{\alpha,\beta}$, for given α, β and σ. The same conclusion applies for subsampling because in the IID case, the difference between resampling with and without replacement is asymptotically negligible if $m^2/n \to 0$ (Politis et al. (1999), page 48). In addition, our simulations in Section 3 indicate that in case of the t- statistic T_n, both m out of n bootstrap and subsampling exhibit large size distortions unless the sample size n is very large.

In this paper we introduce a parametric bootstrap that overcomes the failure of the ordinary nonparametric bootstrap and performs better than the m out of n bootstrap and subsampling. The parametric bootstrap is based on the argument that the distribution $G_{\alpha,\beta}$ is the limiting distribution of T_n for IID $\{Y_j\}_{j=1}^n$ having any distribution $F \in DA(\alpha)$, $1 < \alpha < 2$.

2.2 A parametric bootstrap

The parametric bootstrap for testing (4) that we propose here is described by the following steps.

1. Suppose we have a sample $\mathcal{H} = \{Y_j\}_{j=1}^n$ of IID random variables with distribution $F \in DA(\alpha)$, $1 < \alpha < 2$. Compute the t-statistic T_n.

2. Estimate α and β consistently from the original sample.

3. Draw B bootstrap samples, $Z_1^*, \ldots, Z_n^*$, from $S_{\hat{\alpha}_n,\hat{\beta}_n,1,0}$ (the estimate of $S_{\alpha,\beta,1,0}$) using (3), with $\hat{\alpha}_n$ and $\hat{\beta}_n$ obtained in the previous step. Set $Y_j^* = \mu_0 + Z_j^*$ to satisfy (4).

4. For each sample of stable random variables generated in the previous step compute the bootstrap t-statistic

$$T_n^* = \frac{n^{1/2}(\bar{Y}_n^* - \mu_0)}{\left((n-1)^{-1}\sum_{j=1}^n (Y_j^* - \bar{Y}_n^*)^2\right)^{1/2}} \tag{13}$$

with $\bar{Y}_n^* = n^{-1}\sum_{j=1}^n Y_j^*$.

5. The bootstrap P value is equal to the proportion of bootstrap statistics more extreme than T_n

$$P^*_{B,n,\hat{\alpha}_n,\hat{\beta}_n} = \frac{1}{B}\sum_{i=1}^{B} \mathrm{I}(T^*_{n,i} \leq T_n), \tag{14}$$

where I is the indicator function whose value is 1, when its Boolean argument is true, and 0 when it is false.

As $B \to \infty$, by the strong law of large numbers

$$P^*_{B,n,\hat{\alpha}_n,\hat{\beta}_n} \to^{a.s.} P^*_{n,\hat{\alpha}_n,\hat{\beta}_n} \tag{15}$$

where $P^*_{n,\hat{\alpha}_n,\hat{\beta}_n} = G^*_{n,\hat{\alpha}_n,\hat{\beta}_n}(T_n)$ with $G^*_{n,\hat{\alpha}_n,\hat{\beta}_n}(x) = \Pr(T_n^* \leq x|\mathcal{H})$ the finite sample distribution of T_n^* conditional on the random sample $\mathcal{H}$. The randomness of $G^*_{n,\hat{\alpha}_n,\hat{\beta}_n}$ comes from two sources: from the IID Y^*'s and from the estimators $\hat{\alpha}_n$, $\hat{\beta}_n$ which are functions of $\mathcal{H}$. We also denote by $G_{n,\alpha,\beta}$ the (unknown) finite sample distribution of T_n under (4). The asymptotic validity of the parametric bootstrap relies on Assumptions 1–3.

Assumption 1 *The random variables $\{Y_j\}_{j=1}^n$ are IID and have a distribution $F \in DA(\alpha)$, $1 < \alpha < 2$.*

Assumption 1 implies that the sum tail

$$\frac{1-F(y)}{1-F(y)+F(-y)} \to q_r, \quad \frac{F(-y)}{1-F(y)+F(-y)} \to q_l \tag{16}$$

is balanced, as $y \to \infty$, where q_r and q_l are positive constants, $q_r + q_l = 1$ (Feller (1971), Theorem 2 p.577). Moreover, by Assumption 1,

$$1 - F(y) + F(-y) \approx y^{-\alpha}L(y), \quad y \to \infty, \tag{17}$$

where L is a slowly varying function at infinity, i.e. for any $x > 0$, $\lim_{y\to\infty}[L(xy)/L(y)] = 1$; see Bingham, Goldie, and Teugles (1989). For example L can be a constant, a function converging to a constant, a logarithmic function, an iterated logarithmic function or power of these. Condition (16) guarantees that

$$n\,[1 - F(a_n y)] \to cq_r \frac{2-\alpha}{\alpha} y^{-\alpha} = ry^{-\alpha}, \quad y > 0 \tag{18}$$

$$nF(a_n y) \to cq_l \frac{2-\alpha}{\alpha} y^{-\alpha} = l(-y)^{-\alpha}, \quad y < 0 \tag{19}$$

as $n \to \infty$, $a_n \to \infty$, where c, r and l are positive constants (Feller (1971) p.576). If $q_r = q_l$, then the tails are symmetric.

Denote by $\hat{\alpha}_n$, $\hat{l}_n$ and $\hat{r}_n$ the estimators of α, l and r based on $\mathcal{H}$.

Assumption 2 $\text{plim}_{n\to\infty}\hat{\alpha}_n = \alpha$, $\text{plim}_{n\to\infty}\hat{l}_n = l$, $\text{plim}_{n\to\infty}\hat{r}_n = r$.

The parameter β is linked to the parameters r and l by the following relation

$$\beta = \frac{r-l}{r+l} \tag{20}$$

which follows from noting that for the stable distribution with scale one the tails behave like (18) and (19) with $q_r = (1+\beta)/2$ and $q_l = (1-\beta)/2$ and $c = (2-\alpha)\left(\int_0^\infty y^{-\alpha}\sin y\,\mathrm{d}y\right)^{-1}/\alpha$; see Samorodnitsky and Taqqu (1994) p.16 and Feller (1971) p.576. Hence, $r = c(1+\beta)/2$ and $l = c(1-\beta)/2$ from which (20) follows. Hence under Assumption 2, the parameter β can be consistently estimated by $\hat{\beta}_n = (\hat{r}_n - \hat{l}_n)/(\hat{r}_n + \hat{l}_n)$.

Let $(\tilde{\alpha},\tilde{\beta}) \in \Theta$, a bounded subset $\subset (1,2)\times(1,2)$, and let $\eta(z,\tilde{\alpha}) = \exp(\mathrm{i}v_1 z n^{1/\alpha-1/\tilde{\alpha}} + \mathrm{i}v_2 z^2 n^{2/\alpha-2/\tilde{\alpha}}) - 1$, where $z, v_1, v_2 \in \mathbb{R}$ and $\mathrm{i} = \sqrt{-1}$. Define I_n^* to be $\int_{-\infty}^{\infty}\eta(z,\tilde{\alpha})n\,\mathrm{d}S_{\tilde{\alpha},\tilde{\beta},1,\mu_0}(z)$.

Assumption 3 *The map* $(\tilde{\alpha},\tilde{\beta}) \to I_n^*$ *is continuous at the true value* (α,β).

Assumption 3 is satisfied if the Y^*'s are drawn using (3). This formula generates random variables with characteristic function (2) which is continuous in α and β for $1 < \alpha \leq 2$ (Zolotarev (1986), Samorodnitsky and Taqqu (1994), p.7). Assumption 3 is the analogue of the condition on p.1199 from Bickel and Freedman (1981), which together with Assumptions 1-2 guarantee that I_n^* converges in probability to a nonrandom limit as $n \to \infty$. This result is needed in the proof of the asymptotic validity of the parametric bootstrap.

The weak consistency of the parametric bootstrap (as defined in Shao and Tu (1995), p.72) is showed in Proposition 1 below.

Proposition 1 *Under Assumptions 1–3,*

$$\sup_{x\in\mathbb{R}} \mid G^*_{n,\hat{\alpha}_n,\hat{\beta}_n}(x) - G_{n,\alpha,\beta}(x) \mid \to 0 \tag{21}$$

in probability, as $n \to \infty$.

Proof In this proposition we prove the uniform convergence in probability to zero of the discrepancy between the actual distribution of T_n and the distribution of T_n^*. Since we know the asymptotic distribution of T_n (under Assumption 1 it is derived in LMRS) and since this distribution is continuous, we have that

$$\sup_{x\in\mathbb{R}} \mid G_{n,\alpha,\beta}(x) - G_{\alpha,\beta}(x) \mid \to 0 \tag{22}$$

by Polya's theorem (Serfling (1980), p. 20). It then remains to show that

$$\sup_{x\in\mathbb{R}} \mid G^*_{n,\hat{\alpha}_n,\hat{\beta}_n}(x) - G_{\alpha,\beta}(x) \mid \to 0 \tag{23}$$

in probability under F. We can arrive at this conclusion by imitating the proof of (22). To do this it is convenient to rederive the asymptotic joint characteristic function of τ_n and

$$\lambda_n = \left(\frac{\sum_{j=1}^{n}(Y_j-\mu_0)^2}{a_n^2}\right)^{1/2} \tag{24}$$

for the general case when $F \in DA(\alpha)$, not just for $F = S_{\alpha,\beta,\sigma,0}$, as initially derived in LMRS . The joint characteristic function of τ_n and λ_n^2 is

$$\begin{aligned}
\varphi_n(v_1, v_2) &= \mathrm{E}\left(\mathrm{e}^{\mathrm{i}v_1\tau_n + \mathrm{i}v_2\lambda_n^2}\right) \\
&= \left[\mathrm{E}\exp\left(\mathrm{i}v_1\frac{Y_1-\mu_0}{a_n} + \mathrm{i}v_2\frac{(Y_1-\mu_0)^2}{a_n^2}\right)\right]^n \\
&= \left[1 + \int_{-\infty}^{+\infty}\left[\exp\left(\mathrm{i}v_1\frac{y-\mu_0}{a_n} + \mathrm{i}v_2\frac{(y-\mu_0)^2}{a_n^2}\right) - 1\right]\mathrm{d}F(y)\right]^n \\
&= \left[1 + \frac{1}{n}\int_{-\infty}^{+\infty}\left[\exp\left(\mathrm{i}v_1 z + \mathrm{i}v_2 z^2\right) - 1\right] n\,\mathrm{d}F(a_n z + \mu_0)\right]^n .
\end{aligned}$$

The second equality follows from the IID assumption of the Y's. The fourth equality follows by taking $z = (y - \mu_0)/a_n$. Denote

$$\begin{aligned}
I_n &= \int_{-\infty}^{+\infty}\left(\exp\left(\mathrm{i}v_1 z + \mathrm{i}v_2 z^2\right) - 1\right) n\,\mathrm{d}F(a_n z + \mu_0) \\
&= \int_{-\infty}^{0}\left(\exp\left(\mathrm{i}v_1 z + \mathrm{i}v_2 z^2\right) - 1\right) n\,\mathrm{d}F(a_n z + \mu_0) \\
&+ \int_{0}^{\infty}\left(\exp\left(\mathrm{i}v_1 z + \mathrm{i}v_2 z^2\right) - 1\right) n\,\mathrm{d}F(a_n z + \mu_0).
\end{aligned}$$

From (18) and (19) we have that

$$n\,\mathrm{d}F(a_n z + \mu_0) \to \alpha r z^{-\alpha-1}\,\mathrm{d}z, \quad z > 0 \tag{25}$$

$$n\,\mathrm{d}F(a_n z + \mu_0) \to \alpha l(-z)^{-\alpha-1}\,\mathrm{d}z, \quad z < 0 \tag{26}$$

as $n \to \infty$. Hence

$$\lim_{n\to\infty}\varphi_n(v_1, v_2) = \lim_{n\to\infty}(1 + \frac{1}{n}I_n)^n = \varphi_{\alpha,r,l}(v_1, v_2) \tag{27}$$

with

$$\varphi_{\alpha,r,l}(v_1, v_2) = \exp\left[\alpha l\int_{-\infty}^{0}\frac{\exp\left(\mathrm{i}v_1 z + \mathrm{i}v_2 z^2\right) - 1}{z^{\alpha+1}}\,\mathrm{d}z + \alpha r\int_{0}^{\infty}\frac{\exp\left(\mathrm{i}v_1 z + \mathrm{i}v_2 z^2\right) - 1}{z^{\alpha+1}}\,\mathrm{d}z\right]. \tag{28}$$

We now show that the asymptotic joint characteristic function of the numerator and denominator of

$$t_n^* = \frac{\sum_{j=1}^n (Y_j^* - \mu_0)}{\left(\sum_{j=1}^n (Y_j^* - \mu_0)^2\right)^{1/2}}, \tag{29}$$

appropriately normalised, is given by (28). The numerator and denominator of t_n^* appropriately normalised are

$$\tau_n^* = \frac{\sum_{j=1}^n Y_j^* - \mu_0}{n^{1/\hat{\alpha}_n}}, \qquad \lambda_n^* = \left(\frac{\sum_{j=1}^n ((Y_j^*)^2 - \mu_0)^2}{n^{2/\hat{\alpha}_n}}\right)^{1/2}. \tag{30}$$

The Y_j^*'s are drawn from $S_{\hat{\alpha}_n,\hat{\beta}_n,1,\mu_0}$, hence a_n is estimated by $n^{1/\hat{\alpha}_n}$. The joint characteristic function of τ_n^* and λ_n^{*2} is

$$\begin{aligned}
&\varphi^*_{n,\hat{\alpha}_n,\hat{r}_n,\hat{l}_n}(v_1,v_2) = \mathrm{E}\left(\mathrm{e}^{\mathrm{i}v_1\tau_n^*+\mathrm{i}v_2\lambda_n^{*2}}\middle|\mathcal{H}\right)\\
=&\left[\mathrm{E}\left(\exp\left(\mathrm{i}v_1\frac{Y_1^*-\mu_0}{n^{1/\hat{\alpha}_n}}+\mathrm{i}v_2\frac{(Y_1^*-\mu_0)^2}{n^{2/\hat{\alpha}_n}}\right)\middle|\mathcal{H}\right)\right]^n\\
=&\left[1+\frac{1}{n}\int_{-\infty}^{+\infty}\left(\exp\left(\mathrm{i}v_1\frac{y-\mu_0}{n^{1/\hat{\alpha}_n}}+\mathrm{i}v_2\frac{(y-\mu_0)^2}{n^{2/\hat{\alpha}_n}}\right)-1\right)n\,\mathrm{d}S_{\hat{\alpha}_n,\hat{\beta}_n,1,\mu_0}(y)\right]^n \qquad (31)
\end{aligned}$$

where the conditional expectation $\mathrm{E}(\cdot|\mathcal{H})$ is with respect to the distribution $S_{\hat{\alpha}_n,\hat{\beta}_n,1,\mu_0}$. The second equality follows from the fact that the Y^*'s are conditionally independent. We have to show that the probability limit of (31) is (28). Denote $z=(y-\mu_0)/n^{1/\alpha}$ and let

$$\begin{aligned}
I_n^* &= \int_{-\infty}^{+\infty}\left(\exp\left(\mathrm{i}v_1\frac{y-\mu_0}{n^{1/\hat{\alpha}_n}}+\mathrm{i}v_2\frac{(y-\mu_0)^2}{n^{2/\hat{\alpha}_n}}\right)-1\right)n\,\mathrm{d}S_{\hat{\alpha}_n,\hat{\beta}_n,1,\mu_0}(y)\\
&= \int_{-\infty}^{+\infty}\left[\exp\left(\mathrm{i}v_1 z n^{\frac{1}{\alpha}-\frac{1}{\hat{\alpha}_n}}+\mathrm{i}v_2 z^2 n^{\frac{2}{\alpha}-\frac{2}{\hat{\alpha}_n}}\right)-1\right]n\,\mathrm{d}S_{\hat{\alpha}_n,\hat{\beta}_n,1,\mu_0}(n^{1/\alpha}z+\mu_0).
\end{aligned}$$

By Assumption 2 and from the discussion thereafter we have

$$\hat{\alpha}_n-\alpha=O_p(h_{\hat{\alpha}_n}(n,\alpha)) \quad \text{and} \quad \hat{\beta}_n-\beta=O_p(h_{\hat{\beta}_n}(n,\alpha)), \qquad (32)$$

where the rates of convergence $h_{\hat{\alpha}_n}(n,\alpha)\to 0$ and $h_{\hat{\beta}_n}(n,\alpha)\to 0$ as $n\to\infty$. The rates depend on n and α and they are discussed explicitly in Section 2.4. Denote by $h(n,\alpha)=\max(h_{\hat{\alpha}_n}(n,\alpha),h_{\hat{\beta}_n}(n,\alpha))$. Hence

$$\begin{aligned}
I_n^* =& \int_{-\infty}^{+\infty}\left(\exp\left(\mathrm{i}v_1 z n^{-O_p(h_{\hat{\alpha}_n}(n,\alpha))}+\mathrm{i}v_2 z^2 n^{-2O_p(h_{\hat{\alpha}_n}(n,\alpha))}\right)-1\right)n\,\mathrm{d}\Big(S_{\alpha,\beta,1,\mu_0}(n^{1/\alpha}z+\mu_0)+O_p(h(n,\alpha))\Big)\\
=& \int_{-\infty}^{0}\left(\exp\left(\mathrm{i}v_1 z n^{-O_p(h_{\hat{\alpha}_n}(n,\alpha))}+\mathrm{i}v_2 z^2 n^{-2O_p(h_{\hat{\alpha}_n}(n,\alpha))}\right)-1\right)n\,\mathrm{d}\Big(S_{\alpha,\beta,1,\mu_0}(n^{1/\alpha}z+\mu_0)+O_p(h(n,\alpha))\Big)\\
+& \int_{0}^{+\infty}\left(\exp\left(\mathrm{i}v_1 z n^{-O_p(h_{\hat{\alpha}_n}(n,\alpha))}+\mathrm{i}v_2 z^2 n^{-2O_p(h_{\hat{\alpha}_n}(n,\alpha))}\right)\right)n\,\mathrm{d}\Big(S_{\alpha,\beta,1,\mu_0}(n^{1/\alpha}z+\mu_0)+O_p(h(n,\alpha))\Big).
\end{aligned}$$

From (18) and (19), the discussion after (20) and by Assumption 3, it follows that

$$n\,\mathrm{d}\Big(S_{\alpha,\beta,1,\mu_0}(n^{1/\alpha}z+\mu_0)+O_p(h(n,\alpha))\Big) \quad \to^p \quad \alpha r z^{-\alpha-1}\,\mathrm{d}z, \quad z>0, \qquad (33)$$

$$n\,\mathrm{d}\Big(S_{\alpha,\beta,1,\mu_0}(n^{1/\alpha}z+\mu_0)+O_p(h(n,\alpha))\Big) \quad \to^p \quad \alpha l(-z)^{-\alpha-1}\,\mathrm{d}z, \quad z<0, \qquad (34)$$

as $n\to\infty$. Hence, by Assumption 3, I_n^* converges in probability to a nonrandom limit as $n\to\infty$ and

$$\mathrm{plim}_{n\to\infty}\,\varphi^*_{n,\hat{\alpha}_n,\hat{r}_n,\hat{l}_n}(v_1,v_2)=\mathrm{plim}_{n\to\infty}\left(1+\frac{1}{n}I_n^*\right)^n=\varphi_{\alpha,r,l}(v_1,v_2). \qquad (35)$$

From (27) and (35) we conclude that (τ_n,λ_n^2) and $(\tau_n^*,\lambda_n^{*2})$ have the same asymptotic joint characteristic function. Reiterating the same arguments from LMRS, it follows by the continuity theorem that both (τ_n,λ_n^2) and $(\tau_n^*,\lambda_n^{*2})$ have the same limit distribution. Since λ_n^2 and λ_n^{*2} have the same limit distribution concentrated on the positive axis, *i.e.* the stable distribution $S_{\alpha/2,1,1,\mu_0}$ (Mittnik et al. (1998)), $t_n=\tau_n/\lambda_n$ and $t_n^*=\tau_n^*/\lambda_n^*$ have the same asymptotic distribution $G_{\alpha,\beta}$ derived in LMRS. Moreover, the limiting distributions of t_n and T_n coincide (Efron (1969)). Finally since $G_{\alpha,\beta}$ is continuous, by Polya's theorem we conclude that (23) holds. □

Corollary 1 *Under the true null hypothesis, the bootstrap P value* $P^*_{n,\hat{\alpha}_n,\hat{\beta}_n}$ *(15) has the uniform* $\mathrm{U}(0,1)$ *distribution asymptotically.*

Proof Let

$$t_\infty = \frac{\sum_{j=1}^{\infty}(Y_j - \mu_0)}{\left(\sum_{j=1}^{\infty}(Y_j - \mu_0)^2\right)^{1/2}}$$

which has distribution $G_{\alpha,\beta}$. The distribution of $P^*_{n,\hat{\alpha}_n,\hat{\beta}_n}$ is

$$\begin{aligned}\Pr\left(G^*_{n,\hat{\alpha}_n,\hat{\beta}_n}(T_n) \le u\right) &= \Pr\left[G_{\alpha,\beta}\left(t_\infty + O_p(n^{-1})\right) + O_p(h(n,\alpha)) \le u\right] \\ &= \Pr\left(t_\infty + O_p(n^{-1}) \le G^{-1}_{\alpha,\beta}\left(u - O_p(h(n,\alpha))\right)\right) \\ &= G_{\alpha,\beta}\left[G^{-1}_{\alpha,\beta}\left(u - O_p(h(n,\alpha))\right) - O_p(n^{-1})\right]. \end{aligned} \tag{36}$$

The first equality is based on the fact that $T_n = t_\infty + O_p(n^{-1})$ since the convergence rate of φ_n is n^{-1}; see (27). Also by Assumption 2, $G^*_{n,\hat{\alpha}_n,\hat{\beta}_n} = G_{\alpha,\beta} + O_p(h(n,\alpha))$. For finite n, $G^*_{n,\hat{\alpha}_n,\hat{\beta}_n}$ and T_n are random and dependent through the $Y's$, but asymptotically they are independent as $G^*_{n,\hat{\alpha}_n,\hat{\beta}_n}$ collapses to the nonrandom distribution $G_{\alpha,\beta}$. Hence the second equality follows. As $n \to \infty$, (36) converges in probability to $u \in \mathbb{R}$. □

Remark 1. When $F \in DA(2)$, the limiting distribution of T_n^* and T_n is $G_{2,0}$ which is given by the standard normal distribution, as shown in Gine, Gotze, and Mason (1997). Thus, the parametric bootstrap is based on drawings from the normal distribution with mean 0 and variance 2 (Samorodnitsky and Taqqu (1994) p.20).

Remark 2. The t-statistic T_n is not an asymptotic pivot, since its asymptotic distribution depends on α and β. As suggested by Beran (1988), one way to achieve complete asymptotic pivotalness is to transform T_n by its limiting distribution function. The resulting statistic $G_{\alpha,\beta}(T_n)$ is an asymptotic pivot with limiting uniform $\mathrm{U}(0,1)$ distribution. This procedure entails evaluating integrals with parabolic cylinder functions in the complex plane. One way to avoid this is to estimate the integrals by the EDF, but it would be as computationally intensive as a double bootstrap.

An advantage of the parametric bootstrap of T_n (or t_n) is that its asymptotic distribution applies not only when the Y_j's are generated from a stable distribution, but also whenever they are generated by any distribution in the domain of attraction of a stable law. This leaves us with the practical problem of obtaining good estimates of the parameters. The location and scale parameters are irrelevant for the bootstrap, as we can generate centered simulated variables, and the statistic T_n, being normalized, is invariant to scale.

2.3 Estimation of α and β

The problem of estimating the parameters α and β is hampered by the fact that the stable distributions and the limiting distribution of T_n do not have a closed form. Unless assumptions about the parametric form of the distribution generating the data are made, the only estimation methods that could be employed are those that use just the information in the tails. The most popular estimation method and the one we use in the simulation study of Section 3, is proposed by Hill (1975). Hill's method is based on the k largest order statistics $Y_{1,n} \geq Y_{2,n} \geq \cdots \geq Y_{n,n}$ from a sample of IID random variables, and gives the following estimator

$$\hat{\alpha}_n = \Big(\frac{1}{k-1}\sum_{j=1}^{k-1} \log Y_{j,n} - \log Y_{k,n}\Big)^{-1},$$

where $k = k(n) \to \infty$ in an appropriate way.

On account of (20), for any distribution in the domain of attraction of a stable law $S_{\alpha,\beta,\sigma,\mu}$, we can estimate the asymmetry parameter β if we can estimate r and l, the nonnegative constants given by (18) and (19). If k is the number of order statistics used for the estimation of α in the right tail of the distribution, then r is estimated by $\frac{k}{n}Y_{k,n}^{\hat{\alpha}_n}$. The parameter l is estimated in a similar way, using the order statistics in the left tail of the distribution.

Mason (1982) proves that the Hill estimator is weakly consistent when the cutoff parameter $k = k(n) \to \infty$, $k/n \to 0$ and if and only if $F \in DA(\alpha)$. The condition $k = k(n) \to \infty$ implies that eventually infinitely many order statistics are involved, allowing for the use of the law of large numbers. The requirement $k/n \to 0$ means that the tail and nothing else is estimated. Hall (1982) gives the first result on the asymptotic normality of $\hat{\alpha}_n$. More general results under different sets of conditions on the normalizing constant a_n have been obtained by Davis and Resnick (1984) and Haeusler and Teugles (1985).

The conditions needed for the consistency of the Hill estimates do not offer much guidance on how k should be chosen. In practice it is more useful to use the method of Hall (1990b) or the method of Danielsson, de Haan, Peng, and de Vries (2001) by minimizing the bootstrapped asymptotic mean square error of the Hill estimate of α.

2.4 Rate of convergence of the parametric bootstrap

The rate of convergence of the parametric bootstrap is given by the rate of convergence of the joint characteristic function (31), which is the slower of φ_n, $\hat{\alpha}_n$ and $\hat{\beta}_n$. The rate of convergence of φ_n is n^{-1} as it can be seen from (27). The rate of convergence of $\hat{\alpha}_n$ and $\hat{\beta}_n$ depends on the choice of k which is intimately related to the tails of the distribution F. For example, suppose that the CDF F satisfies

$$1 - F(y) = ry^{-\alpha}\big(1 + d_r y^{-\delta_r} + o(y^{-\delta_r})\big) \quad \text{and} \quad F(-y) = l|y|^{-\alpha}\big(1 + d_l|y|^{-\delta_l} + o(|y|^{-\delta_l})\big) \tag{37}$$

as $y \to \infty$, where $\delta_r, \delta_l > 0$, d_r and d_l are real numbers and r and l are as defined in (18) and (19). Hall (1982) shows that under assumption (37), it is asymptotically optimal to take $k = o(n^{2\delta_r/(2\delta_r+\alpha)})$ (for the right tail) and $k = o(n^{2\delta_l/(2\delta_l+\alpha)})$ (for the left tail) . Then the rate of convergence of the estimators is

$$\hat{\alpha}_n - \alpha = O_p(n^{-\delta_r/(2\delta_r+\alpha)}), \quad \hat{r}_n - r = O_p(n^{-\delta_r/(2\delta_r+\alpha)}\log n, \quad \hat{l}_n - l = O_p(n^{-\delta_l/(2\delta_l+\alpha)}\log n.$$

From (20) it can be seen that the rate of convergence of $\hat{\beta}_n - \beta$ is the slower of $\hat{r}_n - r$ and $\hat{l}_n - l$.

Remark 3. The stable laws themselves satisfy (37) with $\delta_r = \delta_l = \alpha$. For Student's t distribution, $\delta_r = \delta_l = 2$. The t distribution is in the domain of attraction of the stable laws and has infinite variance if the number of degrees of freedom is smaller than or equal to 2.

Remark 4. The assumption (37) is more demanding than just requiring the distribution F to satisfy the general conditions (17). But Hall says that, if one relaxes it, then there does not seem a way to characterize the optimal k and to obtain an algebraic convergence rate for the tail index estimator. Other explicit assumptions about the tails of F are exploited in Haeusler and Teugles (1985) (p.752-754) and they lead to much slower rates of convergence for $\hat{\alpha}_n$ and $\hat{\beta}_n$. In general, a convergence rate of $n^{-1/2}$ cannot be achieved without parametric knowledge of F, as pointed out in Hall and Jing (1998).

Remark 5. The m out of n bootstrap and subsampling based on τ_n have an error of order $n^{-(\alpha-1)(2-\alpha)/\alpha}$ (Hall and Jing (1998)) which is larger than the error of the parametric bootstrap. Simulations in next section support this conclusion. They also show that the parametric bootstrap performs better than the m out of n bootstrap and subsampling based on T_n.

3 Simulation evidence

In this section we investigate the performance of the parametric bootstrap and we compare it with its main competitors: the m out of n bootstrap and subsampling. For these two methods, the choice of m is an important matter. If the bootstrap sample fails to satisfy the conditions $m/n \to 0$ or $m(\log\log n)/n \to 0$, the bootstrap distribution is random and the methods are invalid (as can be concluded from Hall (1990a) and Hall and Yao (2003)). In practice, m is usually estimated using a data-dependent method, rather than using different asymptotic arguments. In this paper we prefer the method of Bickel and Sakov (2008) since it is more suitable for P values. However, simulations not reported here indicate that the m out of n bootstrap and subsampling of τ_n and T_n, with data from the stable law with $\alpha = 1.5$ and $\beta = 0$, do not work well for any choice of m if the sample size is not as large as $2,000$. The subsampling of τ_n works better in this case, but it seems to be very sensitive to the choice of m.

Our simulation study is based on samples of size 100, 400 and 1000 from the t distribution with degrees of freedom 1.1, 1.5, 1.9 and the stable distribution with $\alpha = 1.1$, 1.5, 1.9, $\beta = 0$, 1 and scale $\sigma = 1$. For these distributions we take $a_n = n^{1/\alpha}$ in (5) (see discussion on p.2). We compare the following bootstrap methods

- parametric bootstrap of T_n. The parameters α and β are estimated by Hill's method as described in Section 2.3. The number of order statistics k is estimated using the method of Danielsson et al. (2001) for the stable law and the method of Hall for the t distribution.
- m out of n bootstrap and subsampling of T_n. The choice of m is done using the method of Bickel and Sakov (2008).
- m out of n bootstrap of τ_n. The tail index α is estimated using Hill's method and the number of order statistics k is estimated using the method of Danielsson et al. (2001). The choice of m is made by applying the method of Bickel and Sakov (2008).
- subsampling of τ_n. The tail index is estimated using the method of Bertail, Politis, and Romano (1999) as in Romano and Wolf (1999), while the choice of m is made by applying the method of Bickel and Sakov (2008).

All results are obtained from 10,000 replications of the statistics τ_n and T_n and $B = 399$ bootstrap repetitions. We consider the case in which the null hypothesis $\mu = 0$ is true and the case in which the alternative hypothesis $\mu = -0.5$ is true. The results based on the true null hypothesis are displayed as P value discrepancy plots. The best performance of the tests is achieved when the error in rejection probability (ERP) is close to zero. The results based on the true alternative are displayed as adjusted power functions, by taking into account the actual size of the tests under the true null.

In Figures 1 and 2 the null hypothesis $\mu = 0$ is true. The data were generated from the stable distribution with $\alpha = 1.1$ and $\beta = 0$ for sample sizes of 100 and 1000. It can be seen that the parametric bootstrap has the fastest rate of convergence. If we consider Figure 3, the power is not satisfactory for any bootstrap tests, except for the parametric bootstrap which has a slightly higher power. Figures not displayed here indicate that the same conclusions hold for the t distribution.

In Figures 4, 5, 6 and 7 the data were generated from the stable distribution with $\alpha = 1.5$, $\beta = 0$ and the t distribution with $\alpha = 1.5$ degrees of freedom for samples of size 100 and 400. In contrast to the m out of n bootstrap and subsampling, the parametric bootstrap performs very well for samples as small as 100. Moreover, as Figure 8 shows, the power of the parametric bootstrap is always higher than the power of the other bootstrap tests. Results not included here indicate that the same conclusion holds for samples smaller than the one considered in the figure and also for the stable distribution.

In Figures 9, 10 and 11 the data were generated from the stable distribution with $\alpha = 1.9$ and $\beta = 0$ and from the t distribution with $\alpha = 1.9$ degrees of freedom. It can be seen that the parametric bootstrap has a faster rate of convergence than the other bootstrap methods and performs very well. We did not include the power functions here, but the results are very satisfactory for all bootstrap methods with adjusted power close to one.

The next four figures refer to the case in which the data are heavily skewed, with $\beta = 1$. Figures 12 and 13 show that the m out of n bootstrap and subsampling of T_n work better but not the best if $\alpha = 1.5$. Results not included here reveal that when $\alpha = 1.1$, $\beta = 1$ and $n = 1{,}000$ the ERP of the bootstrap tests can be as high as 0.7 with the m out of n bootstrap and subsampling of T_n having an ERP close to 0.4. In this extreme case, all the tests lack power, as can be seen from Figure 14. The power increases for $\alpha = 1.5$ with the parametric bootstrap having the highest power, as shown in Figure 15. In general, the power is influenced by α and β: the smaller α is and the closer β is to 1 or -1, the lower the power. The highest power is achieved when the distribution is not far from having a finite variance, namely when $\alpha = 1.9$ and $\beta = 0$.

In conclusion, the figures indicate that the parametric bootstrap works better than the m out of n bootstrap and subsampling when α is not close to 1 and β is not close to 1 or -1.

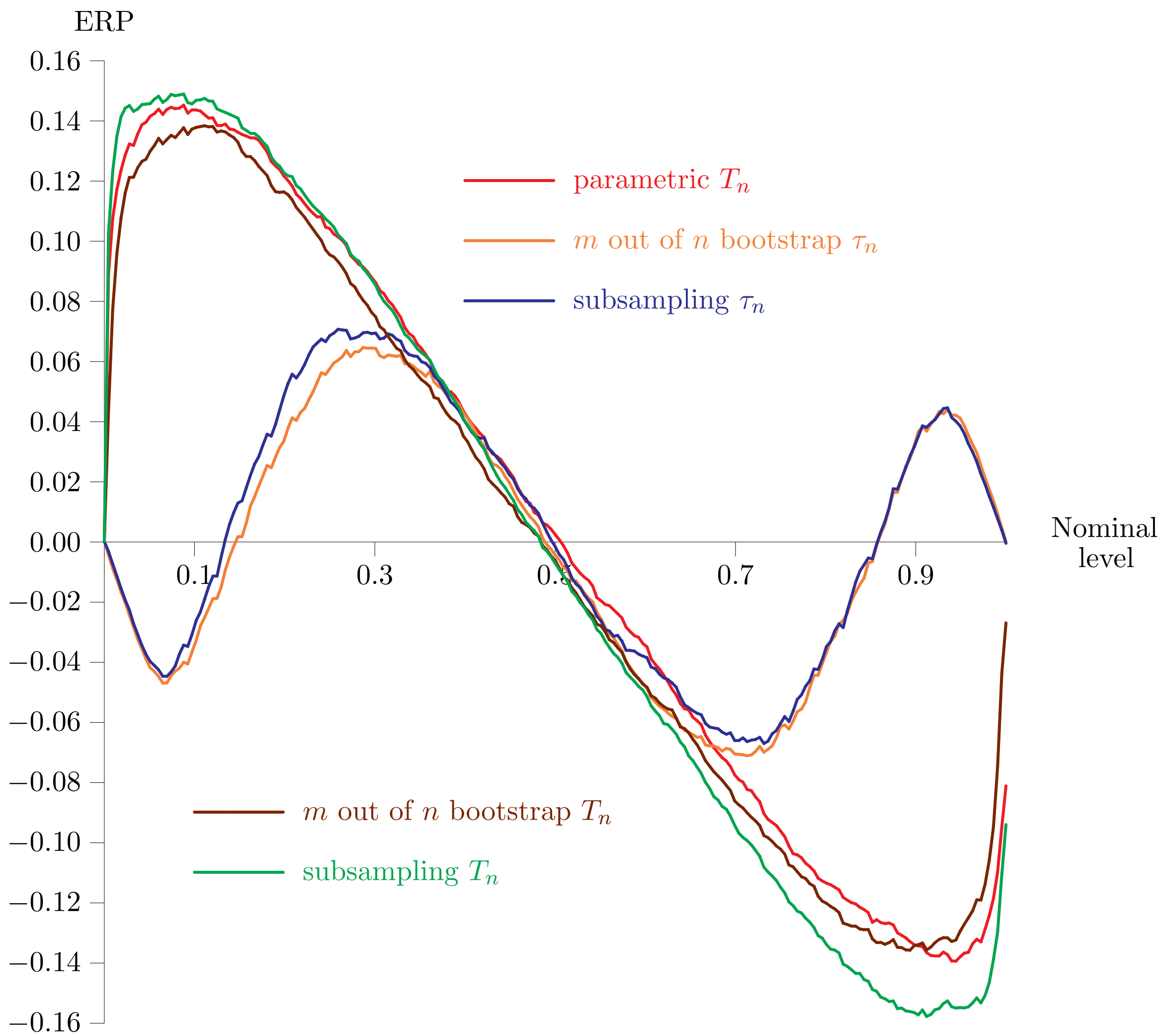

Figure 1: P value discrepancy plots; parametric bootstrap, m out of n bootstrap, subsampling of T_n and τ_n; data from stable law with $\alpha = 1.1$, $\beta = 0$, $n = 100$; m chosen using Bickel and Sakov's method; k chosen using Danielsson's method

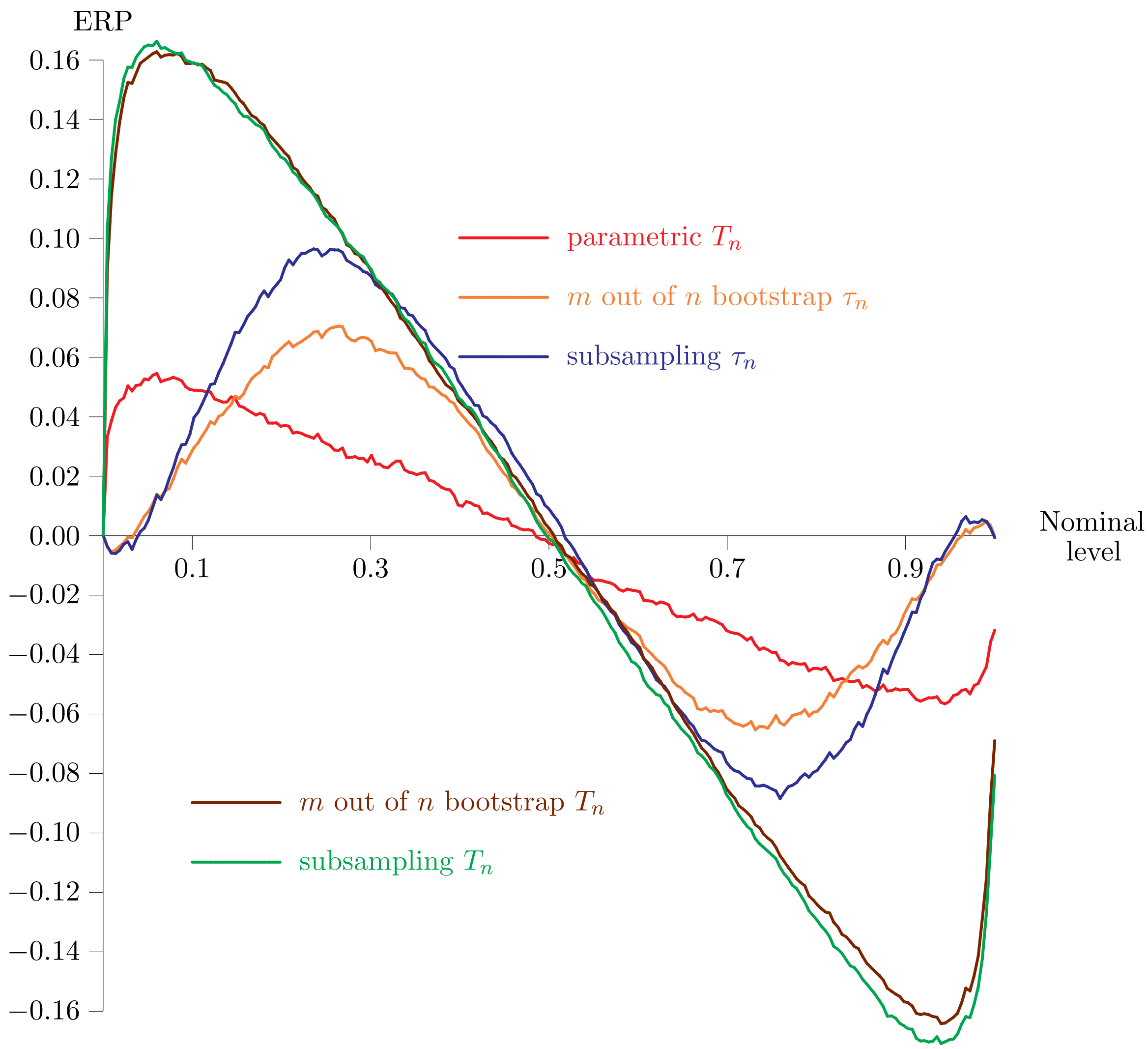

Figure 2: P value discrepancy plots; parametric bootstrap, m out of n bootstrap, subsampling of T_n and τ_n; data from stable law with $\alpha = 1.1$, $\beta = 0$, $n = 1000$; m chosen using Bickel and Sakov's method; k chosen using Danielsson's method

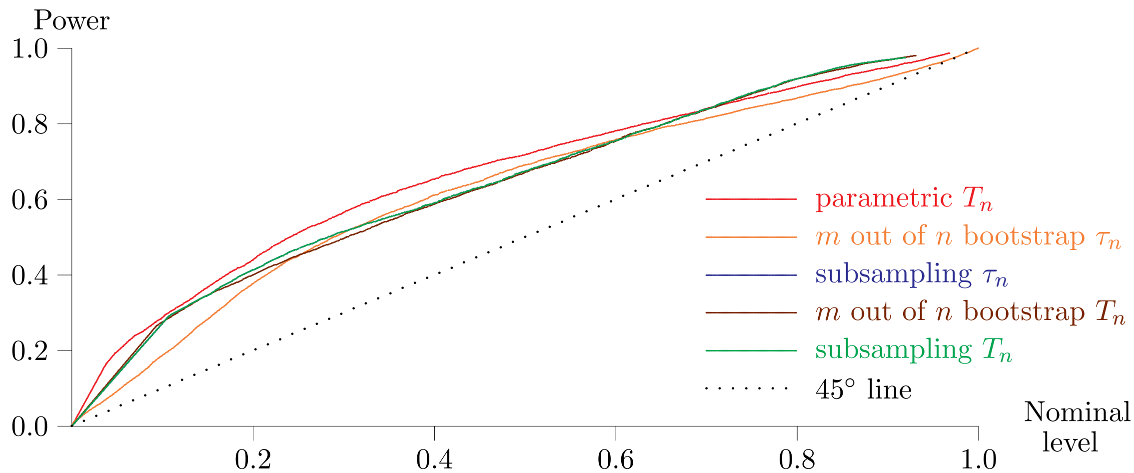

Figure 3: Power; parametric bootstrap, m out of n bootstrap, subsampling of T_n and τ_n; data from stable law with $\alpha = 1.1$, $\beta = 0$, $n = 1000$; m chosen using Bickel and Sakov's method; k chosen using Danielsson's method

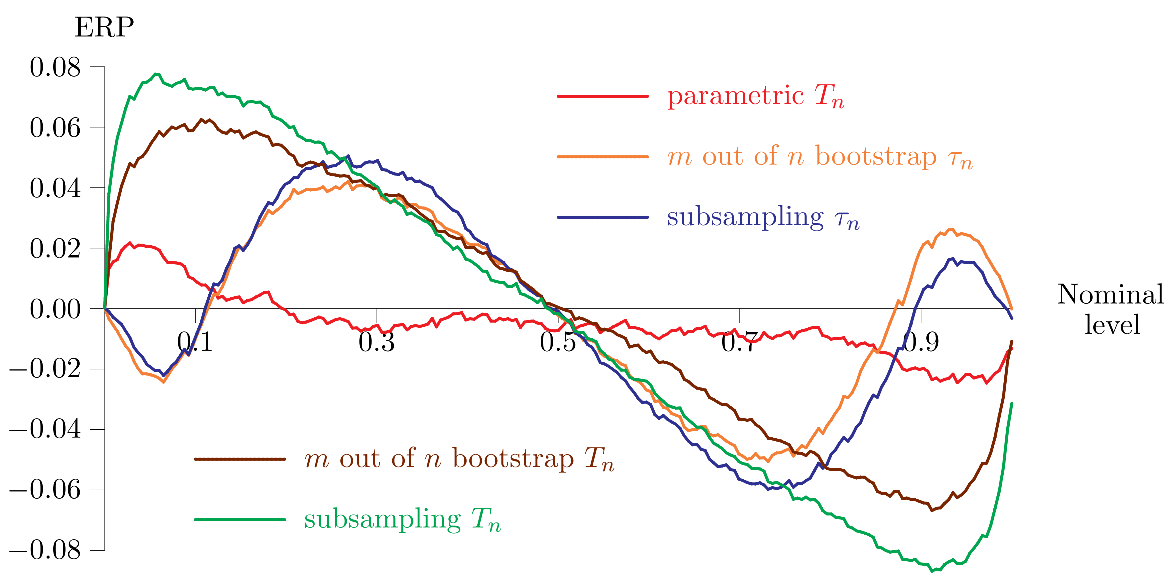

Figure 4: P value discrepancy plots; parametric bootstrap, m out of n bootstrap, subsampling of T_n and τ_n; data from stable law with $\alpha = 1.5$, $\beta = 0$, $n = 100$; m chosen using Bickel and Sakov's method; k chosen using Danielsson's method

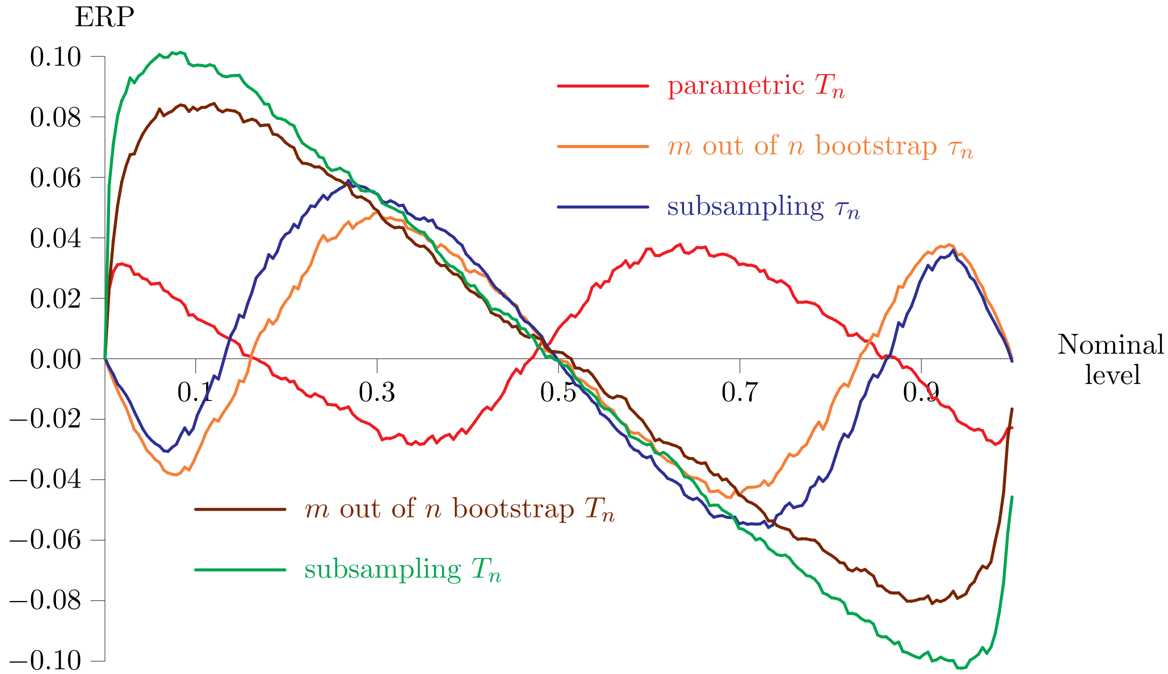

Figure 5: P value discrepancy plots; parametric bootstrap, m out of n bootstrap, subsampling of T_n and τ_n; data from the t distribution with $\alpha = 1.5$ degrees of freedom, $n = 100$; m chosen using Bickel and Sakov's method; k chosen using Hall's method for the parametric bootstrap and Danielsson's method for the m out of n bootstrap and subsampling

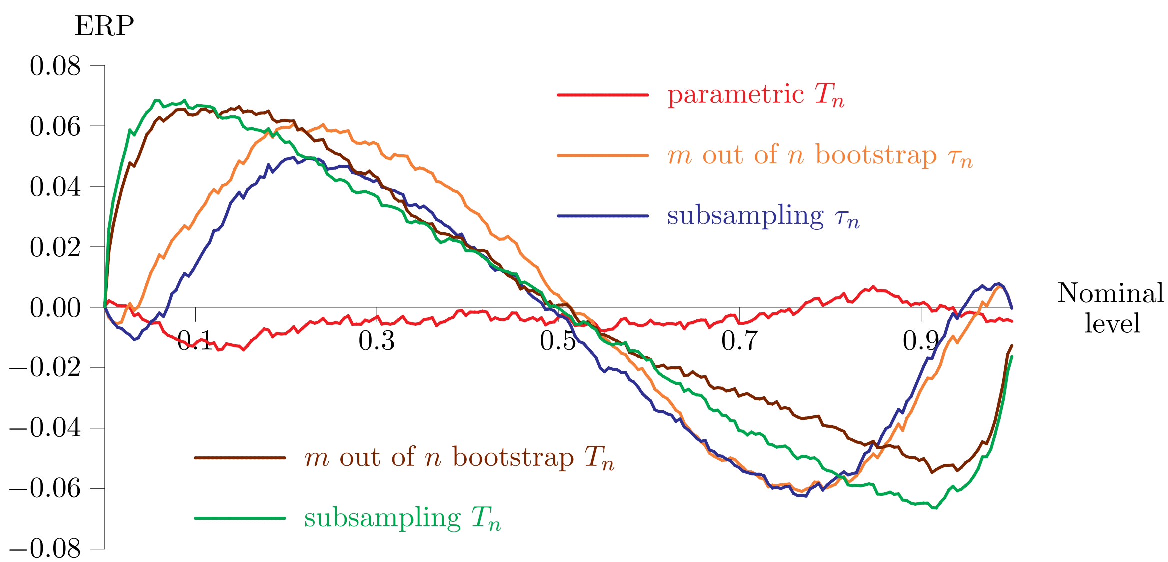

Figure 6: P value discrepancy plots; parametric bootstrap, m out of n bootstrap, subsampling of T_n and τ_n; data from stable law with $\alpha = 1.5$, $\beta = 0$, $n = 400$; m chosen using Bickel and Sakov's method; k chosen using Danielsson's method

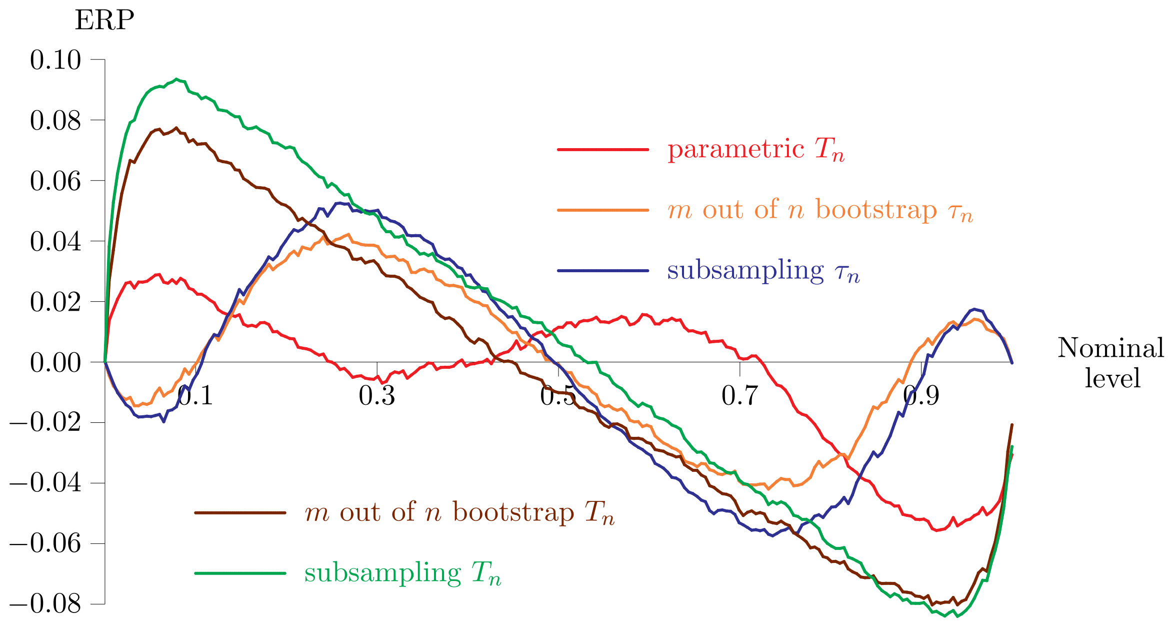

Figure 7: P value discrepancy plots; parametric bootstrap, m out of n bootstrap, subsampling of T_n and τ_n; data from the t distribution with $\alpha = 1.5$ degrees of freedom, $n = 400$; m chosen using Bickel and Sakov's method; k chosen using Hall's method for the parametric bootstrap and Danielsson's method for the m out of n bootstrap and subsampling

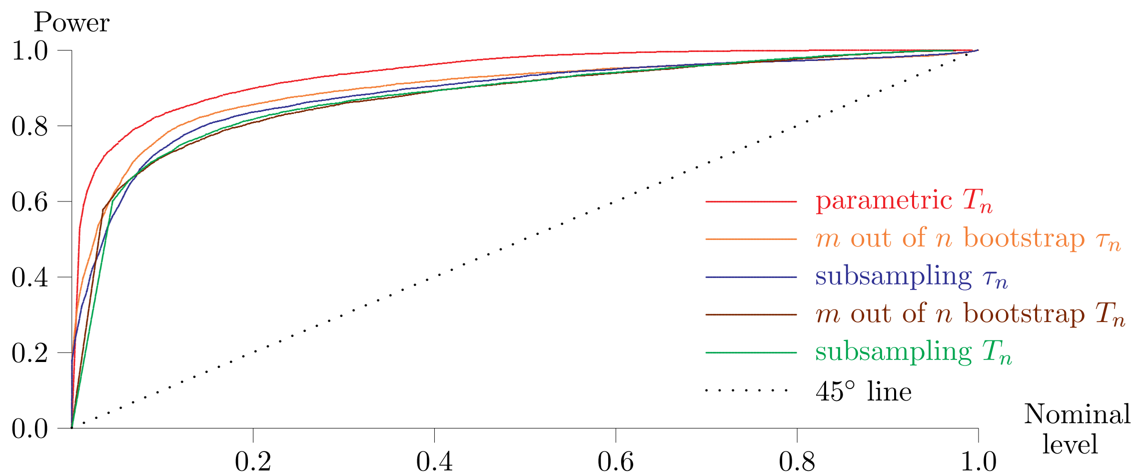

Figure 8: Power; parametric bootstrap, m out of n bootstrap, subsampling of T_n and τ_n; data from the t distribution with $\alpha = 1.5$ degrees of freedom, $n = 1000$; m chosen using Bickel and Sakov's method; k chosen using Danielsson's method

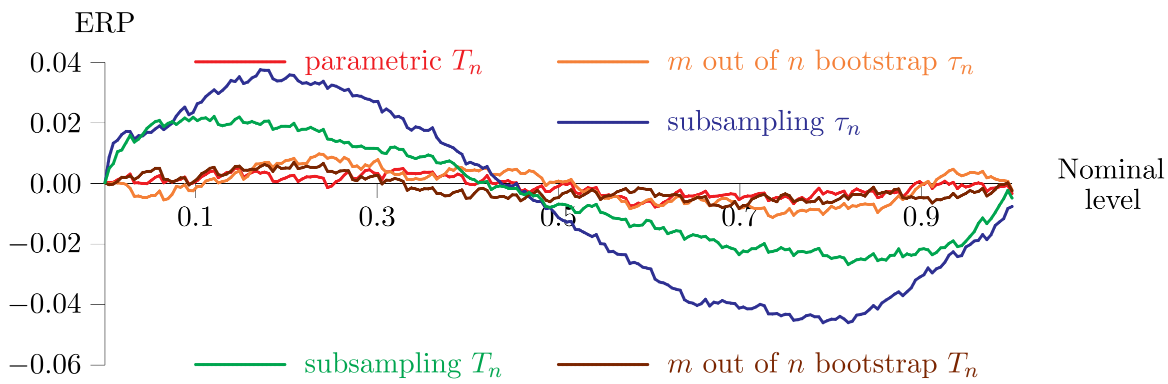

Figure 9: P value discrepancy plots; parametric bootstrap, m out of n bootstrap, subsampling of T_n and τ_n; data from stable law with $\alpha = 1.9$, $\beta = 0$, $n = 100$; m chosen using Bickel and Sakov's method; k chosen using Danielsson's method

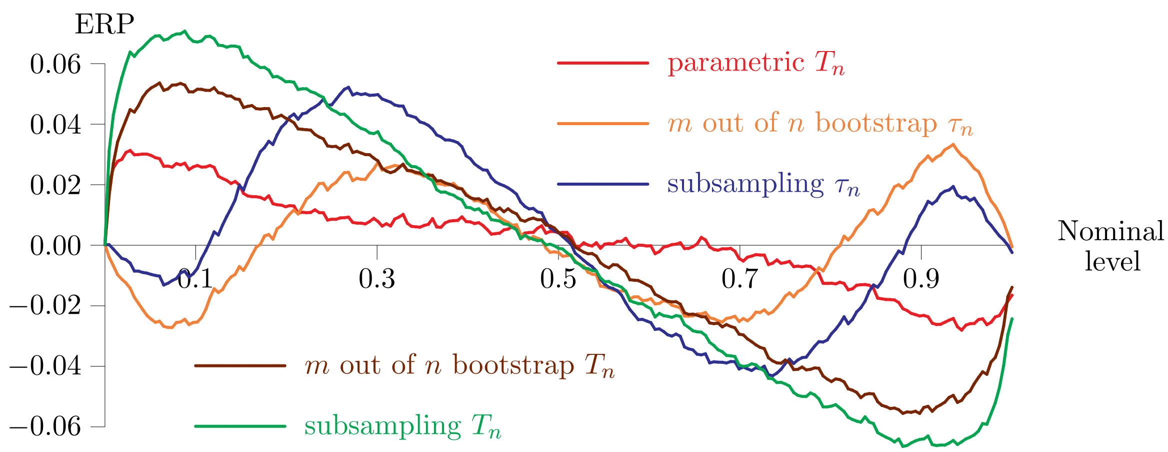

Figure 10: P value discrepancy plots; parametric bootstrap, m out of n bootstrap, subsampling of T_n and τ_n; data from the t distribution with $\alpha = 1.9$ degrees of freedom, $n = 100$; m chosen using Bickel and Sakov's method; k chosen using Danielsson's method

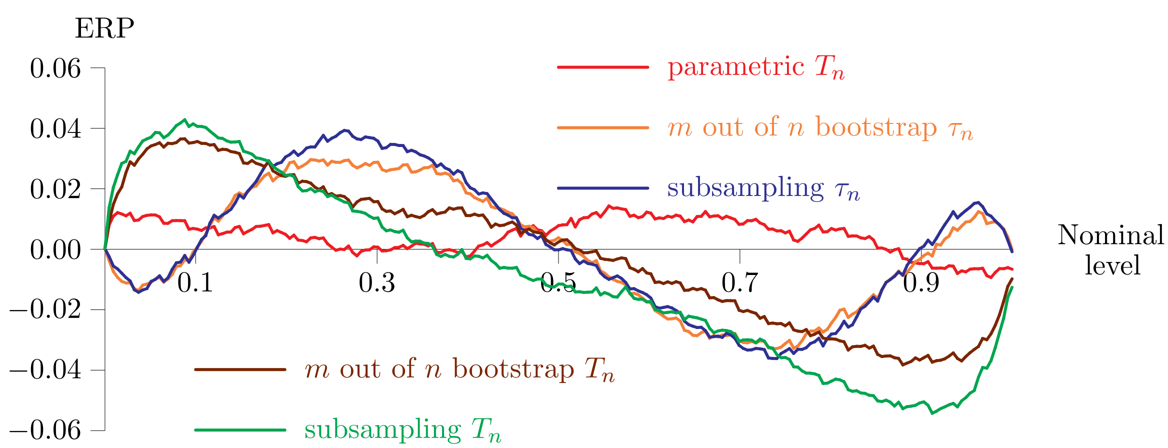

Figure 11: P value discrepancy plots; parametric bootstrap, m out of n bootstrap, subsampling of T_n and τ_n; data from the t distribution with $\alpha = 1.9$ degrees of freedom, $n = 400$; m chosen using Bickel and Sakov's method; k chosen using Danielsson's method

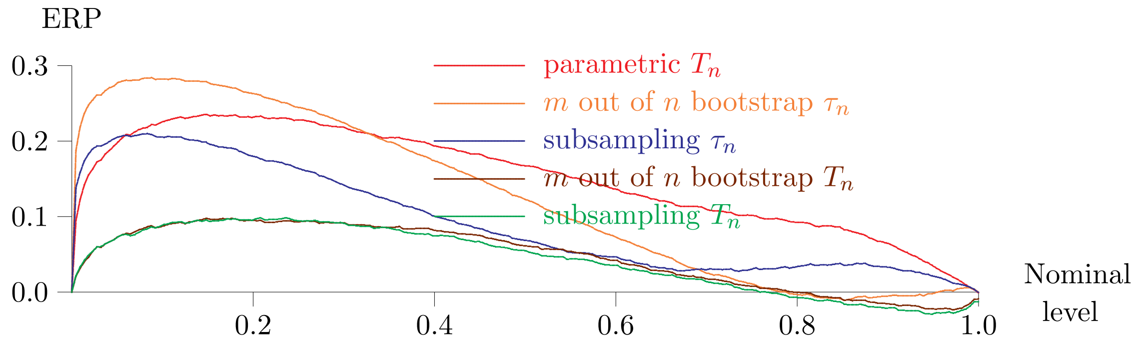

Figure 12: P value discrepancy plots; parametric bootstrap, m out of n bootstrap, subsampling of T_n and τ_n; data from stable law with $\alpha = 1.5$, $\beta = 1$, $n = 1000$; m chosen using Bickel and Sakov's method; k chosen using Danielsson's method

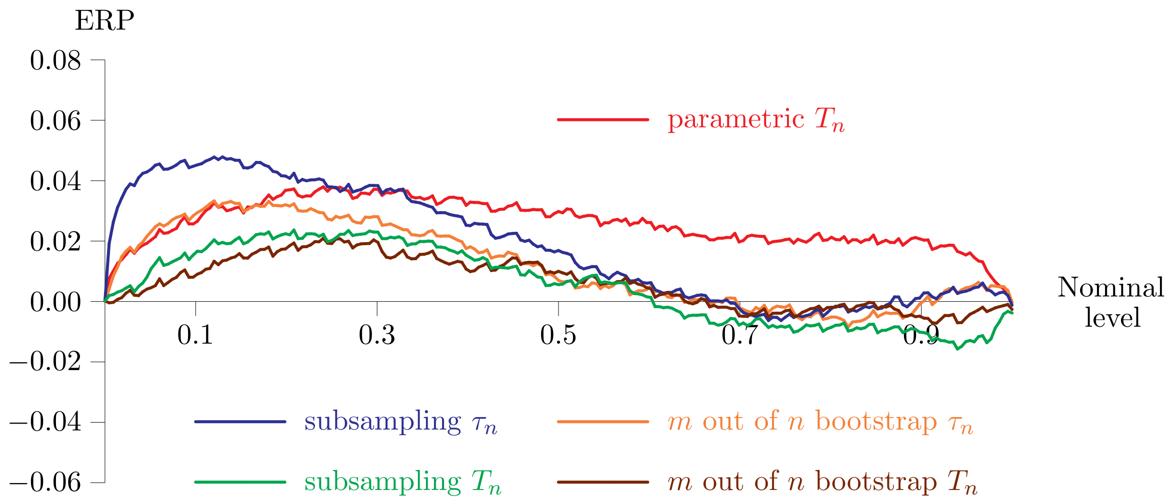

Figure 13: P value discrepancy plots; parametric bootstrap, m out of n bootstrap, subsampling of T_n and τ_n; data from stable law with $\alpha = 1.9$, $\beta = 1$, $n = 1000$; m chosen using Bickel and Sakov's method; k chosen using Danielsson's method

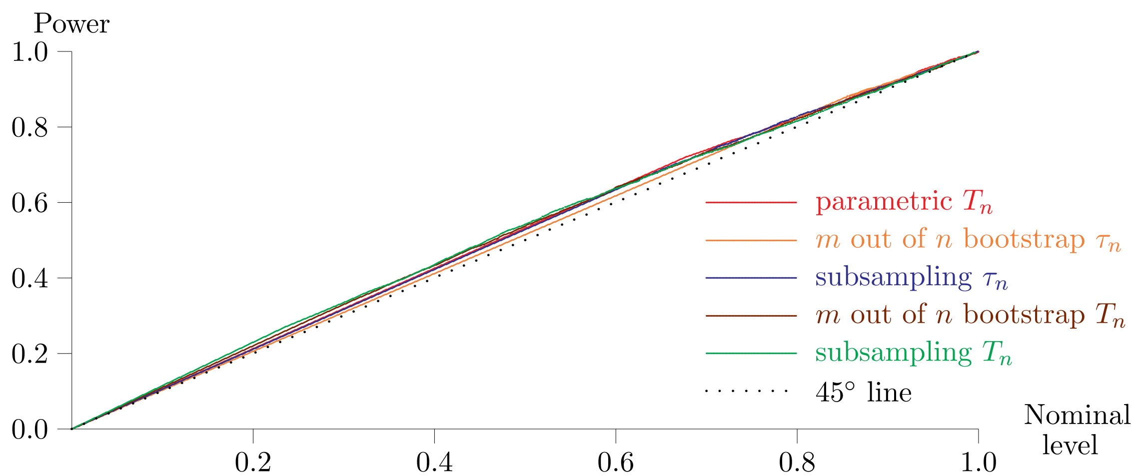

Figure 14: Power; parametric bootstrap, m out of n bootstrap, subsampling of T_n and τ_n; data from stable law with $\alpha = 1.1$, $\beta = 1$, $n = 1000$; m chosen using Bickel and Sakov's method; k chosen using Danielsson's method

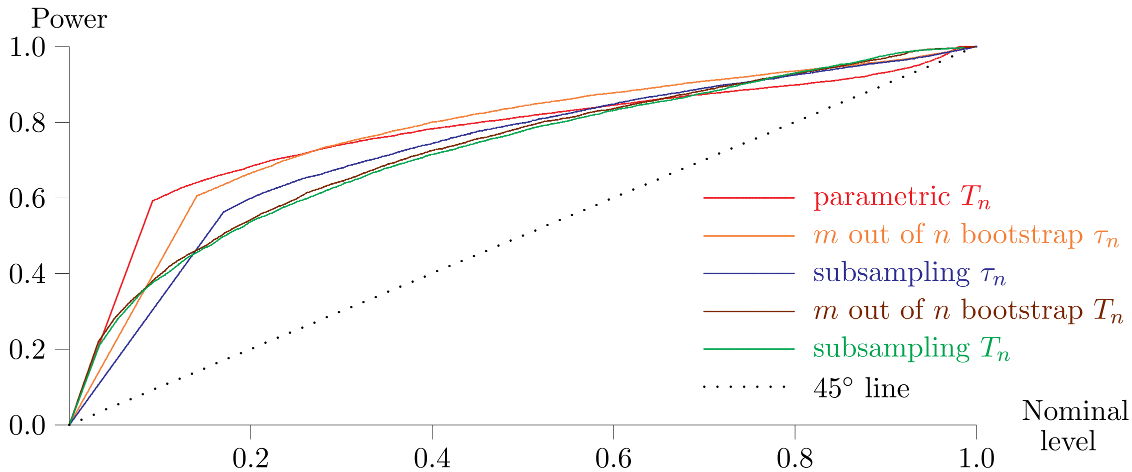

Figure 15: Power; parametric bootstrap, m out of n bootstrap, subsampling of T_n and τ_n; data from stable law with $\alpha = 1.5$, $\beta = 1$, $n = 100$; m chosen using Bickel and Sakov's method; k chosen using Danielsson's method

4 Conclusion

In this paper, we have proposed a parametric bootstrap for the purposes of inference on the expectation of a heavy-tailed distribution when an independent and identically distributed sample generated by that distribution is available. The parametric bootstrap is based on a central-limit argument for self-normalised sums. The bootstrap distribution can be estimated consistently if we can estimate the parameters α and β of the stable law to which the centred and normalized sum of the observations converges. This is most conveniently carried out by simulation, rather than by use of the asymptotic distribution, which, although known, is expressed in terms of integrals of functions of parabolic cylinder functions, and is thus awkward to compute. Our results show that, as long as estimation of α and β is reasonably precise, the parametric bootstrap gives inference with a sample size of 100 that is reliable by any usual standard. Its performance deteriorates when the methods we use to estimate these parameters become imprecise, which happens when the expectation is close to nonexistence, and when the distribution is heavily skewed. We conjecture that it is impossible to devise a reliable method of inference for α close to 1, but it may be possible to find better estimators of β.

Moreover, the parametric bootstrap is a better alternative to the asymptotic test based upon the stable distributions, since it requires the estimation of a smaller number of nuisance parameters under the null hypothesis.

Finally, the parametric bootstrap performs better than its main competitors: subsampling and the m out of n bootstrap, as clearly indicated by our simulations.

5 References

Arcones, M. A. and E. Gine (1989). The bootstrap of the mean with arbitrary bootstrap sample size. *Annales de l'Institut Henri Poincaré 25*(4), 457–481.

Athreya, K. B. (1985). Bootstrap for the mean in the infinite variance case. ii. *Technical Report, Dept. Of Statistics, Iowa State Univ.* (86-21).

Athreya, K. B. (1987). Bootstrap of the mean in the infinite variance case. *Annals of Statistics 15*, 724–731.

Bahadur, R. R. and L. J. Savage (1956). The nonexistence of certain statistical procedures in nonparametric problems. *The Annals of Statistics 27*, 1115–1122.

Beran, R. (1988). Prepivoting test statistics: a bootstrap view of asymptotic refinements. *Journal of the American Statistical Association 83*, 307–27.

Bertail, P., D. N. Politis, and J. P. Romano (1999). On subsampling estimators with unknown rate of convergence. *Journal of the American Statistical Association 94*, 569–579.

Bickel, P. J. and D. A. Freedman (1981). Some asymptotic theory for the bootstrap. *The Annals of Statistics 9*(6), 1196–1217.

Bickel, P. J., F. Gotze, and W. R. van Zwet (1997). Resampling fewer than n observations: gains, loses and remedies for losses. *Statistica Sinica 7*, 1–31.

Bickel, P. J. and A. Sakov (2008). On the choice of m in the m out of n bootstrap and confidence bounds for extrema. *Statistica Sinica 18*, 967–985.

Bingham, N. H., C. M. Goldie, and J. Teugles (1989). *Regular Variation.* Cambridge: Cambridge University Press.

Chambers, J. M., C. Mallows, and C. Stuck (1981). A method for simulating stable random variables. *Journal of the American Statistical Association 71.*

Danielsson, J., L. de Haan, L. Peng, and C. G. de Vries (2001). Using a bootstrap method to choose the sample fraction in tail index estimation. *Journal of Multivariate Analysis 76*, 226–248.

Davis, R. and S. Resnick (1984). Tail estimates motivated by extreme value theory. *Ann. Statist. 12*, 1467–1487.

Efron, B. (1969). Student's t-test under symmetry conditions. *J. Amer. Statist. Assoc. 64*, 1278–1302.

Efron, B. (1979). Bootstrap methods: another look at the jackknife. *Annals of Statistics 7*, 1–26.

Embrechts, P., C. Kluppelberg, and T. Mikosch (1997). *Modelling Extremal Events for Insurance and Finance.* Berlin: Springer-Verlag.

Feller, W. (1971). *An Introduction to Probability Theory and Its Applications vol. II*. New York: John Wiley and Sons.

Gine, E., F. Gotze, and D. M. Mason (1997). When is the student t-statistic asymptotically normal? *Ann. Probab. 25*, 1514–1531.

Gnedenko, B. V. and A. N. Kolmogorov (1954). *Limit Distributions for Sums of Independent Random Variables*. Cambridge, MA: Addison-Wesley Publishing.

Haeusler, E. and J. L. Teugles (1985). On asymptotic normality of hill's estimator for the exponent of regular variation. *Ann. Statist. 13*, 743–756.

Hall, P. (1982). On some simple estimates of an exponent of regular variation. *Journal of the Royal Statistical Society, series B 44*, 37–42.

Hall, P. (1990a). Asymptotic properties of the bootstrap for heavy-tailed distributions. *Annals of probability 18*(2), 1342–1360.

Hall, P. (1990b). Using the bootstrap to estimate mean squared error and select smoothing parameters in nonparametric problems. *Journal of Multivariate Analysis 32*, 177–203.

Hall, P. and B. Jing (1998). Comparison of the bootstrap and asymptotic approximations to the distribution of a heavy-tailed mean. *Statistica Sinica 8*, 887–906.

Hall, P. and Q. Yao (2003). Inference in arch and garch models with heavy-tailed errors. *Econometrica 71(1)*, 285–317.

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

Ibragimov, R. (2007). Efficiency of linear estimators under heavy-tailedness: convolutions of α-symmetric distributions. *Econometric Theory 23*, 501–517.

Ibragimov, R. (2011). On the robustness of location estimators in models of firm growth under heavy-tailedness. *forthcoming Journal of Econometrics*.

Knight, K. (1989). On the bootstrap of the sample mean in the infinite variance case. *The Annals of Statistics 17*, 1168–1175.

Lévy, P. (1925). *Calcul des Probabilités*. Paris: Gauthier-Villars.

Logan, B. F., C. Mallows, S. Rice, and L. A. Shepp (1973). Limit distributions of self-normalized sums. *The Annals of Probability 11*, 788–809.

Mandelbrot, B. (1963). The variation of other speculative prices. *Journal of business 40*, 393–413.

Mason, D. M. (1982). Laws of large numbers for sums of extreme values. *Annals of Probability 10*(3), 754–762.

Mittnik, S. and S. Rachev (2000). *Stable Paretian Models in Finance*. New-York: John Wiley.

Mittnik, S., S. Rachev, and J.-R. Kim (1998). Chi-square-type distributions for heavy-tailed variates. *Econometric Theory 14*, 339–354.

Politis, D. N., J. P. Romano, and M. Wolf (1999). *Subsampling.* New-York: Springer.

Romano, J. P. and M. Wolf (1999). Subsampling inference for the mean in the heavy-tailed case. *Metrika 50*, 55–69.

Samorodnitsky, G. and M. Taqqu (1994). *Stable non-gaussian random processes.* New-York: Chapman and Hall.

Serfling, R. J. (1980). *Approximation Theorems of Mathematical Statistics.* New-York: Wiley.

Shao, J. and D. Tu (1995). *The Jackknife and Bootstrap.* New-York: Springer-Verlag.

Zolotarev, V. M. (1986). *One-dimensional Stable Distributions.* American Mathematical Society: Vol.65 of Translations of mathematical monographs.