

# A Parametric Bootstrap for Heavy-Tailed Distributions

by

**Adriana Cornea**

Tanaka Business School  
Imperial College London  
South Kensington campus  
London SW7 2AZ UK  
**a.cornea@imperial.ac.uk**

and

**Russell Davidson**

Department of Economics and CIREQ  
McGill University  
Montréal, Québec, Canada  
H3A 2T7

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

**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. A parametric bootstrap is derived by considering the distribution of the bootstrap  $P$  value instead of that of the bootstrap statistic. 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, infinite variance

JEL codes: C10, C12, C15

This research was supported by 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. The paper is based on some chapters of the thesis of the first author. We are indebted to Keith Knight for valuable comments on an earlier version.

August 2009

## 1. Introduction

Let  $F$  be the distribution of the independent and identically distributed random variables  $Y_1, \dots, Y_n$ . We are interested in inference on the parameter  $\delta$  in the location model

$$Y_j = \delta + U_j, \quad \mathbb{E}(U_j) = 0, \quad j = 1, \dots, n. \quad (1)$$

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  $U_j$  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  $\alpha$  greater than 1 and smaller than 2. A distribution  $F$  is said to be in the domain of attraction of a stable law with  $\alpha < 2$ , if centred 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 centred and normalized 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  $\alpha$  ( $0 < \alpha \leq 2$ ), the skewness parameter  $\beta$  ( $-1 < \beta < 1$ ), the scale parameter  $c$  ( $c > 0$ ), and the location parameter  $\delta$ . A stable random variable  $X$  can be written as  $X = \delta + cZ$ , where the location parameter of  $Z$  is zero, and its scale parameter unity. We write the distribution of  $Z$  as  $S(\alpha, \beta)$ . When  $0 < \alpha < 2$ , all the moments of  $X$  of order greater than  $\alpha$  do not exist. When  $1 < \alpha \leq 2$ , the parameter  $\delta$  in model (1) can be consistently estimated by the sample mean.

Since there has been some confusion in the literature occasioned by the existence of more than one parametrisation of the stable laws, we specify here that the characteristic function of what we have called the  $S(\alpha, \beta)$  distribution is

$$\mathbb{E}(\exp(itY)) = \exp(-|t|^\alpha [1 - i\beta \tan(\pi\alpha/2)(\text{sign } t)]).$$

In simulation exercises, we generate realisations of this distribution using the algorithm proposed by Chambers, Mallows, and Stuck (1976), their formula modified somewhat to take account of their use of a different parametrisation. Specifically, a drawing from the  $S(\alpha, \beta)$  distribution is given by

$$(1 + \beta^2 \tan^2(\pi\alpha/2))^{1/2\alpha} \frac{\sin(\alpha(U + b(\alpha, \beta)))}{(\cos U)^{1/\alpha}} \left( \frac{\cos(U - \alpha(U + b(\alpha, \beta)))}{W} \right)^{(1-\alpha)/\alpha}$$

where  $U$  is uniformly distributed on  $[-\pi/2, \pi/2]$ ,  $W$  is exponentially distributed with expectation 1, and  $b(\alpha, \beta) = \tan^{-1}(\beta \tan(\pi\alpha/2))/\alpha$ .

It was shown by Athreya (1987) that, when the variance does not exist, the conventional resampling bootstrap of Efron (1979) is not valid, because the bootstrap distribution of the sample mean does not converge as the sample size  $n \rightarrow \infty$  to a deterministic distribution. 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).

A proposed remedy for failure of the conventional bootstrap is the  $m$  out of  $n$  bootstrap; see Arcones and Giné (1989). It is based on the same principle as Efron’s bootstrap, but the bootstrap sample size is  $m$ , smaller than  $n$ . If  $m/n \rightarrow 0$  as  $n \rightarrow \infty$ , this bootstrap is consistent. However, as we will see in simulation experiments, the  $m$  out of  $n$  bootstrap fails to provide reliable inference if the sample size is not very large. 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, and performs somewhat better than the  $m$  out of  $n$  bootstrap.

In this paper, we introduce a parametric bootstrap method that overcomes the failure of bootstrap tests based on resampling, the  $m$  out of  $n$  bootstrap, or subsampling, for the parameter  $\delta$  of the model (1). The method is based on the fact that the distribution of the bootstrap  $P$  value, unlike the distribution of the bootstrap statistic, turns out to have a nonrandom limiting distribution as  $n \rightarrow \infty$ .

## 2. The bootstrap $P$ value

Suppose we wish to test the hypothesis  $\delta = 0$  in model (1). A possible test statistic is

$$\tau = n^{-1/\alpha} \sum_{j=1}^n Y_j. \quad (2)$$

Let the  $Y_j$  be IID realisations of a law in the domain of attraction of the law of  $Y \equiv cZ$ , where  $Z \sim S(\alpha, \beta)$ , and  $1 < \alpha < 2$ . Restricting  $\alpha$  to be greater than 1 ensures that the expectation of  $Y$  exists.

By the Generalized Central Limit Theorem, the asymptotic distribution of  $\tau$  is the stable distribution  $cS(\alpha, \beta)$ . If  $\alpha$ ,  $c$ , and  $\beta$  are known, then we can perform asymptotic inference by comparing the realization of the statistic  $\tau$  with a quantile of the stable distribution  $cS(\alpha, \beta)$ . The asymptotic  $P$  value for a test that rejects in the left tail of the distribution is

$$P = cS(\alpha, \beta)(\tau).$$

Unless the  $Y_i$  actually follow the stable distribution, inference based on this  $P$  value may be unreliable in finite samples.

Now suppose that, despite Athreya and Knight, we bootstrap the statistic  $\tau$  using the conventional resampling bootstrap. This means that, for each bootstrap sample  $Y_1^*, \dots, Y_n^*$ , a bootstrap statistic is computed as

$$\tau^* = n^{-1/\alpha} \sum_{j=1}^n (Y_j^* - \bar{Y}). \quad (3)$$

where  $\bar{Y} = \sum_{j=1}^n Y_j$  is the sample mean. The  $Y_j^*$  are centred using  $\bar{Y}$  because we wish to use the bootstrap to estimate the distribution of the statistic *under the null*, and the sample mean, not 0, is the true mean of the bootstrap distribution. The bootstrap  $P$  value is the fraction of the bootstrap statistics more extreme than  $\tau$ . For ease of exposition, we suppose that “more extreme” means “less than”. Then the bootstrap  $P$  value is

$$P_B^* = \frac{1}{B} \sum_{j=1}^B \mathbf{I}(\tau_j^* < \tau).$$

Note that the presence of the (asymptotic) normalising factor of  $n^{-1/\alpha}$  is no more than cosmetic.

As  $B \rightarrow \infty$ , by the strong law of large numbers, the bootstrap  $P$  value converges almost surely, conditional on the original data, to the random variable

$$p(\mathbf{Y}) = \mathbf{E}^*(\mathbf{I}(\tau^* < \tau)) = \mathbf{E}(\mathbf{I}(\tau^* < \tau) \mid \mathbf{Y}), \quad (4)$$

where  $\mathbf{Y}$  denotes the vector of the  $Y_j$ , and  $\mathbf{E}^*$  denotes an expectation under the bootstrap DGP, that is, conditional on  $\mathbf{Y}$ .  $p(\mathbf{Y})$  is a well-defined random variable, as it is a deterministic measurable function of the data vector  $\mathbf{Y}$ , with a distribution determined by that of  $\mathbf{Y}$ . We will see that as  $n \rightarrow \infty$  this distribution tends to a nonrandom limit.

For convenience in what follows, we let  $\gamma = 1/\alpha$ . Knight (1989) shows that, conditionally on the original data, the bootstrap statistic  $\tau^*$  has the same distribution (in the limit when  $B \rightarrow \infty$ ) as

$$\tau(W) = n^{-\gamma} \sum_{j=1}^n (Y_j W_j - \bar{Y}) = n^{-\gamma} \sum_{j=1}^n Y_j (W_j - 1) = n^{-\gamma} \sum_{j=1}^n (Y_j - \bar{Y})(W_j - 1),$$

where  $W_1, \dots, W_n$  is a multinomial vector with  $n$  trials and each cell probability is  $1/n$ . The last equality follows because  $\sum_j (W_j - 1) = 0$  identically. For large  $n$ , the multinomial vector has approximately the same distribution as a vector of  $n$  independent Poisson random variables  $M_1, \dots, M_n$  with expectation one. Thus, if we make the definition

$$\tau(M) = n^{-\gamma} \sum_{j=1}^n (Y_j - \bar{Y})(M_j - 1),$$

then  $\tau(W) \xrightarrow{d} \tau(M)$  as  $n \rightarrow \infty$  conditionally on  $Y_1, \dots, Y_n$ . We have the following result.

### Proposition 1

Conditionally on the  $Y_j$ , the random variable  $\tau(M)$  has a distribution of which the cumulant-generating function is

$$\sum_{j=1}^n \{ \exp(itn^{-\gamma}(Y_j - \bar{Y})) - 1 \} \quad (5)$$

as a function of  $t$ . The variance of this distribution is  $n^{-2\gamma} \sum_{j=1}^n (Y_j - \bar{Y})^2$ , and its expectation is zero.

The conditional characteristic function of  $\tau(M)$ , as a function of  $t$ , is

$$\begin{aligned} \mathbb{E}^* \{ \exp(it\tau(M)) \} &= \mathbb{E}^* \left[ \exp \left\{ it \sum_{j=1}^n n^{-\gamma}(Y_j - \bar{Y})(M_j - 1) \right\} \right] \\ &= \prod_{j=1}^n \mathbb{E}^* \left[ \exp \{ itn^{-\gamma}(Y_j - \bar{Y})(M_j - 1) \} \right] \end{aligned}$$

We have  $\mathbb{E} \{ \exp(it(M - 1)) \} = \exp(e^{it} - 1 - it)$  for a Poisson variable  $M$  with expectation 1. Thus the above characteristic function is

$$\exp \left[ \sum_{j=1}^n \{ \exp(itn^{-\gamma}(Y_j - \bar{Y})) - 1 - itn^{-\gamma}(Y_j - \bar{Y}) \} \right]$$

The cumulant-generating function (cgf) is the logarithm of the characteristic function, and, since  $\sum_{j=1}^n (Y_j - \bar{Y}) = 0$  by construction, it is equal to (5).

The  $r$ th cumulant of  $\tau(M)$  is the coefficient of  $(it)^r/r!$  in the Taylor expansion of (5) about  $t = 0$ . It is easy to check that the first two moments are as given. ■

**Remark 1:** The function (5) is random, because it depends on the  $Y_j$ .

### Corollary

The distribution of the self-normalised sum

$$t(M) \equiv \frac{n^\gamma \tau(M)}{(\sum_{j=1}^n (Y_j - \bar{Y})^2)^{1/2}} = \frac{\sum_{j=1}^n (Y_j - \bar{Y})(M_j - 1)}{(\sum_{j=1}^n (Y_j - \bar{Y})^2)^{1/2}} \quad (6)$$

has expectation 0 and variance 1 conditional on  $\mathbf{Y}$ , and so also unconditionally.

Let  $F_{\mathbf{Y}}^n$  denote the random CDF of  $t(M)$ . Then, from (4) with  $\tau^*$  replaced by  $\tau(M)$ , we have

$$\begin{aligned} p(\mathbf{Y}) &= \mathbb{E}^* \left( \mathbb{I}(\tau(M) < \tau) \right) = \mathbb{E}^* \left( \mathbb{I} \left( n^{-\gamma} \sum_{j=1}^n (Y_j - \bar{Y})(M_j - 1) < n^{-\gamma} \sum_{j=1}^n Y_j \right) \right) \\ &= \mathbb{E}^* \left[ \mathbb{I} \left( \frac{\sum_{j=1}^n (Y_j - \bar{Y})(M_j - 1)}{(\sum_{j=1}^n (Y_j - \bar{Y})^2)^{1/2}} < \frac{\sum_{j=1}^n Y_j}{(\sum_{j=1}^n (Y_j - \bar{Y})^2)^{1/2}} \right) \right] \\ &= \mathbb{E}^* \left[ \mathbb{I} \left( t(M) < \frac{\sum_{j=1}^n Y_j}{(\sum_{j=1}^n (Y_j - \bar{Y})^2)^{1/2}} \right) \right] = F_{\mathbf{Y}}^n \left( \frac{\sum_{j=1}^n Y_j}{(\sum_{j=1}^n (Y_j - \bar{Y})^2)^{1/2}} \right). \quad (7) \end{aligned}$$

### 3. Asymptotics

The principal questions that asymptotic theory is called on to answer in the context of bootstrapping the mean are:

- (i) Does the distribution with cgf (5) have a nonrandom limit as  $n \rightarrow \infty$ ? and
- (ii) Does the distribution of the bootstrap  $P$  value  $p(\mathbf{Y})$  have a well-defined limit as  $n \rightarrow \infty$ ?

If question (i) has a positive answer, then the cgf (5) must tend in probability to the nonrandom limit, since convergence in distribution to a nonrandom limit implies convergence in probability. Question (ii), on the other hand, requires only convergence in distribution.

A detailed answer to question (i) is found in Hall (1990a). The distribution with cgf (5) has a nonrandom limit if and only if the distribution of the  $Y_j$  either is in the domain of attraction of a normal law or has slowly varying tails one of which completely dominates the other. The former of these possibilities is of no interest for the present paper, where our concern is with heavy-tailed laws. The latter is a special case of what we consider here, but, in that case, as Hall remarks, the nonrandom limit of the bootstrap distribution bears no relation to the actual distribution of the normalised mean.

Regarding question (ii), we have seen that the distribution of  $p(\mathbf{Y})$  is nonrandom, since  $p(\mathbf{Y})$  is the deterministic measurable function of  $\mathbf{Y}$  given by (7). The question is whether the distribution converges to a limiting distribution as  $n \rightarrow \infty$ . A part of the answer is provided by the result of Logan, Mallows, Rice, and Shepp (1973), where it is seen that the self-normalised sum

$$t \equiv \frac{\sum_{j=1}^n Y_j}{(\sum_{j=1}^n (Y_j - \bar{Y})^2)^{1/2}} \quad (8)$$

that appears in (7) has a limiting distribution when  $n \rightarrow \infty$ . In fact, what we have to show here, in order to demonstrate that the bootstrap  $P$  value has a limiting

distribution, is that the self-normalised sum and the CDF  $F_{\mathbf{Y}}^n$  have a limiting joint distribution.

This can be shown by a straightforward extension of the proof in Logan *et al.*, in which it is demonstrated that, when the  $Y_j$  follow a law in  $DA(\alpha)$ , the joint characteristic function of the sum  $n^{-\gamma} \sum_j Y_j$  (recall  $\gamma = 1/\alpha$ ) and any sum of the form  $n^{-p\gamma} \sum_j |X_j|^p$ , with  $p > \alpha$ , tends as  $n \rightarrow \infty$  to a continuous limit. From this it can be deduced from Lévy's continuity theorem that the sums have a limiting joint distribution.

The cumulant-generating function (5) can be written as

$$\phi_{\mathbf{Y}}^n(t) \equiv \sum_{j=1}^n (\exp(in^{-\gamma}tY_j) \exp(-in^{-\gamma}t\bar{Y}) - 1).$$

This is a deterministic function of  $\bar{Y}$  and

$$\tilde{\phi}_{\mathbf{Y}}^n(t) \equiv \sum_{j=1}^n (\exp(in^{-\gamma}tY_j) - 1).$$

Consider the joint characteristic function of  $\tilde{\phi}_{\mathbf{Y}}^n(t)$  and  $U_n \equiv n^{-\gamma} \sum_j Y_j$ . As a function of arguments  $s_1$  and  $s_t$ , it is

$$\mathbb{E}(\exp(is_1U_n + is_t\tilde{\phi}_{\mathbf{Y}}^n(t))) = \left[ \mathbb{E}(\exp(is_1n^{-\gamma}Y + is_t(\exp(in^{-\gamma}tY) - 1))) \right]^n. \quad (9)$$

Let  $g$  be the density of the law of the  $Y_j$ , assumed to be a stable law with tail index  $\alpha$ , and  $1 < \alpha < 2$ . The first moment exists, and is assumed to be zero. Then the expectation in (9) can be written as

$$1 + \int_{-\infty}^{\infty} \left( \exp(is_1n^{-\gamma}y + is_t(\exp(in^{-\gamma}ty) - 1)) - 1 - is_1n^{-\gamma}y + s_ttn^{-\gamma}y \right) g(y) dy, \quad (10)$$

because  $\int_{-\infty}^{\infty} g(y) dy = 1$  and  $\int_{-\infty}^{\infty} yg(y) dy = 0$ . The density  $g$  has the following limiting behaviour:

$$y^{\alpha+1}g(y) \rightarrow r \quad \text{and} \quad y^{\alpha+1}g(-y) \rightarrow l \quad \text{as } y \rightarrow \infty, \quad (11)$$

for nonnegative constants  $r$  and  $l$  determined by the skewness of the distribution, that is, by the parameter  $\beta$  of the stable law.

Following the proof in Logan *et al.*, we change the integration variable to  $x = n^{-\gamma}y$ . The expectation (10) becomes

$$\begin{aligned} & 1 + \int_{-\infty}^{\infty} \left( \exp(is_1x + is_t(\exp(itx) - 1)) - 1 - is_1x + s_ttx \right) g(n^\gamma x) n^\gamma dx \\ &= 1 + \frac{1}{n} \int_{-\infty}^{\infty} \left( \exp(is_1x + is_t(\exp(itx) - 1)) - 1 - is_1x + s_ttx \right) \\ & \quad g(n^\gamma x) (n^\gamma |x|)^{\alpha+1} |x|^{-(1+\alpha)} dx \\ &\sim 1 + \frac{1}{n} \int_{-\infty}^{\infty} \left( \exp(is_1x + is_t(\exp(itx) - 1)) - 1 - is_1x + s_ttx \right) K(x) |x|^{-(1+\alpha)} dx, \end{aligned}$$

where, on account of (11),  $K(x) = l$  for  $x < 0$  and  $K(x) = r$  for  $x > 0$ . Note that, despite the factor of  $|x|^{-(1+\alpha)}$ , with exponent between -2 and -3, the integral is still convergent at  $x = 0$ , because, in the neighbourhood of  $x = 0$ , we have

$$\begin{aligned} & \exp(is_1x + is_t(\exp(itx) - 1)) - 1 - is_1x + s_ttx \\ & \sim is_t(\exp(itx) - 1) + s_ttx + O(x^2) \sim -s_ttx + s_ttx + O(x^2) = O(x^2), \end{aligned}$$

so that the integrand behaves like  $x^{1-\alpha}$  near  $x = 0$ . Since  $1 - \alpha > -1$ , there is no divergence at  $x = 0$ .

We conclude that there is a limiting joint distribution of  $U_n$  and  $\phi_{\mathbf{Y}}^n(t)$  for any real  $t$ . It is clear that the same method of proof, essentially that of Logan *et al.*, equally well establishes the existence of a limiting joint distribution of the self-normalised sum and the CDF  $F_{\mathbf{Y}}$ . This is what we need to conclude that the bootstrap  $P$  value does indeed have a limiting distribution as  $n \rightarrow \infty$ .

#### 4. The $m$ out of $n$ bootstrap and subsampling

The  $m$  out of  $n$  bootstrap is based on the same principle as the ordinary resampling bootstrap, the only difference being that the bootstrap sample size is equal to  $m$ , smaller than  $n$ . As a consequence, the bootstrap statistic  $\tau^*$  of (3) is replaced by

$$\tau_m^* = m^{-\gamma} \sum_{j=1}^m (Y_j^* - \bar{Y}).$$

The  $m$  out of  $n$  bootstrap for the stable distributions was first studied by Athreya (1987), whose pioneering work was continued by Giné and Zinn (1989), Arcones and Giné (1989), Bickel, Gotze, and van Zwet (1997), and Hall and Jing (1998).

The choice of the bootstrap sample size  $m$  is an important matter. It has to be chosen such that the following conditions are satisfied:

$$m \rightarrow \infty, \quad \text{and} \quad m/n \rightarrow 0 \quad \text{or} \quad m(\log \log n)/n \rightarrow 0. \quad (12)$$

The motivation behind the first of these conditions is that it allows us to apply the law of large numbers. In addition, we need the second condition in order for the distribution of  $\tau_m^*$  to converge in probability, or the third for almost sure convergence, to the distribution of  $\tau$ . Proofs of the appropriate large-sample behaviour of the  $m$  out of  $n$  bootstrap can be found in Athreya (1987) and Arcones and Giné (1989). Papers discussing the choice of  $m$  include Datta and McCormick (1995) and Bickel and Sakov (2005).

The subsampling method proposed in Romano and Wolf (1999) is an alternative to the  $m$  out of  $n$  bootstrap. The main difference between the two is that, in the former, resampling is done without replacement.



In this section we show in a simulation study that, despite its consistency, the  $m$  out of  $n$  bootstrap of the statistic  $\tau$  of (2) does not give reliable inference results in other than very large samples. The subsampling method does better, but still suffers from serious distortion if  $m$  is not chosen with care. Figure 1 displays  $P$  value discrepancy plots for the  $m$  out of  $n$  bootstrap. The plots are based on 10,000 realizations of the statistic  $\tau$  from samples of size 100 generated by the symmetric stable distribution with the value of  $\alpha = 1.5$ , supposed known. The bootstrap sample size  $m$  took the values 10, 40, 70, 90. The number of bootstrap replications was  $B = 399$ . Fig. 1 indicates that the  $m$  out of  $n$  bootstrap suffers from considerable size distortions for all values of  $m$ . Moreover, for the usual significance levels of 0.05 and 0.1, it is outperformed by the inconsistent ordinary bootstrap, which here is the bootstrap with  $m = 100$ .

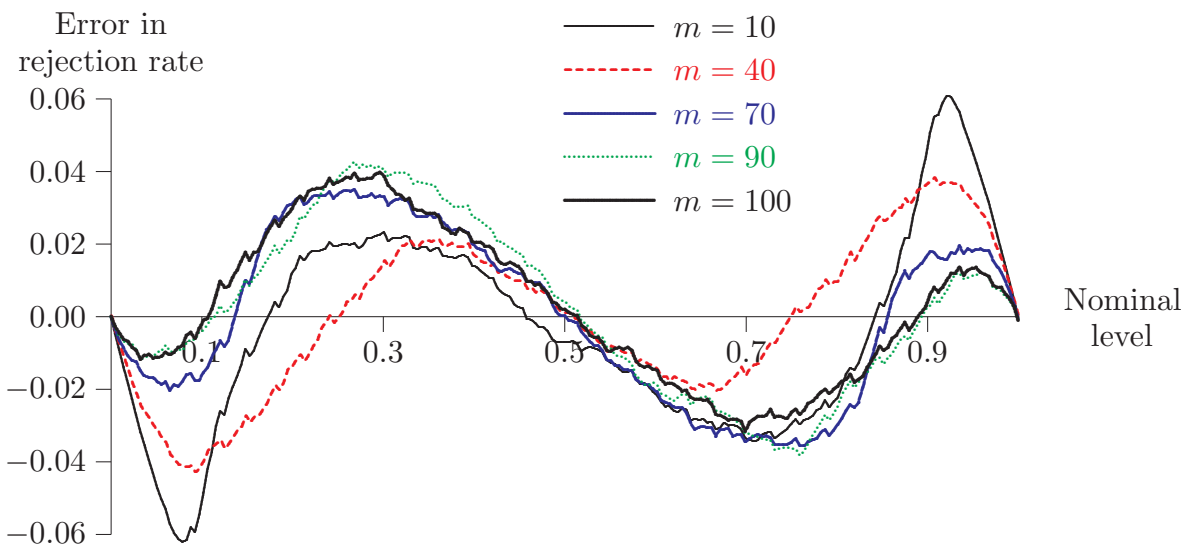


Figure 1:  $P$  value discrepancy plots,  $m$  out of  $n$  bootstrap,  $\alpha = 1.5$ ,  $\beta = 0$ ,  $n = 100$

In Figure 2 we consider the same scenario as in Figure 1, but with  $n = 2,000$ . We show the results only for those values of  $m$  that gave the smallest error in rejection probability. We see that the error in rejection probability of the  $m$  out of  $n$  bootstrap does not vanish even in samples as large as 2,000. The size distortion of the ordinary bootstrap is almost of the same magnitude as that of the  $m$  out of  $n$  bootstrap. These results indicate that the rate of convergence of the  $m$  out of  $n$  bootstrap  $P$  value to its limit distribution is very slow.

Figure 3 shows comparable results for the subsampling bootstrap with  $n = 100$ . For values of  $m$  greater than about 50, the distortions become very large, and so are not shown. Note, though, that for  $m = 37$ , distortion is quite modest. But even a small difference in  $m$  can, as is seen in the figure, give rise to considerable distortion. The size distortions of the subsampling and the  $m$  out of  $n$  bootstrap are even larger when we consider data from an asymmetric stable distribution with  $\alpha = 1.5$  and  $\beta = 0.5$ .

The simulation results presented here are a manifestation of a result of Hall and Jing (1998), which shows that the difference between the estimate of the distribution of the sample mean given by the  $m$  out of  $n$  bootstrap and the true distribution tends to

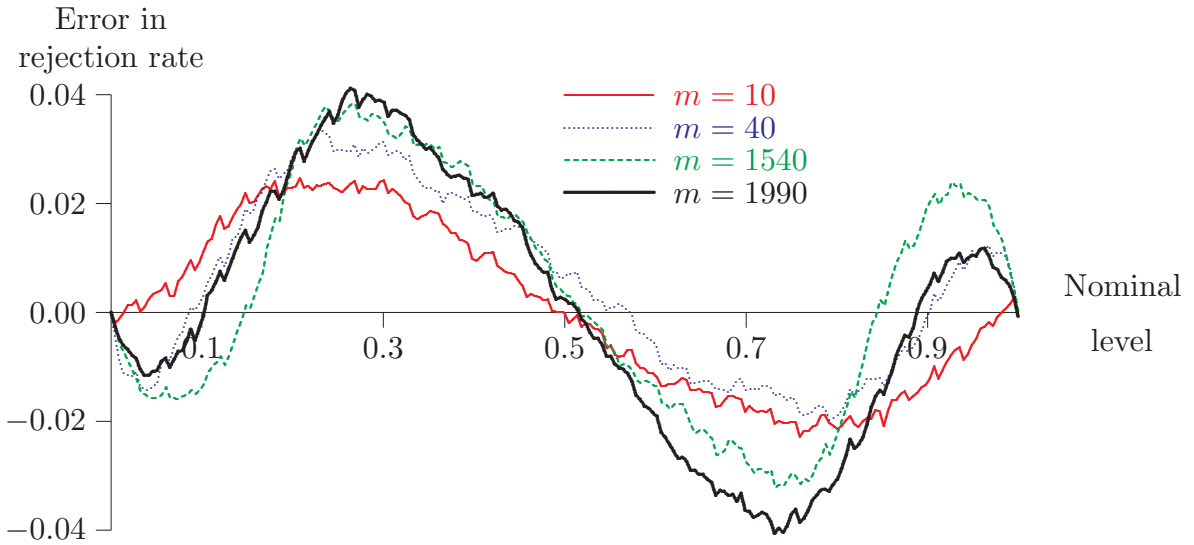


Figure 2:  $P$  value discrepancy plots,  $m$  out of  $n$  bootstrap,  $\alpha = 1.5$ ,  $n = 2000$

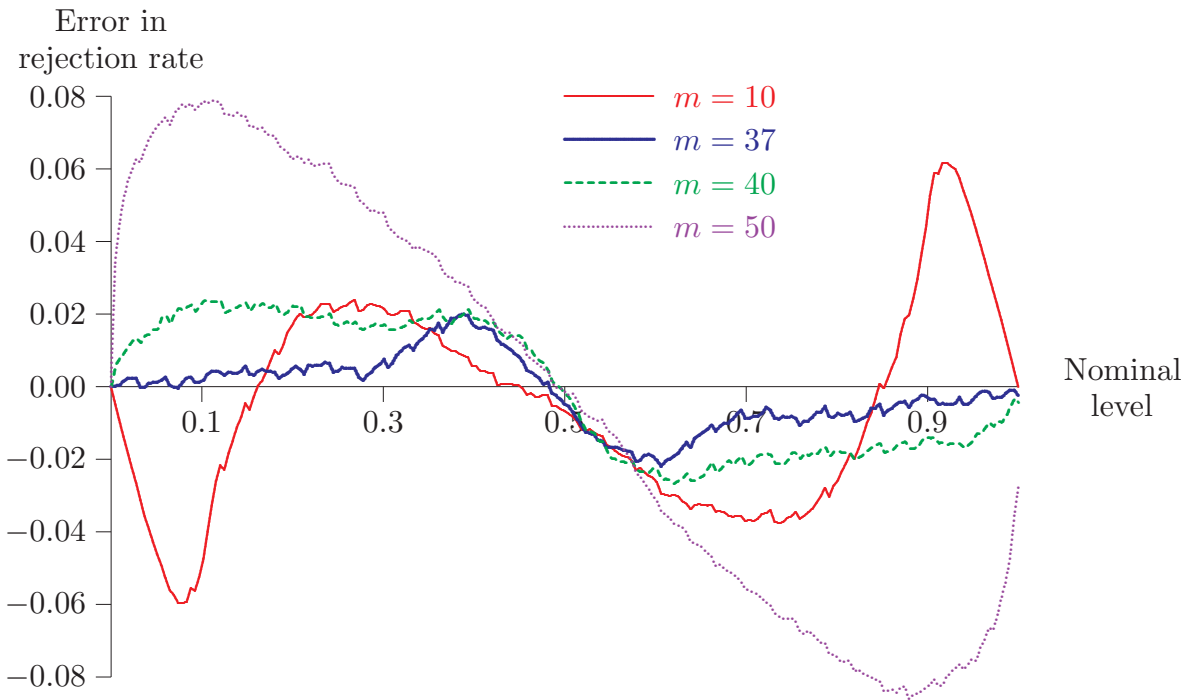


Figure 3:  $P$  value discrepancy plots, subsampling,  $\alpha = 1.5$ ,  $\beta = 0$ ,  $n = 100$

zero as  $n \rightarrow \infty$  more slowly than that of an estimate that assumes a stable law with consistently estimated parameters.

## 5. Studentized statistics

The statistic  $\tau$  depends on the tail index  $\alpha$ , which in practice would have to be esti-

mated. A natural way to avoid this is to employ the studentized statistic:

$$t = \frac{n^{1/2}\bar{Y}}{\{(n-1)^{-1} \sum_{j=1}^n (Y_j - \bar{Y})^2\}^{1/2}}.$$

Note that this is equal, apart from a multiplicative constant, to the self-normalised sum in the argument of  $F_{\bar{Y}}^n$  in (7). The statistic for the usual resampling bootstrap is

$$t^* = \frac{n^{1/2}(\bar{Y}^* - \bar{Y})}{\{(n-1)^{-1} \sum_{j=1}^n (Y_j^* - \bar{Y}^*)^2\}^{1/2}},$$

with  $n$  replaced by  $m$  for the  $m$  out of  $n$  bootstrap and subsampling. Hall (1990a) shows that the ordinary bootstrap of the  $t$  statistic is not consistent if  $F \in DA(\alpha)$  with  $1 < \alpha < 2$ . But Hall and LePage (1996) prove that the  $m$  out of  $n$  bootstrap is justified asymptotically. The same is shown by Romano and Wolf(1999) for subsampling.

Figure 4 displays the  $P$  value discrepancy plots for the  $m$  out of  $n$  bootstrap based on the  $t$  statistic, using samples of size  $n = 100$  from the stable distribution with  $\alpha = 1.5$  and  $\beta = 0$ , these values again assumed known. This bootstrap suffers from large size distortions for any value of  $m$  between 10 and 100. Again, the rate of convergence to the limit is very slow. In addition, the resampling bootstrap performs just as badly. Compared with the bootstrap of the nonstudentized statistic  $\tau$ , which underrejects for conventional significance levels, the bootstrap of the  $t$  statistic systematically overrejects. Figure 5 shows similar results for subsampling. Distortions are smallest for very small values of  $m$ , but increase very quickly as  $m$  gets larger.

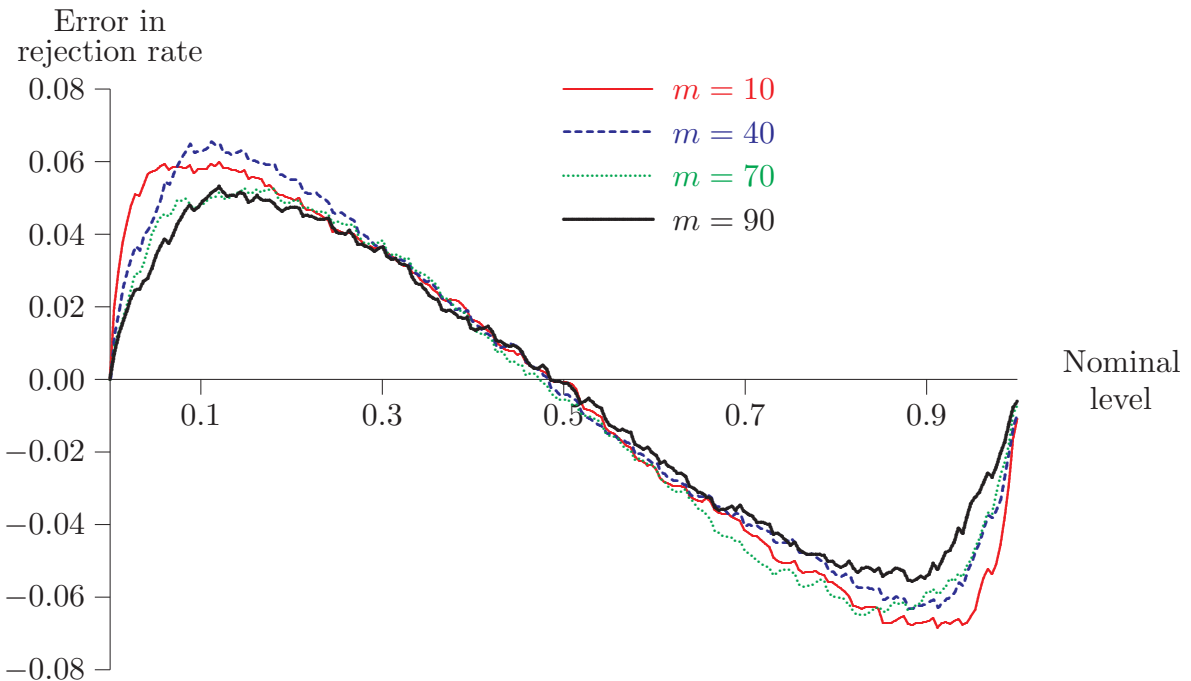


Figure 4:  $P$  value discrepancy plots,  $m$  out of  $n$  bootstrap for  $t$ ,  $\alpha = 1.5$ ,  $\beta = 0$ ,  $n = 100$

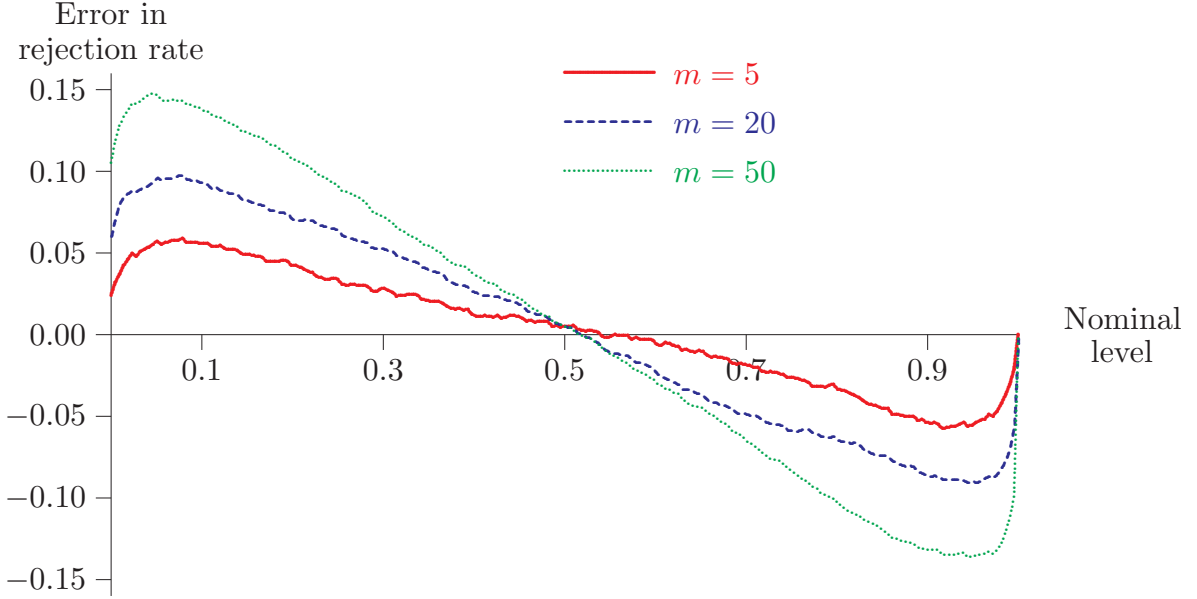


Figure 5:  $P$  value discrepancy plots, subsampling for  $t$ ,  $\alpha = 1.5$ ,  $\beta = 0$ ,  $n = 100$

The results of this and the previous section indicate that, although consistency is necessary to avoid bootstrap failure, it does not guarantee reliable inference in samples of moderate size.

## 6. A parametric bootstrap

In the [Corollary](#) of Section 2, the distribution of the statistic  $t(M)$  of (6) was shown to have expectation 0 and variance 1. A simulation study not reported in detail here shows that, for values of  $n$  in the range from 20 to 2,000, the distribution is not too far removed from standard normal. Suppose for a moment that the CDF of  $t(M)$  is actually equal to  $\Phi$ , the standard normal CDF. Then the bootstrap  $P$  value  $p(\mathbf{Y})$  of (7) would be  $\Phi(t)$ , where  $t$  is given by (8), and its CDF would be

$$\Pr(p(\mathbf{Y}) \leq u) = \Pr(\Phi(t) \leq u) = \Pr(t \leq \Phi^{-1}(u)).$$

Recall that Logan, Mallows, Rice, and Shepp ([1973](#)) have shown that  $t$  has a limiting distribution. This distribution depends on the parameters  $\alpha$  and  $\beta$  of the distribution of the  $Y_j$ ; let us denote its CDF by  $G_{\alpha,\beta}$ . The limiting distribution of  $p(\mathbf{Y})$  would thus have CDF  $G_{\alpha,\beta} \circ \Phi^{-1}$ . Provided that  $\alpha$  and  $\beta$  can be estimated consistently, an asymptotically valid test of the hypothesis that the expectation of the  $Y_j$  is zero could be based on  $p(\mathbf{Y})$  and the estimated CDF  $G_{\hat{\alpha},\hat{\beta}} \circ \Phi^{-1}$ .

Such a test would be equivalent to a test based on  $t$  itself and its estimated limiting distribution  $G_{\hat{\alpha},\hat{\beta}}$ . However, for a test based on  $t$ , there is no need to assume that the conditional distribution of  $t(M)$  is standard normal. Even so, the asymptotic distribution function  $G_{\alpha,\beta}$  is characterized by a complex integral involving parabolic cylinder functions, and so computing it is a nontrivial task. For a finite sample,

therefore, it is easier and preferable to estimate the distribution of  $t$  consistently by simulation of self-normalized sums from samples of stable random variables with  $\alpha$  and  $\beta$  consistently estimated from the original sample. This amounts to a parametric bootstrap of  $t$ . It is not absolutely equivalent to a parametric bootstrap of  $p(\mathbf{Y})$  itself, which would be as computationally intensive as a double bootstrap. It is not unrelated to, but is still different from, a suggestion in Hall and Jing (1998) that, for a symmetric distribution, one might bootstrap  $\hat{c}S(\hat{\alpha}, 0)(n^{-1/\hat{\alpha}}(\bar{Y} - \delta_0))$ , a procedure which they acknowledge to be time-consuming.

An advantage of a parametric bootstrap of  $t$  is that its asymptotic distribution applies not only when the  $Y_j$  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 centred simulated variables, and the statistic  $t$ , being normalized, is invariant to scale.

The proposed bootstrap is described by the following steps:

1. Given the sample of random variables  $Y_1, \dots, Y_n$  with distribution  $F \in DA(\alpha)$ , compute the self-normalized sum  $t$ .
2. Estimate  $\alpha$  and  $\beta$  consistently from the original sample.
3. Draw  $B$  samples of size  $n$  from  $S(\hat{\alpha}, \hat{\beta})$  with  $\hat{\alpha}$  and  $\hat{\beta}$  obtained in the previous step.
4. For each sample of the stable random variables compute the bootstrap self-normalized sum,

$$t^* = \frac{\sum_{j=1}^n Y_j^*}{\left(\sum_{j=1}^n (Y_j^* - \bar{Y}^*)^2\right)^{1/2}}.$$

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

## Proposition 2

The distribution of  $t^*$ , conditional on the sample  $Y_1, \dots, Y_n$ , approaches that of  $t$  as  $n \rightarrow \infty$  when the  $Y_j$  are drawn from a distribution in the domain of attraction of a non-Gaussian stable law  $S(\alpha, \beta)$ .

The result follows from three facts: first, the consistency of the estimators  $\hat{\alpha}$  and  $\hat{\beta}$ , second, the continuity of the stable distributions with respect to  $\alpha$  and  $\beta$ , and, third, the result of Logan *et al.* that shows that the self-normalized sum has the same asymptotic distribution for all laws in the domain of attraction of a given stable law  $S(\alpha, \beta)$ . ■

### Rate of convergence

If the distribution of the variables  $Y_j$  is in the domain of attraction of  $cS(\alpha, \beta)$ , Logan *et al.* (1973) show that the joint characteristic function of the numerator and denominator of  $t$ , as a function of arguments  $s_n$  and  $s_d$ , is  $(1 + I_n/n)^n$ , where  $f$  is the density of the law of the  $Y_j$ , and

$$I_n = \int_{-\infty}^{\infty} \frac{\exp(ixs_n + ix^2s_d) - 1 - ix s_n}{|x|^{\alpha+1}} (n^{1/\alpha}|x|)^{\alpha+1} f(n^{1/\alpha}x) dx. \quad (13)$$

Let  $I_\infty$  denote the limit of  $I_n$ . Then  $(1 + I_\infty/n)^n = \exp I_\infty + O(n^{-1})$ .

Hall (1982) introduces a restrictive condition on laws in the domain of attraction of a stable law, as follows. The CDF  $F$  of the distribution satisfies

$$\begin{aligned} 1 - F(y) &= ry^{-\alpha}(1 + d_r y^{-\delta_r} + o(y^{-\delta_r})) \quad \text{and} \\ F(y) &= l|y|^{-\alpha}(1 + d_l |y|^{-\delta_l} + o(|y|^{-\delta_l})) \end{aligned} \quad (14)$$

as  $y \rightarrow \infty$ , where  $\delta_r, \delta_l > 0$ ,  $d_r$  and  $d_l$  are real numbers and  $r$  and  $l$  are as defined in (11). If Hall's condition on the distribution holds, then, for the density in the right-hand tail of (13), we have  $y^{\alpha+1} f(y) = \alpha r(1 + d_r y^{-\delta_r} + o(y^{-\delta_r}))$ , and so

$$(n^{1/\alpha}x)^{\alpha+1} f(n^{1/\alpha}x) = \alpha r(1 + d_r n^{-\delta_r/\alpha} x^{-\delta_r} + o(n^{-\delta_r/\alpha})),$$

where  $y = n^{1/\alpha}x$ . A similar relation holds in the left-hand tail. Thus the rate of convergence of the integrand in  $I_n$  is that of  $n^{-\delta/\alpha}$  where  $\delta = \min(\delta_r, \delta_l)$ . The rate of convergence of the joint characteristic function itself is thus the slower of  $n^{-1}$  and  $n^{-\delta/\alpha}$ , as is therefore the rate of convergence of the distribution of  $t$ . In the cases studied by Hall (1982),  $\delta \geq \alpha$ , and so in those cases the rate is  $n^{-1}$ , from which it follows that the rate of convergence of the bootstrap distribution is that of the estimators  $\hat{\alpha}$  and  $\hat{\beta}$ .

### Remark 3

If Hall's condition (14) is not satisfied, then the rate of convergence can be much slower, as we will see in the simulation study of Section 8.

### Remark 4

The test statistic  $t$  is neither a pivot nor an asymptotic pivot, since its asymptotic distribution depends upon the parameters  $\alpha$  and  $\beta$ . As suggested by Beran (1988), one way to achieve asymptotic pivotalness is to transform  $t$  by its limiting distribution function. Here that function is  $G_{\alpha,\beta}$ , which can be consistently estimated. The resulting statistic is an asymptotic pivot with limiting distribution the uniform  $U(0,1)$  distribution. Of course, this procedure entails evaluating the integral with parabolic cylinder functions.

## 7. Methods for the estimation of the parameters

The problem of estimating the parameters  $\alpha$  and  $\beta$  is hampered by the fact that the density and distribution function of a non-Gaussian stable law have no closed-form expression. Nevertheless, there are estimation procedures for the stable laws, and also more general procedures that use only the information in the tails of the distributions. In the first category we have maximum likelihood (DuMouchel (1973), Nolan (2001)), characteristic function methods (Koutrouvelis (1980), Kogon and Williams (1998)), the quantile method of McCulloch (1986), the indirect inference method of Garcia, Renault, and Veredas (2006), and the continuous generalized method of moments of Carrasco and Florens (2000). In the second category we may cite Pickands's estimator (Pickands (1975)), Hill's estimator (Hill (1975)), and the estimator of de Haan and Resnick (de Haan and Resnick (1980)). We survey some of these in this section.

For the tails of the stable law  $cS(\alpha, \beta)$  we have

$$\lim_{y \rightarrow \infty} y^\alpha \Pr(Y > y) = C_\alpha \frac{1 + \beta}{2} c^\alpha, \quad \lim_{y \rightarrow \infty} y^\alpha \Pr(Y < -y) = C_\alpha \frac{1 - \beta}{2} c^\alpha,$$

where  $C_\alpha^{-1}$  is  $\int_0^\infty y^{-\alpha} \sin y \, dy$ ; see Samorodnitsky and Taqqu (1994). It follows from (11) that

$$\frac{r}{l} = \frac{1 + \beta}{1 - \beta}. \quad (15)$$

On account of (15), for any distribution in the domain of attraction of a stable law  $cS(\alpha, \beta)$ , we can estimate the skewness parameter  $\beta$  if we can estimate  $r$  and  $l$ .

### Hill's method

The best known estimator that uses only the information in the tails of the distribution is the Hill estimator. It is based on the  $k$  largest of the order statistics of a sample of independent and identically distributed random variables  $Y_{(1)} > Y_{(2)} > \dots > Y_{(n)}$ , and is equal to

$$\hat{\alpha}_{\text{Hill}} = \left( \frac{1}{k-1} \sum_{j=1}^{k-1} \log Y_{(j)} - \log Y_{(k)} \right)^{-1}.$$

If  $k$  is the number of order statistics used for the estimation of  $\alpha$  in the right tail of the distribution, then  $r$  is estimated by  $kY_{(k)}^{\hat{\alpha}_{\text{Hill}}}/n$ . The parameter  $l$  is estimated in a similar way, using the information in the left tail of the distribution.

A variant for the estimation of  $r$  and  $l$  is introduced by Aban and Meerschaert (2004). They define

$$\hat{\mu} = \log Y_{(k)} - \hat{\alpha}_{\text{Hill}}^{-1} \sum_{j=k}^n \frac{1}{j},$$

and then the estimate of  $r$  is  $\exp(\hat{\alpha}_{\text{Hill}}\hat{\mu})$ . An estimate of  $l$  is obtained similarly.

## Quantile method

One of the methods for estimating the parameters  $\alpha$  and  $\beta$  was introduced by Fama and Roll (1971) for symmetric stable distributions, and extended by McCulloch (1986) for the general asymmetric case. Using the 0.05, 0.25, 0.5, 0.75, and 0.95 quantiles, one computes the following indices:

$$\nu_\alpha = \frac{q_{0.95} - q_{0.05}}{q_{0.75} - q_{0.25}}, \quad \nu_\beta = \frac{q_{0.95} + q_{0.05} - q_{0.5}}{q_{0.95} - q_{0.05}}.$$

The indices are then inverted using tables in McCulloch's paper in order to obtain consistent estimators of  $\alpha$  and  $\beta$ . McCulloch suggests that these estimators can be used as a very good starting point for more sophisticated and theoretically superior estimators, such as maximum likelihood.

## Maximum likelihood

The maximum likelihood estimators of the parameters  $\alpha$  and  $\beta$  are based on numerical approximations to the density of the stable distributions. With no closed-form representation of the density, this is not a trivial task. The MLE was first obtained by DuMouchel (1973) by using the Fast Fourier Transform and Bergstrom series expansions for the tails. He showed that the standard theory, in terms of root- $n$  asymptotic normality and Cramér-Rao bounds, applies for the maximum likelihood estimators of the parameters of the stable laws.

Nolan (2001) continued the pioneering work of DuMouchel and optimized the method by employing direct numerical integration of the stable density derived from one of the parametrizations of the characteristic function in Zolotarev (1986)).

## Characteristic function methods

The one-one correspondence between the density and the characteristic function motivates the characteristic function approaches for the estimation of the parameters of stable distributions. The methods that fall into this category and have proved to have the best performance are those proposed by Koutrouvelis and Kogon-Williams, by using different expressions of the cumulant generating function of the stable law. Let  $\hat{\zeta}(s) = n^{-1} \sum_{j=1}^n e^{isY_j}$  be the empirical characteristic function. Koutrouvelis' method is based on a parametrization of the characteristic function given in Zolotarev (1986) that is not continuous in all parameters. The tail index  $\alpha$  is estimated by ordinary least squares as the coefficient of  $\log s$  in the following regression:

$$\log(-\log |\hat{\zeta}(s)|) = \alpha \log c + \alpha \log s + u_s, \quad s = 1, \dots, K. \quad (16)$$

where  $u_s$  is a disturbance term. The values of  $s$  are chosen, following Koutrouvelis, as  $s = \pi k/25$ ,  $k = 1, \dots, K$ . Some experimentation showed us that, for sample size  $n = 100$ , the mean squared error of the estimator is minimised near  $K = 16$ . The parameter  $c$  can be estimated using the estimated constant  $\hat{a}$  from (16) as  $\hat{c} = \exp(\hat{a}/\hat{\alpha})$ .



Then  $\beta$  and  $\delta$  are estimated by running the regression

$$\text{Arg } \hat{\zeta}(s) = \delta s + \beta(cs)^\alpha \tan \frac{\pi\alpha}{2} + u_s, \quad (17)$$

where  $\text{Arg}$  denotes the principal argument of the complex number  $\hat{\zeta}(s)$ , that is, the angle  $\theta$  such that  $\hat{\zeta}(s) = |\hat{\zeta}(s)|e^{i\theta}$  and  $-\pi < \theta \leq \pi$ . In regression (17),  $\text{Arg } \hat{\zeta}(s)$  is regressed on  $s$  and  $s^{\hat{\alpha}}$ , where  $\hat{\alpha}$  is from (16). Then the estimated coefficient of  $s$  is the estimate of  $\delta$ , and the estimate of  $\beta$  is  $\hat{\beta} = \hat{b}\hat{c}^{-\hat{\alpha}} \cot(\pi\hat{\alpha}/2)$ , where  $\hat{b}$  is the estimated coefficient of  $s^{\hat{\alpha}}$  and  $\hat{c}$  is from (16). Koutrouvelis recommends setting the values of  $s$  in (17) as  $s = \pi l/50$ ,  $l = 1, \dots, L$ .

The method of Kogon and Williams is based on another of Zolotarev's parametrizations of the characteristic function, one that is continuous in all parameters. The regression for the estimation of  $\alpha$  is the same as in Koutrouvelis' method. The parameter  $\beta$  is estimated by ordinary least squares from the following regression

$$\text{Arg } \hat{\zeta}(s) = \left(\delta + \beta c^\alpha \tan \frac{\pi\alpha}{2}\right)s + \beta cs \tan \frac{\pi\alpha}{2} ((cs)^{\alpha-1} - 1) + u_s.$$

## Comments

For the estimation of  $\alpha$  we prefer the method of Koutrouvelis in cases in which the underlying distribution is stable. Simulation results, not reported here, indicate that Koutrouvelis' method performs as well as maximum likelihood in terms of the root mean squared error and bias, and is much less time-consuming. The more robust method of Hill is fast and performs well provided the optimal number of order statistics is used, which in practice can be obtained by employing the Hill plot – see Hill (1975) – or the  $m$  out of  $n$  bootstrap – see Hall (1990b) and Caers and Dyck (1999). For the estimation of the skewness parameter  $\beta$ , provided the underlying distribution is in fact a stable law, the quantiles method is the best, since it does not depend on the estimate of  $\alpha$ , unlike Koutrouvelis' method. For laws that are not stable but belong to the domain of attraction of a stable law, the Aban-Meerschaert estimates of  $r$  and  $l$  seem to give the best estimates of  $\beta$ . We made no use of any other of the methods mentioned above.

## Rate of convergence of parameter estimates

The distribution of the elements of a sample drawn using the parametric bootstrap is the stable law with parameters  $\hat{\alpha}$  and  $\hat{\beta}$ , and the true distribution under the null is a law in the domain of attraction of the stable law with parameters  $\alpha$  and  $\beta$ . These distributions differ both because their parameters are different, and because the true distribution is not necessarily a stable law. Therefore the rate of convergence of the distribution of  $t^*$  to that of  $t$  is the slower of

- (i) the rate of convergence of the estimators  $\hat{\alpha}$  and  $\hat{\beta}$  to their true values, and

- (ii) the rate of convergence of the distribution of the self-normalized sum  $t$  to its limit distribution.

We dealt with point (ii) in the last section. If we assume that the data are distributed according to a stable distribution, the maximum likelihood and the characteristic function methods give the usual parametric root- $n$  rate. If we suppose only that the distribution is somewhere in the entire domain of attraction of the stable laws, then Hill's method could be used. In this situation an optimal number of order statistics  $k$  has to be chosen. Hall (1982) shows that, if the underlying distribution  $F$  satisfies conditions (14), then it is asymptotically optimal to choose  $k$  such that  $k = O(n^{2\delta/(2\delta+\alpha)})$ . In that case,

$$\hat{\alpha} - \alpha = O_p(k^{-1/2}) \quad \text{and} \quad \hat{r} - r = O_p(k^{-1/2} \log(n/k)). \quad (18)$$

From (15) it can be seen that the order of  $\hat{\beta} - \beta$  is that of  $\hat{r} - r$ .

Hall's condition is more demanding than just requiring the distribution  $F$  to be in the domain of attraction of a stable law or even in the domain of normal attraction. 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. The stable laws themselves satisfy Hall's condition with  $\delta = \alpha$ . For Student's  $t$ ,  $\delta = 2$ . Hall lists some other distributions that satisfy his condition.

The optimal number of extreme values used by Hill's method depends on unknown properties of the tails. In practice  $k$  must itself be estimated from the sample. Simulations show that the exponential test suggested by Hill (1975) has very low power and tends to overestimate  $k$ . Hall and Welsh (1985) use adaptive methods in order to estimate the size of the extreme subsample. Under conditions (14), the effect of the estimation of  $k$  is that

$$\hat{\alpha} - \alpha = O_p(n^{-\gamma/(2\gamma+\alpha)}) \quad \text{and} \quad \hat{r} - r = O_p(n^{-\gamma/(2\gamma+\alpha)} \log n).$$

Thus the rate of convergence of the estimators is slower compared with the case (18) when  $k$  is assumed to be known.

The rate of convergence (18) does not hold for all distributions in the domains of attraction of the stable laws. Csorgo and Viharos (1997) and the references therein, extended Hill's estimators to the entire domain of attraction  $DA(\alpha)$ . However, they show that the price paid for this generality is a lower rate of convergence. In addition, their results support Hall and Jing's (1998) point of view indicating that "sophisticated" methods based upon extreme order statistics can improve the rate of convergence, but not up to the parametric rate of  $n^{-1/2}$  without sacrificing robustness.

## 8. Simulation evidence

In this section we investigate the performance of the parametric bootstrap in samples of size  $n = 100$  in a simulation study. First, we consider the case in which the null hypothesis  $\delta = 0$  is true. All results are based on 10,000 replications of the statistic  $t$  and 399 bootstrap repetitions in step 3 of the bootstrap algorithm. They are displayed graphically as  $P$  value discrepancy plots. The error in rejection probability (ERP), that is, the difference between the actual rejection rate based on the bootstrap  $P$  value and the nominal significance level should be close to zero if the parametric bootstrap works well.

In Figure 6, the data were generated from the stable distribution with  $\alpha = 1.5$  and  $\beta = 0, 0.5$  and  $1$ , all assumed known. The parametric bootstrap performs very well in cases in which, as the simulation results from Section 4 indicate, the subsampling method and the  $m$  out of  $n$  bootstrap have large ERPs, especially when  $\beta \neq 0$ .

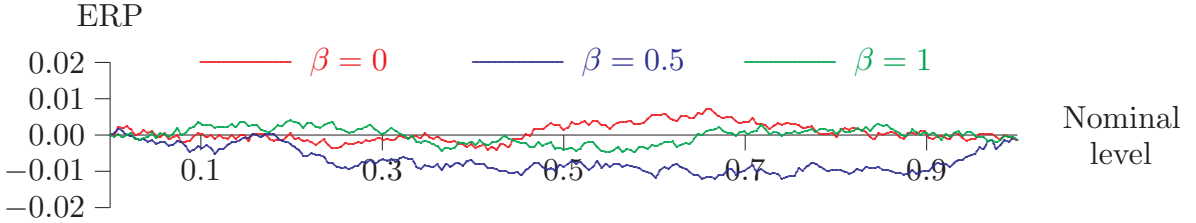


Figure 6:  $P$  value discrepancy plots,  $\alpha = 1.5$ ,  $n = 100$

In complete contrast to the results shown in Figure 6, the  $P$  value discrepancy plots in Figure 7 show what happens when the data are generated from a loggamma distribution. If  $X$  is a random variable from a gamma distribution with density  $x^{a-1} \exp(-x/b)/(b^a \Gamma(a))$  (here  $\Gamma$  is the gamma function), where  $a$  is the shape parameter and  $b$  is the scale, then  $Y = \exp X$  has the loggamma distribution, with density  $(\log y)^{a-1} y^{-1/b-1}/(b^a \Gamma(a))$ , where  $1/b = \alpha$  is the tail index and  $a$  is the scale. The distribution of  $Y$  has tails that behave like

$$1 - F(y) \sim \frac{\alpha^{a-1}}{\Gamma(a)} (\log y)^{a-1} y^{-\alpha},$$

as  $y \rightarrow +\infty$ . For  $a \neq 1$ , the loggamma distribution is in  $DA(\alpha)$ , but not in domain of normal attraction. In particular, Hall's conditions (14) are not satisfied. For the figure,  $a = 1.7$  and  $b = 0.666667$ , so that  $1/b = \alpha = 1.5$ , assumed known. Since the loggamma distribution is totally skewed to the right, we take  $\beta = 1$ .

When  $n = 100$ , the smallest ERP for the subsampling method and the  $m$  out of  $n$  bootstrap was achieved for  $m = 10$ . However, even with this optimal choice of the bootstrap sample size, the size distortions are very severe, for instance 0.3 for the 0.05 significance level. The parametric bootstrap performs much better, but the ERPs are not close to zero. For the 0.05 significance level the size distortion is around 0.14 for a sample size of 100 and 0.08 for a sample size of 1,000. Since the parameters of the

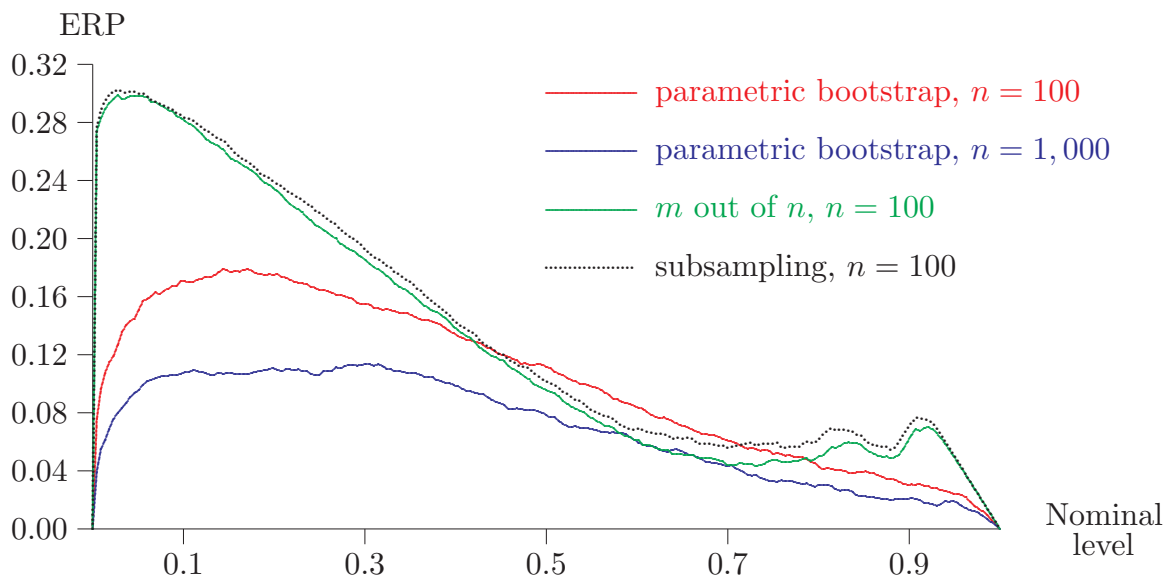


Figure 7:  $P$  value discrepancy plots, loggamma law,  $n = 100$

bootstrap DGP in the simulations are the true ones, the ERP of the parametric bootstrap is due to the very slow convergence of the distribution of  $t$  under the loggamma distribution to its distribution under the limiting stable law.

For Figure 8, the data were generated from the stable distribution, with  $\alpha = 1.1$ , 1.5 and 1.9, and  $\beta = 0$ . For bootstrapping, the symmetry of the distribution was assumed to be known, and so the skewness parameter  $\beta$  was set to zero in step 4 of the parametric bootstrap algorithm, and was not estimated in step 3. The other parameter,  $\alpha$ , was estimated by Hill's method applied to the absolute values of the data, in order to take advantage of symmetry. The parameter  $k$  was set to 33 for  $\alpha = 1.1$ , to 42 for  $\alpha = 1.5$ , and to 44 for  $\alpha = 1.9$ , these values having been determined by some preliminary experimentation to find the best choices.

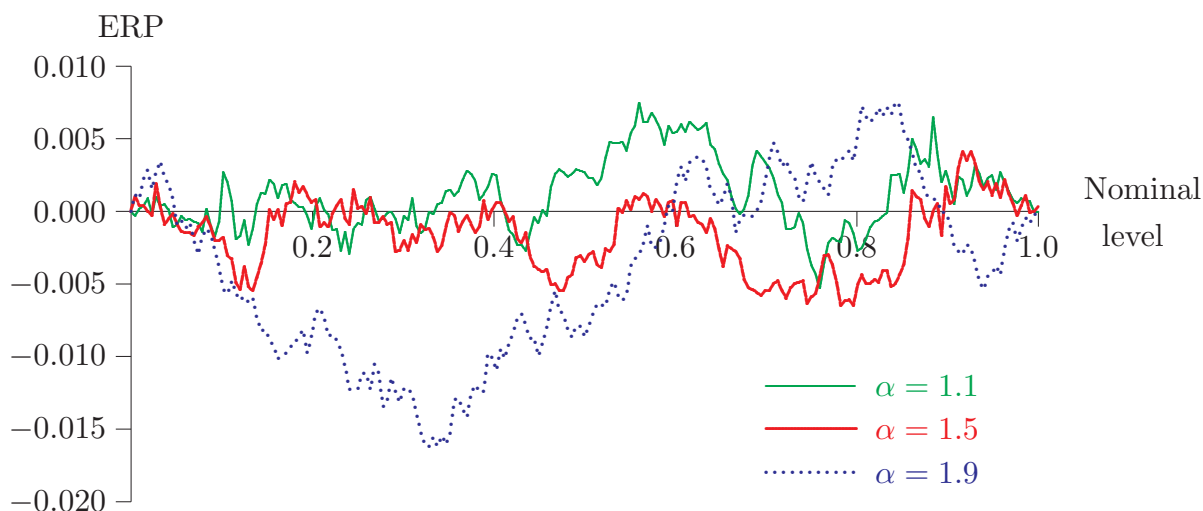


Figure 8:  $P$  value discrepancy plots,  $\hat{\alpha}_{Hill}$ , stable law,  $\beta = 0$

For Figure 9, the data came from Student's  $t$  distribution, with  $\alpha$ , which is now identified with the degrees-of-freedom parameter, set to 1.1, 1.5, and 1.9. Again the symmetry is assumed known in bootstrapping. For Hill's method, we used  $k = 30, 22,$  and  $17$  for  $\alpha = 1.1, 1.5,$  and  $1.9$ . Although the estimator of Koutrouvelis' method would be more precise when the data are generated by a stable law, it does not perform well with the Student's  $t$  data, since it is specific to the stable law. It is in any case clear from Figures 8 and 9 that the parametric bootstrap performs very well under the true null hypothesis, even though  $\alpha$  is estimated, and though the data come from a distribution which, although in the domain of normal attraction of a stable law, is not a stable law.

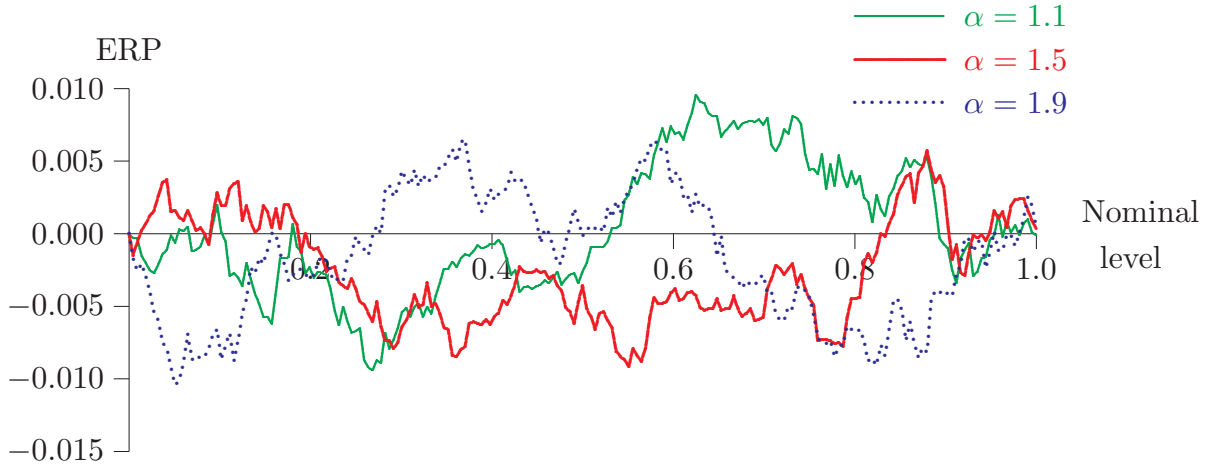


Figure 9:  $P$  value discrepancy plots,  $\hat{\alpha}_{Hill}$ , Student's  $t$

In Figure 10, the data are again from the stable law, with  $\alpha = 1.3, 1.5,$  and  $1.9$ , and the most extreme case of asymmetry, namely  $\beta = 1$ . For bootstrapping,  $\alpha$  was estimated by Koutrouvelis' method, and  $\beta$  by the quantile method. Hill's method seems to work poorly with such asymmetric data, and Koutrouvelis' method for  $\beta$  is less precise than the quantile method. We see that the performance of the parametric bootstrap is much influenced by the value of the skewness parameter  $\beta$ . Compared with the symmetric case, the ERPs are significantly larger, especially for smaller values of  $\alpha$ . Indeed, with  $\alpha = 1.1$  (not shown in the figure), the ERP is unacceptably large.

Even when the data are generated by a symmetric law, an investigator may not be prepared to assume this. Thus it is of practical importance to see how estimating  $\beta$  affects the performance of the parametric bootstrap. The parameters  $\alpha$  and  $\beta$  used in the parametric bootstrap that gave the  $P$  value discrepancy plots in Figures 11 and 12 were both estimated by Hill's method, although the true value of  $\beta$  is zero. There is little difference between the results for the stable law in Figure 11 and those for Student's  $t$  in Figure 12, but the size distortions are greater than when it is assumed that  $\beta = 0$ . They are greater the smaller is  $\alpha$ , and for  $\alpha = 1.1$ , they are again unacceptably large.

Matters can be improved a little by replacing the Hill estimate of  $\beta$  by the variant of Aban and Meerschaert, which we denote by  $\hat{\beta}_{AM}$ . Results analogous to those of

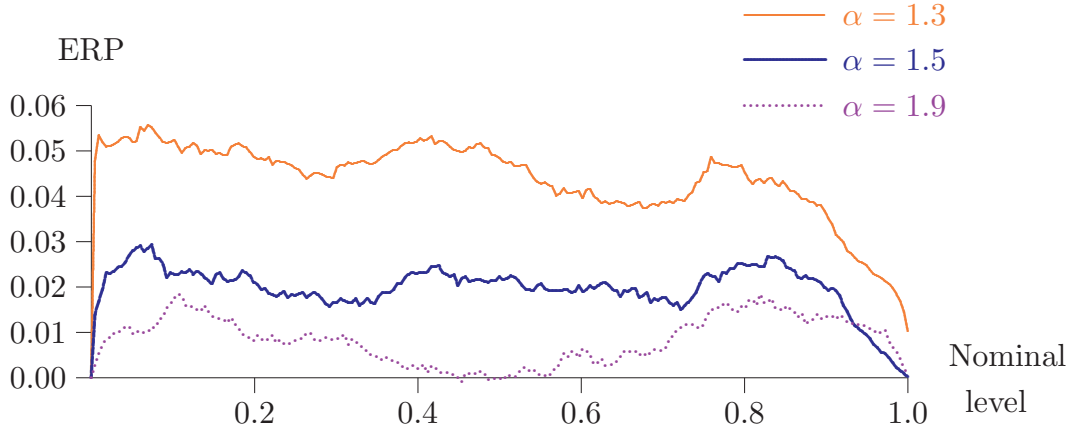


Figure 10:  $P$  value discrepancy plots,  $\hat{\alpha}_{Koutrouvelis}$ ,  $\hat{\beta}_{quantiles}$ , stable law  $\beta = 1$

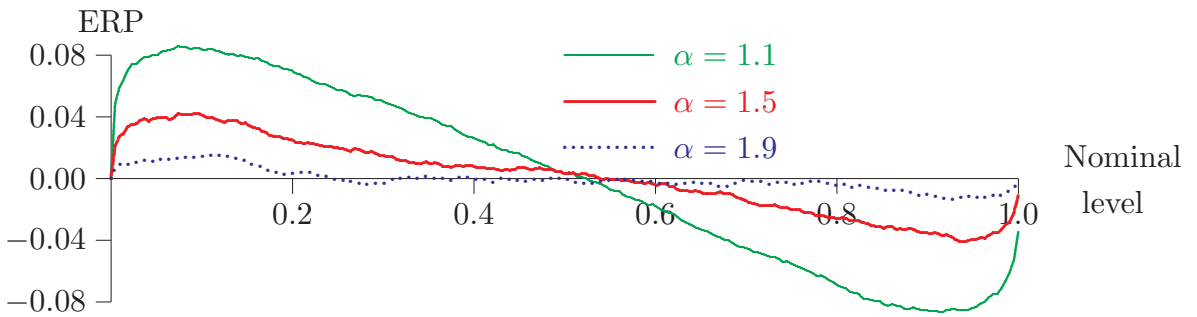


Figure 11:  $P$  value discrepancy plots,  $\hat{\alpha}_{Hill}$ ,  $\hat{\beta}_{Hill}$ , stable law

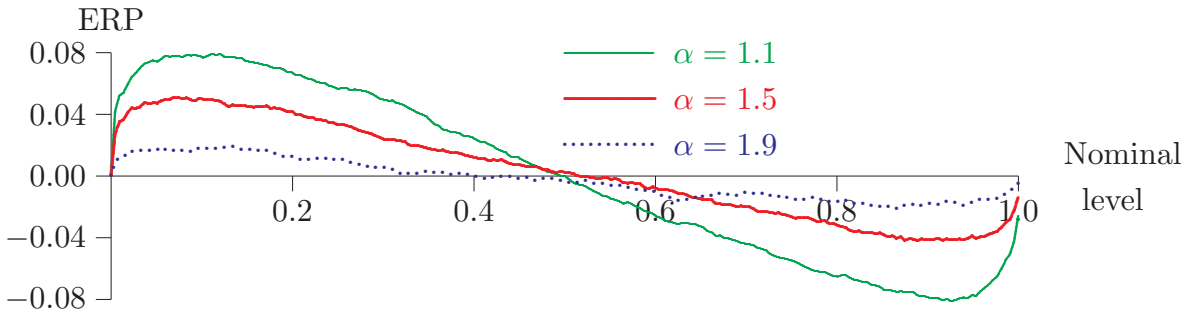


Figure 12:  $P$  value discrepancy plots,  $\hat{\alpha}_{Hill}$ ,  $\hat{\beta}_{Hill}$ , Student's  $t$  law

Figures 11 and 12 are given in Figures 13 and 14. It can be seen that the improvement is slight and that the qualitative features are the same as with Hill's estimator of  $\beta$ .

If  $\beta$  is nonzero, performance becomes still worse. The size distortions that we see in Figure 15, where the true  $\beta$  is 0.5, are enormous for small  $\alpha$ . In Figure 16, we have  $\beta = 1$ , and the distortions are even greater. For both these figures, we use  $\hat{\alpha}_{Hill}$  and  $\hat{\beta}_{AM}$ . The reason for the large distortions is that, for large  $\beta$  and  $\alpha$  close to 1, the distribution  $G_{\alpha,\beta}$  of the self-normalized sum is exceedingly sensitive to the value of  $\alpha$ . None of the estimates of  $\alpha$  that we have considered is precise enough to overcome this phenomenon. In this case, the failure of the parametric bootstrap is due less to any deficiency of its own than to the imprecision of the estimates of  $\alpha$ .

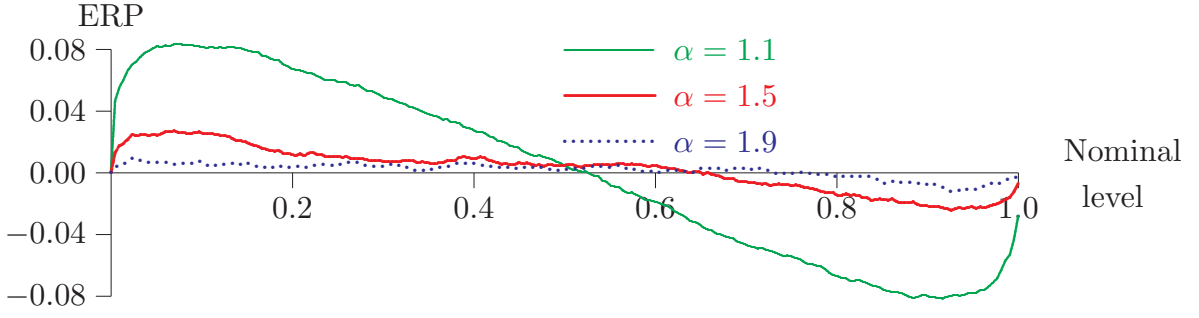


Figure 13:  $P$  value discrepancy plots,  $\hat{\alpha}_{\text{Hill}}$ ,  $\hat{\beta}_{\text{AM}}$ , stable law

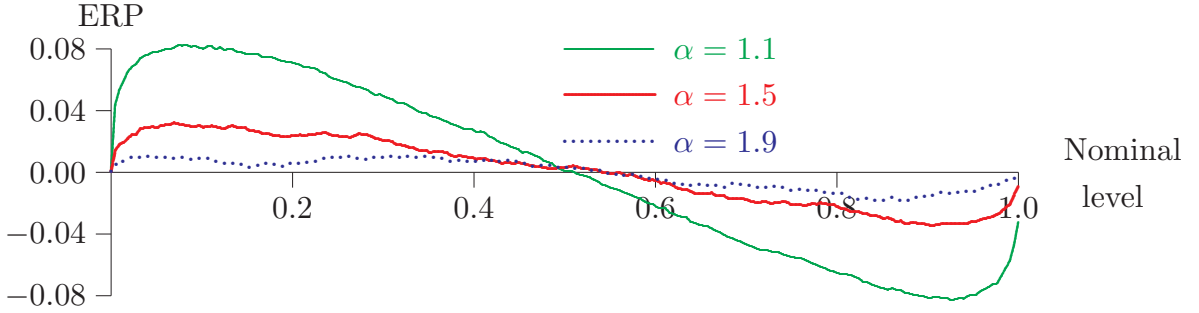


Figure 14:  $P$  value discrepancy plots,  $\hat{\alpha}_{\text{Hill}}$ ,  $\hat{\beta}_{\text{AM}}$ , Student's  $t$  law

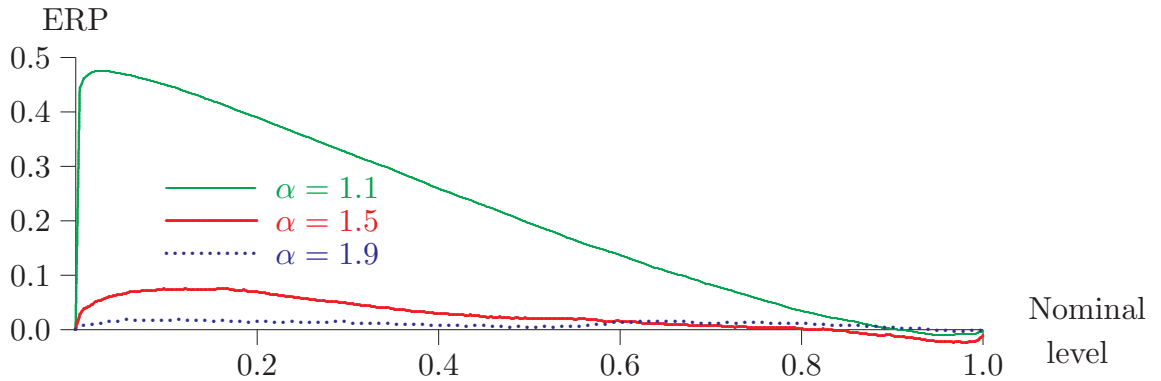


Figure 15:  $P$  value discrepancy plots,  $\hat{\alpha}_{\text{Hill}}$ ,  $\hat{\beta}_{\text{AM}}$ , stable law, true  $\beta = 0.5$

We end this section with some evidence about the behaviour of the parametric bootstrap when the null hypothesis  $\delta = 0$  is false, the true  $\delta$  being  $-0.5$ . Except for this change, the setups for Figures 17 and 18 are the same as for Figures 8 and 9. The results are shown as plots of the estimated rejection rates as a function of nominal level. Several interesting facts emerge from the figures. First, power is influenced by the tail index  $\alpha$ . The smaller  $\alpha$ , the lower the power. Second, compared with the stable distribution, the power is slightly lower when the data come from Student's  $t$ . From Figure 19, we see that power is also influenced by the skewness parameter: the smaller  $\alpha$  and the larger  $\beta$ , the lower the power of the bootstrap test. For this last figure, since for  $\alpha$  close to 1 the size distortion is considerable, we used the *true* value of  $\alpha$  for bootstrapping, so as to have a better notion of the theoretical power. The skewness

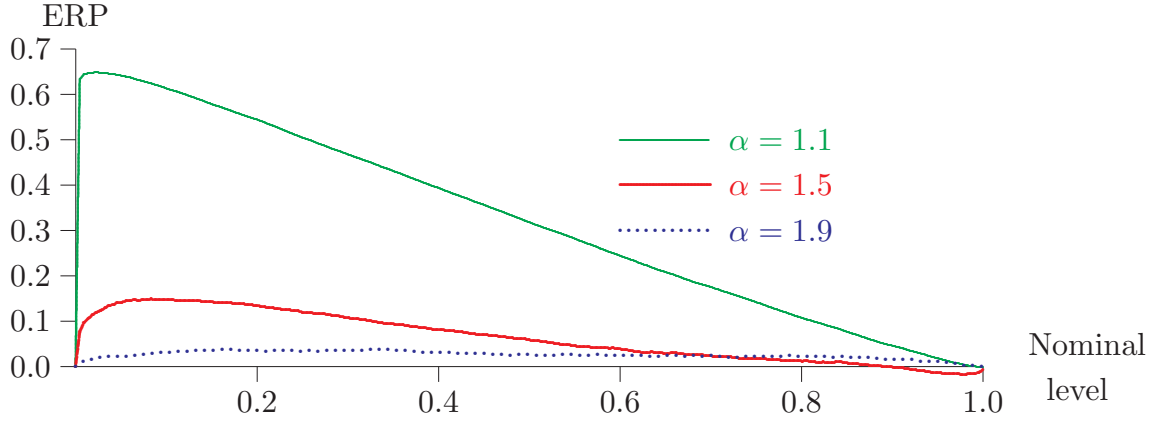


Figure 16:  $P$  value discrepancy plots,  $\hat{\alpha}_{\text{Hill}}$ ,  $\hat{\beta}_{\text{AM}}$ , stable law, true  $\beta = 1$

parameter was estimated by the quantile method. Simulation results not shown here indicate that the power of the tests based upon the subsampling method and of the  $m$  out of  $n$  bootstrap is very similar to the power of the parametric bootstrap test.

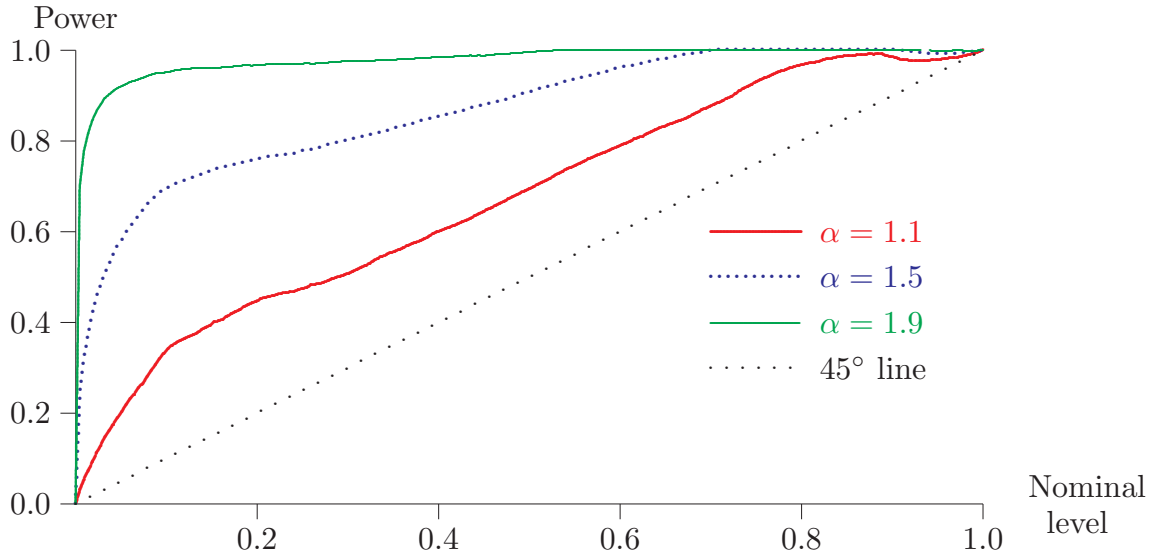


Figure 17: Power,  $\hat{\alpha}_{\text{Hill}}$ , symmetric stable law,  $\delta = -0.5$

## 9. 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. We show that, unlike the conventional nonparametric bootstrap statistic, the bootstrap  $P$  value has a non-random limiting distribution when the sample size tends to infinity. By itself, this is no guarantee that inference based on the bootstrap  $P$  value is reliable in finite samples, as we see by looking at the finite-sample performance of the  $m$  out of  $n$  and subsampling bootstraps, which are both consistent but are unreliable with a sample size of 100.



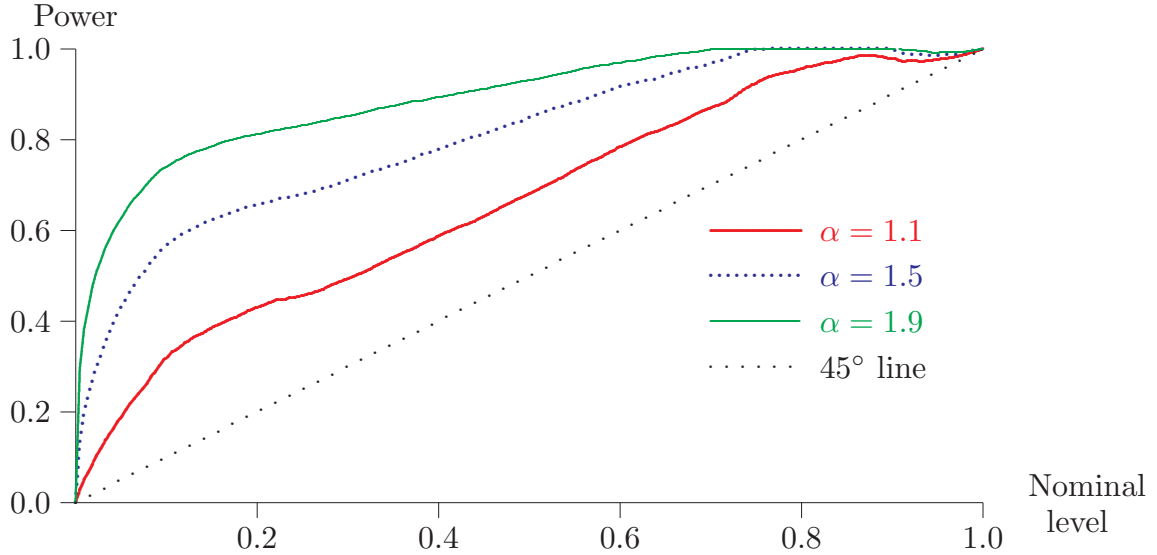


Figure 18: Power,  $\hat{\alpha}_{Hill}$ , Student's  $t$ ,  $\delta = -0.5$ .

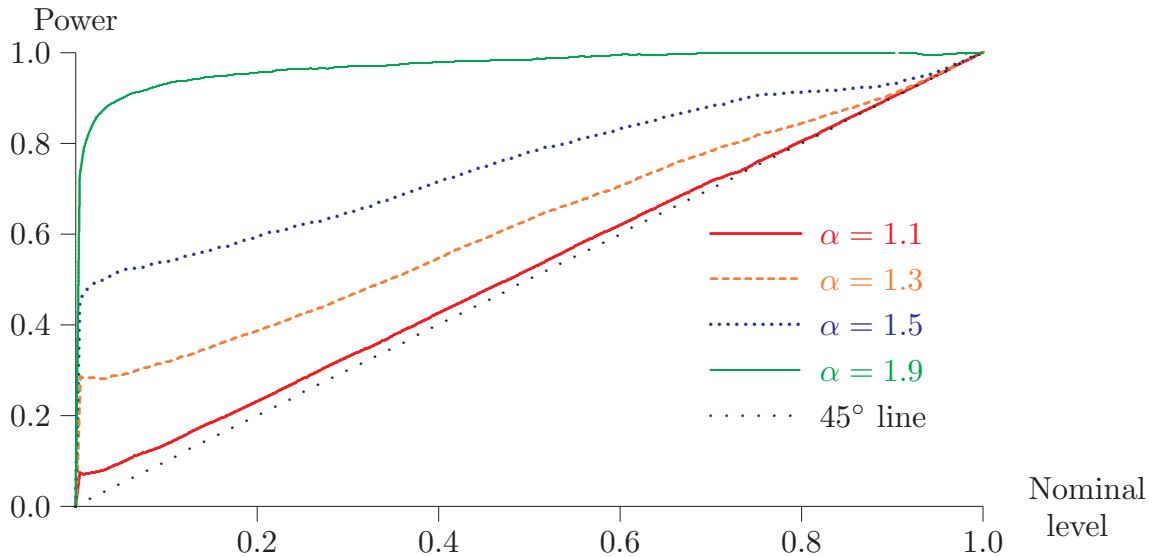


Figure 19: Power, true  $\alpha$ ,  $\hat{\beta}_{quantiles}$ , stable law,  $\beta = 1$ .

The study of the distribution of the bootstrap  $P$  value shows that, to a good approximation, it is a deterministic function of the self-normalised sum of the observations of the sample. This sum also has a nonrandom limiting distribution, and is a better candidate for bootstrapping than the bootstrap  $P$  value itself, which would require computation as intensive as a double bootstrap. The distribution of the self-normalised sum can be estimated consistently if we can estimate the parameters  $\alpha$  and  $\beta$  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  $\alpha$  and  $\beta$  is reasonably precise, the parametric bootstrap

gives inference with a sample size of 100 that is reliable by any usual standard. Its performance degrades 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  $\alpha$  close to 1, but it may be possible to find better estimators of  $\beta$ .

Moreover, we have shown that 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. The asymptotic test relies on the estimation of a scale parameter and a slowly varying function, for which it seems that no general estimation methods exist, without use of parametric assumptions about the underlying distribution.

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

## References

- Aban, I. B. and M. M. Meerschaert (2004). “Generalized least-squares estimators for the thickness of heavy-tails”, *Journal of Statistical Planning and Inference* **119**, 341–352.
- Arcones, M. A. and E. Giné, (1989). “The bootstrap of the mean with arbitrary bootstrap sample size”, *Annales de l’Institut Henri Poincaré* **25**, 457–481.
- 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”, *Annals of Statistics*, **27**, 1115–22.
- Beran, R. (1988) “Prepivoting test statistics: a bootstrap view of asymptotic refinements”, *Journal of the American Statistical Association*, **83**, 307–27.
- Bickel, P. J., Gotze, F., and van Zwet, W. R. (1997). “Resampling fewer than  $n$  observations: gains, losses and remedies for losses”, *Statistica Sinica* **7**, 1–31.
- Bickel, P. J. and Sakov, A. (2005). “On the choice of  $m$  in the  $m$  out of  $n$  bootstrap and its applications to confidence bounds for extreme percentiles”, Unpublished manuscript.
- Caers, J. and Dyck, J. V. (1999). “Nonparametric tail estimation using a double bootstrap method”, *Computational Statistics and Data Analysis* **29**, 191–211.
- Carrasco, M. and Florens, J.-P. (2000). “Generalization of GMM to a continuum of moment conditions”, *Econometric Theory* **16**, 797–834.

- Chambers, C. M., C. L. Mallows, and B. W. Stuck (1976). “A method for simulating stable random variables”, *Journal of the American Statistical Association*, **71**, 340–4.
- Csorgo, S. and Viharos, L. (1997). “Asymptotic normality of least-squares estimators of tail indices”, *Bernoulli* **3**, 351–70.
- Datta, S. and McCormick, W. P. (1995). “Bootstrap inference for a first-order autoregression with positive innovations”, *Journal of the American Statistical Association* **90**, 1289–1301.
- DuMouchel, W. H. (1973). “On the asymptotic normality of the maximum-likelihood estimate when sampling from a stable distribution”, *Annals of Statistics* **1**, 948–957.
- Efron, B. (1979). “Bootstrap methods: Another look at the jackknife”, *Annals of Statistics*, **7**, 1–26.
- Fama, E. F. and Roll, R. (1971). “Parameter estimates for symmetric stable distributions”, *Journal of the American Statistical Association* **66**, 331–338.
- Garcia, R., Renault, E., and Veredas, D. (2006). “Estimation of stable distributions by indirect inference”, CORE DP 112.
- Giné, E. and Zinn, J. (1989). “Necessary conditions for the bootstrap of the mean”, *Annals of Statistics* **17**, 684–691.
- de Haan, L. and Resnick, S. I. (1980). “A simple asymptotic estimate for the index of a stable distribution”, *Journal of the Royal Statistical Society, series B* **42**, 83–87.
- 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**, 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 Jing, B. (1998). “Comparison of the bootstrap and asymptotic approximations to the distribution of a heavy-tailed mean”, *Statistica Sinica* **8**, 887–906.
- Hall, P. and LePage, R. (1996). “On bootstrap estimation of the distribution of the studentized mean”, *Annals of the Institute of Statistical Mathematics* **48**, 403–421.
- Hall, P. and Welsh, A. H. (1985). “Adaptive estimates of parameters of regular variation”, *Annals of Statistics*, **13**, 331–41.

- Hill, B. M. (1975). “A simple general approach to inference about the tail of a distribution”, *Annals of Statistics* **3**, 1163–1174.
- Knight, K. (1989). “On the bootstrap of the sample mean in the infinite variance case”, *Annals of Statistics* **17**, 1168–1175.
- Kogon, S. M. and Williams, D. B. (1998). *A practical guide to heavy tails: statistical techniques and applications: Characteristic function based estimation of stable distribution parameters*, pp. 311–335, Boston: Birkhäuser.
- Koutrouvelis, I. A. (1980). “Regression-type estimation of the parameters of stable laws”, *Journal of American Statistical Association* **75**, 918–928.
- Lévy, P. (1925). *Calcul des Probabilités*, Paris: Gauthier-Villars.
- Logan, B. F., Mallows, C., Rice, S., and Shepp, L. A. (1973). “Limit distributions of self-normalized sums”, *Annals of Probability* **11**, 788–809.
- McCulloch, J. H. (1986). “Simple consistent estimators of stable distribution parameters”, *Communications in Statistics – Simulation and Computation* **15**, 1109–1136.
- Nolan, J. P. (2001). *Lévy Processes: Maximum Likelihood Estimation and Diagnostics for Stable Distributions*, Boston: Birkhäuser.
- Pickands, J. (1975). “Statistical inference using extreme order statistics”, *Annals of Statistics* **3**, 119–131.
- Romano, J. P. and Wolf, M. (1999). “Subsampling inference for the mean in the heavy-tailed case”, *Metrika* **50**, 55–69.
- Samorodnitsky, G. and Taqqu, M. (1994). *Stable non-gaussian random processes*, New-York: Chapman and Hall.
- Zolotarev, V. M. (1986). *One-dimensional Stable Distributions*. American Mathematical Society: Vol. 65 of Translations of mathematical monographs.