# Statistical Inference in the Presence of Heavy Tails

by

**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

Income distributions are usually characterised by a heavy right-hand tail. Apart from any ethical considerations raised by the presence among us of the very rich, statistical inference is complicated by the need to consider distributions of which the moments may not exist. In extreme cases, no valid inference about expectations is possible until restrictions are imposed on the class of distributions admitted by econometric models. It is therefore important to determine the limits of conventional inference in the presence of heavy tails, and, in particular, of bootstrap inference. In this paper, recent progress in the field is reviewed, and examples given of how inference may fail, and of the sorts of conditions that can be imposed to ensure valid inference.

Keywords: Heavy tails, inference, bootstrap, wild bootstrap

JEL codes: C10, C14, C15, I39

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 first version of this paper was prepared for the Special Session entitled "Econometrics of Inequality" sponsored by *The Econometrics Journal* at the 2010 Royal Economic Society Conference, March 2010, at the University of Surrey. I thank Stefan Bache for provoking me to develop the material in Section 7.

November 2010

## 1. Introduction

In the study of income distribution, it is more common to adopt a distribution-free approach than to impose a parametric functional form on the cumulative distribution function (CDF) of the distribution. Income distributions often have a heavy right-hand tail, and, when this is so, distribution-free approaches can lead to very unreliable inference. There are some theoretical reasons for this, one of which was explained long ago in a paper by Bahadur and Savage (1956) that is still not very widely known. Although parametric models can perform better in the presence of heavy tails, they too can have problems.

The Earth is finite, and so it is in principle impossible for any income distribution not to be bounded above; this is *a fortiori* true for any observed sample of incomes. However, it is sometimes true that the best-fitting models of income distributions imply the non-existence of some moments, sometimes including the variance, or even the expectation itself. For instance, the three-parameter Singh-Maddala distribution – Singh and Maddala (1976) – lacks some moments, just which ones depending on the parameter values. It is therefore pertinent to take account of the possible non-existence of moments in the study of income distribution.

In this paper, I review work in which it is seen that heavy tails pose problems, and in which some efforts are made to avoid these problems. In the next section, I state the main theorem of the previously cited paper of Bahadur and Savage, and give a concrete example of the impossibility result of the theorem. It is pointed out that the problem arises in models where the mapping from the set of distributions that constitute the model to the moment of interest is not continuous. Section 3 reviews some recent work of mine on the Gini index, where it appears that heavy tails undermine reliability of inference. There, the bootstrap turns out to be useful in at least mitigating this difficulty. In Section 4, a suggestion made in Davidson and Flachaire (2007) is reviewed, whereby a conventional resampling bootstrap can be combined with a parametric bootstrap in the right-hand tail. This bootstrap works well with distributions similar to those in developed countries, but its performance degrades when the second or somewhat higher moment of the distribution does not exist. Then, in Section 5, I sketch so-far unpublished recent work by Adriana Cornea and me, available as the discussion paper Cornea and Davidson (2008), in which we propose a purely parametric bootstrap for distributions in the domain of attraction of a non-Gaussian stable law. This bootstrap works reasonably well provided that the distribution has a moment higher than the first (the variance never exists) and is not too skewed.

Sections 6 and 7 deal with methods based on quantiles. For obvious reasons, these are much less disturbed by heavy tails than methods based on moments. Section 6 presents a family of measures of goodness of fit, based on a measure of distribution change proposed by Cowell (1985) – see also Cowell, Flachaire, and Bandyopadhyay (2009) – but adapted to use quantiles instead of moments. The bootstrap turns out to give very reliable inference with this new measure in circumstances that made inference for the Gini index unreliable. Finally, in Section 7, I discuss some speculative work in which the wild bootstrap is applied to quantile regression.

## 2. The Result of Bahadur and Savage

The paper of Bahadur and Savage (1956) (BS) contains a number of impossibility results about inference on the expectation of a distribution based on an IID sample drawn from it. The thrust of all the results is that, unless some restrictions, over and above the mere existence of the expectation of the distribution, are placed on the class of distributions that constitute the model, such inference is impossible. Impossible in the sense that the size of a test of a specific value for the expectation is independent of the significance level, and that no valid confidence intervals exist.

The model for which these impossibility results hold must be reasonably general, and the precise regularity conditions made by Bahadur and Savage are as follows. Each DGP of the model is characterised by a CDF, $F$ say. The class $\mathcal{F}$ of those $F$ that the model contains is such that

(i) For all $F \in \mathcal{F}$, $\mu_F \equiv \int_{-\infty}^{\infty} x \, \mathrm{d}F(x)$ exists and is finite;

(ii) For every real number $m$, there is $F \in \mathcal{F}$ with $\mu_F = m$;

(iii) $\mathcal{F}$ is convex.

Let $\mathcal{F}_m$ be the subset of $\mathcal{F}$ for which $\mu_F = m$. Then Bahadur and Savage prove the following theorem.

### Theorem 1

For every bounded real-valued function $\phi$ defined on the sample space (that is, $\mathbb{R}^n$ for a sample of size $n$), the quantities $\inf_{F \in \mathcal{F}_m} \mathrm{E}\phi$ and $\sup_{F \in \mathcal{F}_m} \mathrm{E}\phi$ are independent of $m$.

From this, the main results of their paper can be derived. The argument is based on the fact that the mapping from $\mathcal{F}$, endowed with the topology of weak convergence, to the real line, with the usual topology, that maps a CDF $F$ to its expectation $\mu_F$ is not continuous.[1]

Rather than work at the high level of generality of BS's paper, I present a one-parameter family of distributions, all with zero expectation. If an IID sample of size $n$ is drawn from a distribution that is a member of this family, one can construct the usual $t$ statistic for testing whether the expectation of the distribution is zero. I will then show that, for *any* finite critical value, the probability that the $t$ statistic exceeds that value tends to one as the parameter of the family tends to zero. It follows that, if all the DGPs of the sequence are included in $\mathcal{F}$, the $t$ test has size one.

[1] Actually, Bahadur and Savage use a seemingly different topology, based on the metric of **absolute-variational distance**, defined for two CDFs $F$ and $G$ as

$$\delta(F, G) = \sup_{\phi \in \Phi} |\mathrm{E}_F\phi - \mathrm{E}_G\phi|,$$

where $\Phi$ is the set of real-valued functions of the sample space taking values in the interval $[0, 1]$.

Each distribution in the family is characterised by a parameter $p$ with $0 < p < 1$. A random variable from the distribution can be written as

$$U = Y/p^2 + (1 - Y)W - 1/p \tag{1}$$

where $W \sim \mathrm{N}(0, 1)$ and

$$Y = \begin{cases} 1 & \text{with probability } p \\ 0 & \text{with probability } 1 - p, \end{cases}$$

$W$ and $Y$ being independent. It is evident that $\mathrm{E}_p U = 0$.

Now consider a sample of IID drawings $U_t$, each from the above distribution for given $p$. Let $N$ be $\sum_{t=1}^n Y_t$. The value of $N$ is thus the number of drawings with value $(1 - p)/p^2$. We see that

$$\Pr(N = 0) = (1 - p)^n. \tag{2}$$

The $t$ statistic for a test of the hypothesis that $\mathrm{E}\,U = 0$ can be written as

$$T = \frac{\hat{\mu}}{\hat{\sigma}_\mu}, \text{ where } \hat{\mu} = \frac{1}{n}\sum_{t=1}^n U_t, \text{ and } \hat{\sigma}_\mu^2 = \frac{1}{n(n-1)}\sum_{t=1}^n (U_t - \hat{\mu})^2.$$

Conditional on $N = 0$, $\hat{\mu} = -1/p + \bar{W}$, where $\bar{W} = n^{-1}\sum W_t$ is the mean of the $W_t$. Thus the conditional distribution of $n^{1/2}\hat{\mu}$ is $\mathrm{N}(-n^{1/2}/p, 1)$. Then, since $U_t - \hat{\mu} = W_t - \bar{W}$ if $N = 0$, the conditional distribution of $n\hat{\sigma}_\mu^2$ is $\chi^2_{n-1}/(n - 1)$. Consequently, the conditional distribution of $T$ is noncentral $t_{n-1}$, with noncentrality parameter $-n^{1/2}/p$. We can compute as follows for $c > 0$:

$$\Pr(|T| > c) > \Pr(T < -c) > \Pr(T < -c \text{ and } N = 0) = \Pr(N = 0) \Pr(T < -c \,|\, N = 0). \tag{3}$$

Now

$$\Pr(T < -c \,|\, N = 0) = F_{n-1,-n^{1/2}/p}(-c), \tag{4}$$

where $F_{n-1,-n^{1/2}/p}$ is the CDF of noncentral $t$ with $n - 1$ degrees of freedom and noncentrality parameter $-n^{1/2}/p$.

For fixed $c$ and $n$, let $p \to 0$. Then from (2) we see that $\Pr(N = 0) \to 1$. From (4), it is clear that $\Pr(T < -c \,|\, N = 0)$ also tends to 1, since the noncentrality parameter tends to $-\infty$, which means that the probability mass to the left of any fixed value tends to 1. It follows from (3) that the rejection probability tends to 1 whatever the critical value $c$, and so the test has size 1 if DGPs characterised by random variables distributed according to (1) are admitted to the null hypothesis. A similar, more complicated, argument shows that a test based on a resampling bootstrap DGP delivers a $P$ value that, in a sample of size $n$, tends to $n^{-(n-1)}$ as $p \to 0$, and so the size of this bootstrap test tends to $1 - n^{-(n-1)}$, regardless of the significance level.

Since the distribution (1) has all its moments finite for positive $p$, imposing conditions on the existence of moments does not prevent all the distributions characterised by (1) from being present in the null model, with the unfortunate consequences predicted by the theorem of BS. As remarked above, the problem is essentially due to the fact that the expectation is not a continuous function of the distribution with the usual topologies. In the example with the distributions given by (1), we see that for $p > 0$, the first moment is zero. In the limit, however, it is infinite.

If valid inference is to be possible, we must impose a restriction that excludes the distributions (1) with small $p$. We are therefore led to consider a *uniform* bound on some moment of order at least 1. Such a bound implies a bound on the absolute first moment as well, but it seems necessary to bound a moment of order strictly greater than 1. We want to show that such a bound renders the mapping from $\mathcal{F}$ to the expectation continuous. Suppose then, that $\mathcal{F}$ is restricted so as to contain only distributions such that, for some $\theta > 0$, $\mathrm{E}|U|^{1+\theta} < K$, for some specified $K$. The proof of Lemma 1 in the Appendix shows that this restriction is enough to make the mapping continuous. Note, however, that in order to compute the size of a test about the expectation, the actual, numerical, values of $K$ and $\theta$ must be known.

## 3. Illustration with the Gini Index

Most of this section is borrowed from one of my recent papers, Davidson (2009), in which I develop methods for performing inference, both asymptotic and bootstrap, for the Gini index. The methods rely on the assumption that the estimation error of the sample Gini, divided by its standard error, is asymptotically standard normal. In order to see whether the asymptotic normality assumption yields a good approximation, simulations were undertaken with drawings from the exponential distribution, with CDF $F(x) = 1 - \mathrm{e}^{-x}$, $x \geq 0$. The true value $G_0$ of the Gini index for this distribution is easily shown to be one half. In Figure 1, graphs are shown of the EDF of 10,000 realisations of the statistic $\tau = (\hat{G} - G_0)/\hat{\sigma}_G$, using the bias-corrected version of $\hat{G}$ and the standard error $\hat{\sigma}_G$ derived in Davidson (2009), for sample sizes $n = 10$ and 100. The graph of the standard normal CDF is also given as a benchmark.

It can be seen that, even for a very small sample size, the asymptotic standard normal approximation is good. The greatest absolute differences between the empirical distributions of the $\tau$ and the standard normal CDF were 0.0331 and 0.0208 for $n = 10$ and $n = 100$ respectively.

The exponential distribution may well be fairly characteristic of distributions encountered in practice, but its tail is not heavy. Heavy-tailed distributions are notorious for causing problems for both asymptotic and bootstrap inference, and so in Figure 2 we show empirical distributions for the standardised statistic $\tau$ with data generated by the Pareto distribution, of which the CDF is $F_{\text{Pareto}}(x) = 1 - x^{-\lambda}$, $x \geq 1$, $\lambda > 1$. The second moment of the distribution is $\lambda/(\lambda - 2)$, provided that $\lambda > 2$, so that, if $\lambda \leq 2$, no reasonable inference about the Gini index is possible. If $\lambda > 1$, the true Gini index is $1/(2\lambda - 1)$. Plots of the distribution of $\tau$ are shown in Figure 2 for $n = 100$ and $\lambda = 100, 5, 3, 2$. For values of $\lambda$ greater than about 50, the distribution does not change much, which implies that there is a distortion of the standard error with the heavy tail even if the tail index is large.

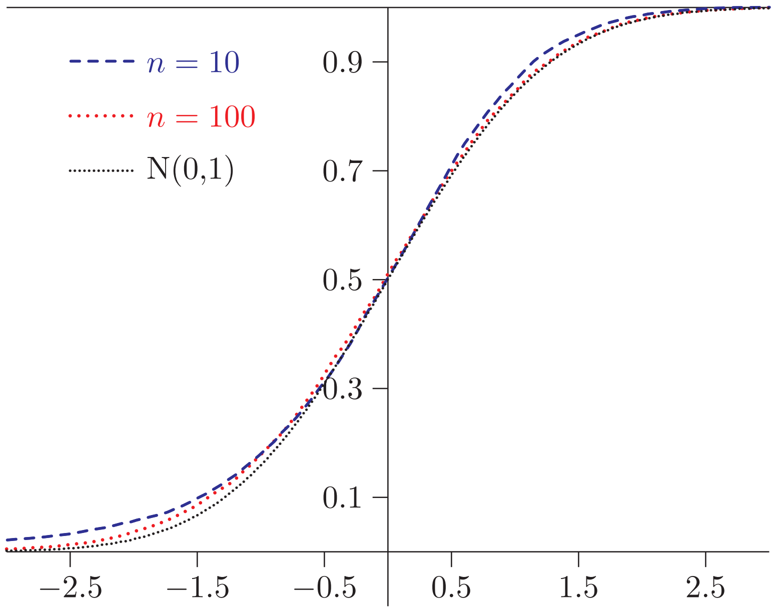


**Figure 1. Distribution of the standardised statistic; exponential distribution**

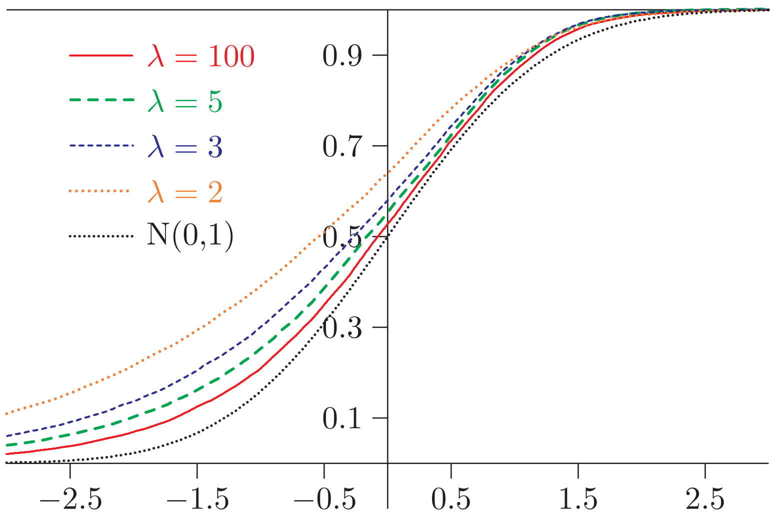


**Figure 2. Distribution of the standardised statistic; Pareto distribution, $n$=100**

Table 1 shows how the bias of $\tau$, its variance, and the greatest absolute deviation of its distribution from standard normal vary with $\lambda$. It is plain from the table that the usual difficulties with heavy-tailed distributions are just as present here as in other circumstances.

| $\lambda$ | Bias | Variance | Divergence from N(0,1) |
|---|---|---|---|
| 100 | -0.1940 | 1.3579 | 0.0586 |
| 20 | -0.2170 | 1.4067 | 0.0647 |
| 10 | -0.2503 | 1.4798 | 0.0742 |
| 5 | -0.3362 | 1.6777 | 0.0965 |
| 4 | -0.3910 | 1.8104 | 0.1121 |
| 3 | -0.5046 | 2.1011 | 0.1435 |
| 2 | -0.8477 | 3.1216 | 0.2345 |

Table 1. Summary statistics for Pareto distribution

The lognormal distribution is not usually considered as heavy-tailed, since it has all its moments. It is nonetheless often used in the modelling of income distributions. Since the Gini index is scale invariant, we consider only lognormal variables of the form $e^{\sigma W}$, where $W$ is standard normal. In Figure 3 the distribution of $\tau$ is shown for $n = 100$ and $\sigma = 0, 5, 1.0, 1.5$. We can see that, as $\sigma$ increases, distortion is about as bad as with the genuinely heavy-tailed Pareto distribution. The comparison is perhaps not entirely fair, since, even for the worst case with $\lambda = 2$ for the Pareto distribution, $G = 1/3$. However, for $\sigma = 1$, the index for the lognormal distribution is 0.521, and for $\sigma = 1.5$ there is a great deal of inequality, with $G = 0.711$.

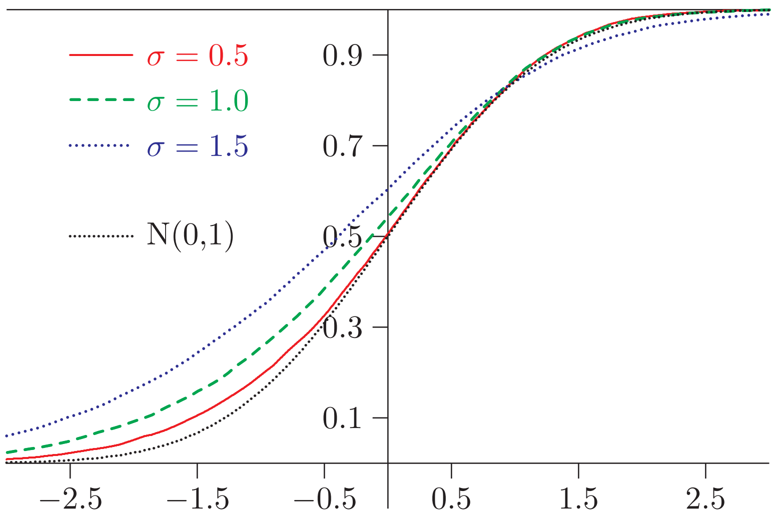


**Figure 3. Distribution of $\tau$; lognormal distribution, $n$=100**

We end this section with some evidence about the behaviour of the bootstrap. In Table 2, coverage rates of percentile-$t$ bootstrap confidence intervals are given for $n = 100$ and for nominal confidence levels from 90% to 99%. The successive rows of the table correspond, first, to the exponential distribution, then to the Pareto distribution for $\lambda = 10, 5, 2$, and finally to the lognormal distribution for $\sigma = 0.5, 1.0, 1.5$. The numbers are based on 10,000 replications with 399 bootstrap repetitions each.

| Level | 90% | 92% | 95% | 97% | 99% |
|---|---|---|---|---|---|
| Exponential | 0.889 | 0.912 | 0.943 | 0.965 | 0.989 |
| $\lambda = 10$ | 0.890 | 0.910 | 0.942 | 0.964 | 0.984 |
| $\lambda = 5$ | 0.880 | 0.905 | 0.937 | 0.957 | 0.982 |
| $\lambda = 2$ | 0.831 | 0.855 | 0.891 | 0.918 | 0.954 |
| $\sigma = 0.5$ | 0.895 | 0.918 | 0.949 | 0.969 | 0.989 |
| $\sigma = 1.0$ | 0.876 | 0.898 | 0.932 | 0.956 | 0.981 |
| $\sigma = 1.5$ | 0.829 | 0.851 | 0.888 | 0.914 | 0.951 |

Table 2. Coverage of percentile-$t$ confidence intervals

Apart from the expected serious distortions when $\lambda = 2$, and when $\sigma = 1.5$, the coverage rate of these confidence intervals is remarkably close to nominal. It seems that, unless the tails are very heavy indeed, or the Gini index itself large, the bootstrap can yield acceptably reliable inference in circumstances in which the asymptotic distribution does not.

## 4. Measures of Inequality

In this section, largely borrowed from Davidson and Flachaire (2007), we consider a bootstrap DGP which combines a parametric estimate of the upper tail with a nonparametric estimate of the rest of the distribution. This approach is based on finding a parametric estimate of the index of stability of the right-hand tail of the income distribution. The approach is inspired by the paper by Schluter and Trede (2002), in which they make use of an estimator proposed by Hill (1975) for the index of stability. The estimator is based on the $k$ greatest order statistics of a sample of size $n$, for some integer $k \leq n$. If we denote the estimator by $\hat{\alpha}$, it is defined as follows:

$$\hat{\alpha} = H_{k,n}^{-1}; \qquad H_{k,n} = k^{-1} \sum_{i=0}^{k-1} \log Y_{(n-i)} - \log Y_{(n-k+1)}, \tag{5}$$

where $Y_{(j)}$ is the $j^{\text{th}}$ order statistic of the sample. The estimator (5) is the maximum likelihood estimator of the parameter $\alpha$ of the Pareto distribution with tail behaviour of the CDF like $1 - cy^{-\alpha}$, $c > 0$, $\alpha > 0$, but is applicable more generally; see Schluter and Trede (2002). Modelling upper tail distributions is not new in the literature on extreme value distribution, a good introduction to this work is Coles (2001).

The choice of $k$ is a question of trade-off between bias and variance. If the number of observations $k$ on which the estimator $\hat{\alpha}$ is based is too small, the estimator is very noisy, but if $k$ is too great, the estimator is contaminated by properties of the distribution that have nothing to do with its tail behaviour. A standard approach consists of plotting $\hat{\alpha}$ for different values of $k$, and selecting a value of $k$ for which the parameter estimate $\hat{\alpha}$ does not vary significantly, see Coles (2001) and Gilleland and Katz (2005). Experiments with this graphical method for samples of different sizes $n = 100, 500, 1000, 2000, 3000, 4000, 5000$, led us to choose $k$ to be the square root of the sample size: the parameter estimate $\hat{\alpha}$ is stable with this choice and it satisfies the requirements that $k \to \infty$ and $k/n \to 0$ as $n \to \infty$. The observations in the experiments were drawn from the Singh-Maddala distribution, with CDF

$$F(y) = 1 - \frac{1}{(1 + ay^b)^c} \tag{6}$$

and parameter values $a = 100$, $b = 2.8$, $c = 1.7$, a choice that closely mimics the net income distribution of German households, apart from a scale factor. Note that the automatic choice of $k$ is an area of active research; for instance Caers and Van Dyck (1999) proposed an adaptive procedure based on a $m$ out of $n$ bootstrap method.

Bootstrap samples are drawn from a distribution defined as a function of a probability mass $p_{\text{tail}}$ that is considered to constitute the tail of the distribution. Each observation of a bootstrap sample is, with probability $p_{\text{tail}}$, a drawing from the distribution with CDF

$$F(y) = 1 - (y/y_0)^{-\hat{\alpha}}, \quad y > y_0, \tag{7}$$

where $y_0$ is the order statistic of rank $n(1 - p_{\text{tail}})$ of the sample, and, with probability $1 - p_{\text{tail}}$, a drawing from the empirical distribution of the sample of smallest $n(1 - p_{\text{tail}})$ order statistics. Thus this bootstrap is just like the ordinary resampling bootstrap for all but the right-hand tail, and uses the distribution (7) for the tail. If $\hat{\alpha} < 2$, this means that variance of the bootstrap distribution is infinite.

Suppose that we wish to perform inference on some index of inequality that depends sensitively on the details of the right-hand tail. In order for the bootstrap statistics to test a true null hypothesis, we must compute the value of the index for the bootstrap distribution defined above. Indices of interest are functionals of the income distribution. Denote by $T(F)$ the value of the index for the distribution with CDF $F$. The estimate of the index from an IID sample is then $T(\hat{F})$, where $\hat{F}$ is the empirical distribution function of the sample. The CDF of the bootstrap distribution can be written as

$$F_{\text{bs}}(y) = \frac{1}{n} \sum_{i=1}^{n(1-p_{\text{tail}})} \mathrm{I}(Y_{(i)} \le y) + \mathrm{I}(y \ge y_0) p_{\text{tail}} \big(1 - (y/y_0)^{-\hat{\alpha}}\big),$$

where I is the indicator function, and $Y_{(i)}$ is order statistic $i$ from the sample. From this the index for the bootstrap distribution, $T(F_{\text{bs}})$, can be computed.

It is desirable in practice to choose $p_{\text{tail}}$ such that $np_{\text{tail}}$ is an integer, but this is not absolutely necessary. In our simulations, we set $p_{\text{tail}} = hk/n$, for $h = 0.3, 0.4, 0.6, 0.8$, and 1.0. Results suggest that the best choice is somewhere in the middle of the explored range, but a more detailed study of the optimal choice of $p_{\text{tail}}$ remains for future work. The bootstrap procedure is set out as an algorithm below.

### Semiparametric Bootstrap Algorithm

In order to test the hypothesis that the true value of the index is equal to $T_0$:

1. With the original sample, of size $n$, compute the index of interest, $\hat{T}$, and the $t$-type statistic
$$W = (\hat{T} - T_0)/[\hat{V}(\hat{T})]^{1/2}, \tag{8}$$
where $\hat{V}(\hat{T})$ denotes a variance estimate, usually based on asymptotic theory.
2. Select $k$ with graphical or adaptive methods, select a suitable value for $h$, set $p_{\text{tail}} = hk/n$, and determine $y_0$ as the order statistic of rank $n(1 - p_{\text{tail}})$ from the sample.
3. Fit a Pareto distribution to the $k$ largest incomes, with the estimator $\hat{\alpha}$ defined in (5). Compute the true value of the index, $T_0^*$, for the bootstrap distribution as $T(F_{\text{bs}})$.
4. Generate a bootstrap sample as follows: construct $n$ independent Bernoulli variables $X_i^*$, $i = 1, \ldots, n$, each equal to 1 with probability $p_{\text{tail}}$ and to 0 with probability $1-p_{\text{tail}}$. The income $Y_i^*$ of the bootstrap sample is a drawing from the distribution (7) if $X_i = 1$, and a drawing from the empirical distribution of the $n(1 - p_{\text{tail}})$ smallest order statistics $Y_{(j)}$, $j = 1, \ldots, n(1 - p_{\text{tail}})$, if $X_i = 0$.
5. With the bootstrap sample, compute the index $\hat{T}^\star$, its variance estimate $\hat{V}(\hat{T}^\star)$, and the bootstrap statistic $W^\star = (\hat{T}^\star - T_0^*)/[\hat{V}(\hat{T}^\star)]^{1/2}$.
6. Repeat steps 4 and 5 $B$ times, obtaining the bootstrap statistics $W_j^\star, j = 1, \ldots, B$. The bootstrap $P$-value is computed as the proportion of $W_j^\star, j = 1, \ldots, B$, that are smaller than $W$.

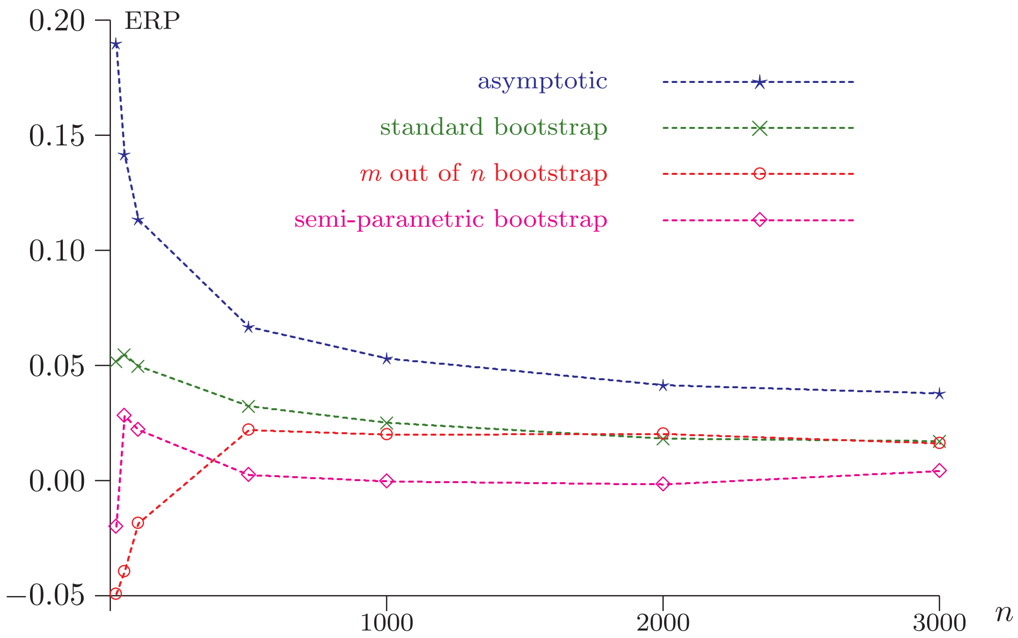


**Figure 4. ERPs in left-hand tail**

In Figure 4, the errors in rejection probability (ERPs) in the left-hand tail are plotted for tests at a nominal significance level of 5%; the asymptotic test, the standard percentile-$t$ bootstrap, the $m$ out of $n$ bootstrap, and the semi-parametric bootstrap just described, with $h = 0.4$. The index used was Theil's index

$$T(F) = \int \frac{y}{\mu_F} \log\left(\frac{y}{\mu_F}\right) \mathrm{d}F(y).$$

Figure 5 shows comparable results for the right-hand tail.

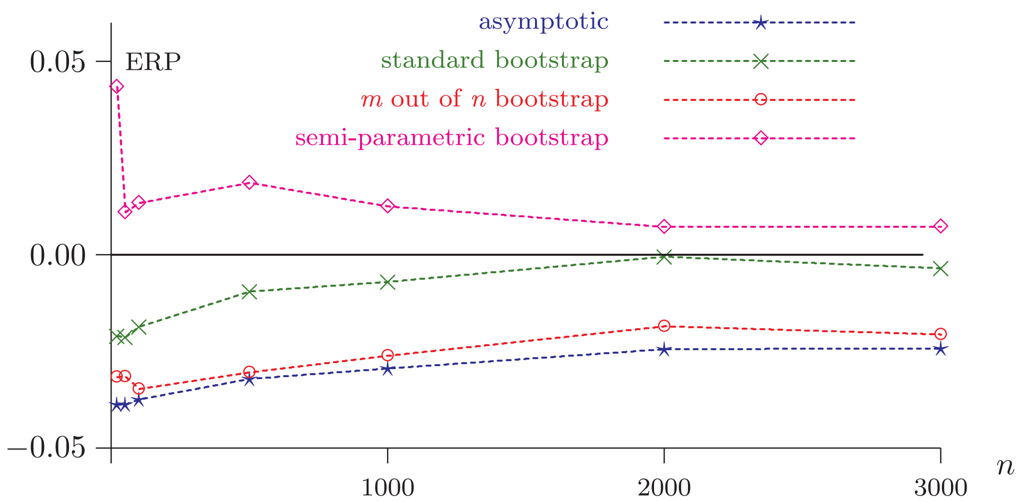


**Figure 5. ERPs in right-hand tail**

Some rather straightforward conclusions can be drawn from these Figures. In the troublesome left-hand tail, the $m$ out of $n$ bootstrap provides some slight improvement over the standard percentile-$t$ bootstrap, notably by converting the overrejection for small sample sizes to underrejection. For larger samples, the performances of the standard and $m$ out of $n$ bootstraps are very similar. The semi-parametric bootstrap, on the other hand, provides a dramatic reduction in the ERP for all sample sizes considered, the ERP never exceeding 0.033 for a sample size of 50. In the much better-behaved right-hand tail, both the $m$ out of $n$ and semi-parametric bootstraps perform worse than the standard bootstrap, although their ERPs remain very modest for all sample sizes. This less good performance is probably due to the extra noise they introduce relative to the standard bootstrap.

### Heavier tails

Although the bootstrap distribution of the statistic $W$ of (8) converges to a random distribution when the variance of the income distribution does not exist, it is still possible that at least one of the bootstrap tests we have considered may have correct asymptotic behaviour, if, for instance, the rejection probability averaged over the random bootstrap distribution tends to the nominal level as $n \to \infty$. We do not pursue this question here.

Finite-sample behaviour, however, is easily investigated by simulation. In Table 3, we show the ERPs in the left and right-hand tails at nominal level 0.05 for all the procedures considered, for sample size $n = 100$, for two sets of parameter values. These are, first, $b = 2.1$ and $c = 1$, with index of stability $\alpha = 2.1$, and, second, $b = 1.9$ and $c = 1$, with index $\alpha = 1.9$. In the first case, the variance of the income distribution exists; in the second it does not.

| | asymptotic | std bootstrap | $m$ out of $n$ | semi-parametric |
|---|---|---|---|---|
| $b = 2.1, c = 1$ | 0.41 | 0.24 | 0.15 | 0.16 |
| | -0.03 | -0.04 | -0.03 | 0.04 |
| $b = 1.9, c = 1$ | 0.48 | 0.28 | 0.20 | 0.18 |
| | -0.03 | -0.04 | -0.02 | 0.06 |

Table 3. ERPs for very heavy tails: left above, right below

Although the variance estimate in the denominator of (8) is meaningless if the variance does not exist, we see from the Table that the ERPs seem to be continuous across the boundary at $\alpha = 2$. This does not alter the fact that the ERPs in the left-hand tail are unacceptably large for all procedures.

## 5. A Parametric Bootstrap for the Domains of Stable Laws

This section is borrowed from Cornea and Davidson (2008). In that paper, we develop a procedure for bootstrapping the mean of distributions in the domain of attraction of a stable law. We show that the $m$ out of $n$ bootstrap is no better in this context than in that of the preceding section, and that subsampling barely helps. The formal results that show that both of these procedures are asymptotically valid seem to apply only for exceedingly large samples; beyond anything in our simulations.

The stable laws, introduced by Lévy (1925), are the only possible limiting laws for suitably centred 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 characterised 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.

Suppose we wish to test the hypothesis $\delta = 0$ in the model

$$Y_j = \delta + U_j, \quad \mathrm{E}(U_j) = 0, \quad j = 1, \ldots, n. \tag{9}$$

We suppose that the disturbances $U_j$ follow a distribution in the domain of attraction of a stable law $cS(\alpha, \beta)$ with location parameter 0. When $1 < \alpha \leq 2$, the parameter $\delta$ in model (9) can be consistently estimated by the sample mean. A possible test statistic is

$$\tau = n^{-1/\alpha} \sum_{j=1}^{n} Y_j. \tag{10}$$

By the Generalised 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 realisation 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, rather than a distribution in the domain of attraction, inference based on this $P$ value may be unreliable in finite samples.

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 to a deterministic distribution as the sample size $n \to \infty$. 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 by Knight (1989).

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

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

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

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

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

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

$$p(\boldsymbol{Y}) = \mathrm{E}^*\big(\mathrm{I}(\tau^* < \tau)\big) = \mathrm{E}\big(\mathrm{I}(\tau^* < \tau) \bigm| \boldsymbol{Y}\big), \tag{11}$$

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

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

where the $M_j$ are $n$ independent Poisson random variables with expectation one. The cumulant-generating function (cgf) of the distribution of $\tau(M)$ is

$$\sum_{j=1}^{n}\bigl\{\exp\bigl(\mathrm{i}tn^{-\gamma}(Y_j - \bar{Y})\bigr) - 1\bigr\} \tag{12}$$

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. Note that the function (12) is random, because it depends on the $Y_j$.

It follows that the distribution of the self-normalised sum

$$t(M) \equiv \frac{\sum_{j=1}^{n}(Y_j - \bar{Y})(M_j - 1)}{\bigl(\sum_{j=1}^{n}(Y_j - \bar{Y})^2\bigr)^{1/2}} \tag{13}$$

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

Let $F_{\boldsymbol{Y}}^{n}$ denote the random CDF of $t(M)$. Then, from (11) with $\tau^*$ replaced by $\tau(M)$, we have

$$p(\boldsymbol{Y}) = F_{\boldsymbol{Y}}^{n}\Bigl(\frac{\sum_{j=1}^{n}Y_j}{\bigl(\sum_{j=1}^{n}(Y_j - \bar{Y})^2\bigr)^{1/2}}\Bigr). \tag{14}$$

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 (12) have a nonrandom limit as $n \to \infty$? and

(ii) Does the distribution of the bootstrap $P$ value $p(\boldsymbol{Y})$ have a well-defined limit as $n \to \infty$?

If question (i) has a positive answer, then the cgf (12) 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 (1990). The distribution with cgf (12) 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 $p(\boldsymbol{Y})$ is nonrandom, since $p(\boldsymbol{Y})$ is the deterministic measurable function of $\boldsymbol{Y}$ given by (14). The question is whether the distribution converges to a limiting distribution as $n \to \infty$. 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

$$t \equiv \frac{\sum_{j=1}^{n} Y_j}{\left(\sum_{j=1}^{n}(Y_j - \bar{Y})^2\right)^{1/2}} \tag{15}$$

that appears in (14) has a limiting distribution when $n \to \infty$. 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 $F^n_{\boldsymbol{Y}}$ have a limiting joint distribution, and this can be shown by a straightforward extension of the proof in Logan *et al.*. This is what we need to conclude that the bootstrap $P$ value does indeed have a limiting distribution as $n \to \infty$. Of course, asymptotic inference is possible only if we know what that limiting distribution actually is.

We stated earlier that the distribution of the statistic $t(M)$ of (13) has 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 $p(\boldsymbol{Y})$ of (14) would be $\Phi(t)$, where $t$ is given by (15), and its CDF would be

$$\Pr\bigl(p(\boldsymbol{Y}) \le u\bigr) = \Pr\bigl(\Phi(t) \le u\bigr) = \Pr\bigl(t \le \Phi^{-1}(u)\bigr).$$

Denote the CDF of the limiting distribution of $t$ by $G_{\alpha,\beta}$. The limiting distribution of $p(\boldsymbol{Y})$ 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 $p(\boldsymbol{Y})$ and the estimated CDF $G_{\hat{\alpha},\hat{\beta}} \circ \Phi^{-1}$.

The asymptotic distribution function $G_{\alpha,\beta}$ is characterised 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-normalised sums from samples of stable random variables with $\alpha$ and $\beta$ consistently estimated from the original sample. This amounts to a parametric bootstrap of $t$, without reference to $p(\boldsymbol{Y})$.

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 normalised, is invariant to scale.

The proposed bootstrap is described by the following steps:

1. Given the sample of random variables $Y_1, ..., Y_n$ with distribution $F$ in the domain of attraction of the stable law $cS(\alpha, \beta)$, compute the self-normalised 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-normalised 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 $t$.

### Theorem 2

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

### Proof:

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-normalised sum has the same asymptotic distribution for all laws in the domain of attraction of a given stable law $S(\alpha, \beta)$. ■

## 6. Goodness of Fit

This section is based on ongoing work joint with Emmanuel Flachaire, Frank Cowell, and Sanghamitra Bandyopadhyay. The idea is to develop a goodness-of-fit test based on a measure of distance between two distributions. Usually, one of the distributions is the empirical distribution of a sample; the other might be the empirical distribution of another sample, or else a theoretical distribution, in which case we suppose that it is absolutely continuous. It is desired to have a test that satisfies a number of criteria: among them robustness to heavy tails, and also the possibility to tailor the test statistic to maximise power in certain regions of the distribution.

For two samples of the same size, $\{X_i\}$ and $\{Y_i\}$, $i = 1, \ldots, n$, Cowell, in Cowell (1985), proposed the measure

$$J_\alpha = \frac{1}{n\alpha(\alpha-1)} \sum_{i=1}^n \left\{ \left(\frac{X_i}{\mu_1}\right)^\alpha \left(\frac{Y_i}{\mu_2}\right)^{1-\alpha} \right\},$$

where $\mu_1$ and $\mu_2$ are respectively the means of the $X$ and $Y$ samples, and $\alpha$, which may take on any real value, determines the part of the distribution to be weighted most heavily.

The measure can be adapted so that the $Y$ sample is replaced by a theoretical distribution with CDF $F$, as follows:

$$J_\alpha = \frac{1}{n\alpha(\alpha-1)} \sum_{i=1}^{n} \Big\{ \Big(\frac{X_i}{\mu_1}\Big)^\alpha \Big(\frac{F^{-1}(i/(n+1))}{\mu_F}\Big)^{1-\alpha} \Big\}, \tag{16}$$

where $F^{-1}$ is the quantile function for distribution $F$, and $\mu_F$ can be either the mathematical expectation of that distribution, or else the mean of the quantiles $F^{-1}(i/(n+1))$.

For the purposes of inference, it is necessary to know the distribution of $J_\alpha$ in (16) under the null hypothesis that the $X$ sample is an IID sample drawn from distribution $F$, or, failing that, the limiting distribution as $n \to \infty$. Under regularity conditions that are very restrictive regarding the right-hand tail of the distribution, it can be shown that the limiting distribution of $nJ_\alpha$ is that of

$$\frac{1}{2\mu_F} \Big[ \int_0^1 \frac{B^2(t)\,\mathrm{d}t}{F^{-1}(t) f^2(F^{-1}(t))} - \frac{1}{\mu_F} \Big( \int_0^1 \frac{B(t)\,\mathrm{d}t}{f(F^{-1}(t))} \Big)^2 \Big]. \tag{17}$$

Here $f = F'$ is the density of distribution $F$, $\mu_F$ is its expectation, and $B(t)$ is a Brownian bridge. Unfortunately, even for a distribution as well-behaved as the exponential, the random variable (17) has an infinite expectation. The divergence of the expectation as $n \to \infty$ is very slow, like $\log \log n$, but divergence at any rate whatever invalidates asymptotic inference, and makes bootstrap inference hard to justify. Only if $F$ has a bounded support does the limiting distribution have reasonable properties.

When $F$ is a known distribution, the $X_i$ of the original sample can be transformed to $F(X_i)$, which, under the null hypothesis, is distributed as U(0,1), that is, uniformly on the interval $[0, 1]$. The statistic that compares the $F(X_i)$ and the uniform distribution is, analogously to (17), but without the denominator of $n$,

$$G_\alpha \equiv \frac{1}{\alpha(\alpha-1)} \sum_{i=1}^{n} \Big\{ \Big(\frac{F(X_{(i)})}{\hat\mu_U}\Big)^\alpha \Big(\frac{i/(n+1)}{1/2}\Big)^{1-\alpha} - 1 \Big\}, \tag{18}$$

where $\hat\mu_U \equiv n^{-1} \sum_i F(X_i)$, and the $X_{(i)}$ are the order statistics. Since the $F(X_i)$ are IID drawings from U(0,1), the distribution of $G_\alpha$ in (18) is the same as that of the variable in which the $F_{(i)}$ are replaced by the order statistics of an IID sample of size $n$ from U(0,1). Thus the distribution of (18) under the null depends only on $n$ and $\alpha$, and is quite unaffected by the heaviness or otherwise of the tail of $F$. Further, the limiting distribution as $n \to \infty$ exists and has finite moments.

Matters are slightly more complicated if $F$ is known only up to the values of some parameters that can be consistently estimated. Suppose we have a family of distributions $F(\boldsymbol{\theta})$ and a vector of consistent estimates $\hat{\boldsymbol{\theta}}$. It can be shown that the distribution of $\hat G_\alpha$, in which $F$ is replaced by $F(\hat{\boldsymbol{\theta}})$, is well defined, and the limiting distribution exists. In fact, under certain conditions on the family $F(\boldsymbol{\theta})$, $\hat G_\alpha$ is an asymptotic pivot, a fact that justifies the use of the bootstrap.

In a simulation study with the lognormal distribution, variables $X$ were generated by the formula $X = \exp(\mu + \sigma W)$, $W \sim \mathrm{N}(0, 1)$. For each of $N$ samples of $n$ IID drawings, estimates $\hat{\mu}$ and $\hat{\sigma}$ of the parameters were obtained, and the estimated distribution, with CDF $\Phi\big((\log x - \hat{\mu})/\hat{\sigma}\big)$, used to construct a realisation of the $\hat{G}_\alpha$ just described. Next, $B$ bootstrap samples of size $n$ were generated by the formula $X^* = \exp(\hat{\mu} + \hat{\sigma} W^*)$, $W^* \sim \mathrm{N}(0, 1)$, and, for each bootstrap sample, estimates $\mu^*$ and $\sigma^*$ were obtained and used to compute a bootstrap statistic $G^*_\alpha$. The full set of bootstrap statistics was then used to form a bootstrap $P$ value, as the proportion of the $B$ statistics greater than the $\hat{G}_\alpha$ obtained from the original sample. The nominal distribution of the bootstrap $P$ value is U(0,1). Table 4 shows the maximum discrepancies of the empirical distributions of these $P$ values, based on $N$ replications, for $N = 10,000$, $B = 399$, $\mu = 0$, $\sigma = 1$, and $\alpha = 2$, as a function of sample size $n$. Except for the very small sample size with $n = 16$, the discrepancies are insignificantly different from zero.

| $n$ | 16 | 32 | 64 | 128 |
|---|---|---|---|---|
| max discrepancy | 0.0147 | 0.0048 | 0.0065 | 0.0049 |

Table 4: $P$ value discrepancies for $\hat{G}_\alpha$; lognormal distribution

## 7. A Wild Bootstrap for Quantile Regression

The content of this section jumps away from that of the rest of the paper, by considering quantile-based inference and, in particular, quantile regression. Basing inference on quantiles is often a good way to avoid the difficulties posed by the possible presence of heavy tails. In particular, quantile regression provides a way to obtain quite detailed information about the conditional distribution of a dependent variable. In Koenker and Xiao (2002), there is considerable discussion of the problem of performing reliable inference on the results of quantile regression. Although Koenker and Hallock (2000) mention heteroskedasticity as a “peril” for quantile regression, there does not seem to be a great deal in the literature about heteroskedasticity-robust inference for quantile regression, beyond suggestions that the Eicker-White approach can be extended to it. See also the excellent textbook treatment in Koenker (2005).

With least-squares regression, and indeed in many other contexts, a bootstrap technique that offers a degree of robustness against heteroskedasticity is the wild bootstrap. In Davidson and Flachaire (2008), it is suggested that, for ordinary least squares, the most reliable way to implement the wild bootstrap is to use the Rademacher distribution:

$$\varepsilon_t = \begin{cases} 1 & \text{with probability } 1/2 \\ -1 & \text{with probability } 1/2. \end{cases} \tag{19}$$

in order to form the bootstrap disturbances as $u^*_t = \hat{u}_t \varepsilon_t$, where the $\hat{u}_t$ are the OLS residuals. This means that the bootstrap disturbances are just the residuals multiplied by a random sign.

If one forgets for a moment the difference between residuals and true disturbances, the wild bootstrap with (19) conditions on the absolute values of the disturbances, and generates the bootstrap distribution by varying their signs. This procedure can be expected to work well if the disturbances are symmetrically distributed about zero, because in that case the sign of the disturbance is independent of its absolute magnitude. A recent working paper by Cavaliere, Georgiev, and Taylor (2009) uses this fact to justify a wild bootstrap for the mean when the variance need not exist. In fact their procedure works for the median even when the mean itself does not exist. When the disturbances of a regression model are skewed, however, the symmetric wild bootstrap can be expected to work less well, which is why the original suggestion of Mammen (1993) was to use an asymmetric distribution instead of (19).

With a skewed distribution with median zero, it is still possible to decompose a drawing into a sign and another variable independent of the sign. We have

**Lemma 2**

Let $F$ be an absolutely continuous CDF. The random variable $X$ given by

$$\begin{aligned} &X = SF^{-1}(U) + (1-S)F^{-1}(1-U), \\ &U \sim \mathrm{U}(0, 0.5), \quad S = \begin{cases} 1 & \text{with probability } 0.5 \\ 0 & \text{with probability } 0.5 \end{cases}, \quad U \perp\!\!\!\perp S, \end{aligned} \tag{20}$$

follows the distribution with CDF $F$. Conversely, if $X$ has CDF $F$, and if $U$ and $S$ are defined by

$$S = \mathrm{I}\big(X \le F^{-1}(0.5)\big) \quad \text{and} \quad U = SF(X) + (1-S)(1-F(X)), \tag{21}$$

then $X$, $U$, and $S$ satisfy (20). In addition, the unconditional median of $X$ is also a median conditional on $U$.

**Proof:** In Appendix

A wild bootstrap procedure for median regression can be based on this lemma. Let the regression model be

$$y_t = \boldsymbol{X}_t\boldsymbol{\beta} + u_t, \qquad t = 1, \ldots, n, \tag{22}$$

where the disturbance terms $u_t$ have zero median, so that the median of $y_t$ conditional on the exogenous explanatory variables $\boldsymbol{X}_t$ is $\boldsymbol{X}_t\boldsymbol{\beta}$. The quantile regression estimator $\hat{\boldsymbol{\beta}}$ for the median is just the least-absolute-deviation (LAD) estimator, which minimises the sum

$$\sum_{t=1}^{n} |y_t - \boldsymbol{X}_t\boldsymbol{\beta}|.$$

The wild bootstrap DGP first centres the LAD residuals by subtracting their median from each of them. Then the residuals along with the corresponding $\boldsymbol{X}_t$ variables are sorted in increasing order of the residuals, thereby keeping the pairing between explanatory variables and residuals. Denote by $(\hat{u}_{(t)}, \boldsymbol{X}_{(t)})$ the pair in position $t$ of the sorted array. The next step is to generate a sequence $S_t^*$ of IID drawings from the binary distribution (19), and then, for each $t$ such that $S_t^* = 1$, set $u_t^* = \hat{u}_{(t)}$, and, for each $t$ such that $S_t^* = -1$, set $u_t^* = \hat{u}_{(n-t)}$ for $n$ even, or $u_{(n+1-t)}$ for $n$ odd. The bootstrap sample is generated by the equation

$$y_t^* = \boldsymbol{X}_{(t)}\hat{\boldsymbol{\beta}} + u_t^*.$$

In this way, the bootstrap DGP implements (20) with $F$ given by the empirical distribution of the recentred LAD residuals. Zero is a median of each $u_t^*$ conditional on $\boldsymbol{X}_{(t)}$, and so the bootstrap DGP is a special case of the model (22). Note that the reordering of the observations due to the sorting by the residuals is of no importance.

It must be noted that this new wild bootstrap is *not* appropriate for least-squares regression, since, the mean of $u_t^*$ conditional on $\boldsymbol{X}_{(t)}$ is not zero, unless the underlying distribution is symmetric.

All the bootstrap DGPs we look at next have the form

$$y_t^* = \boldsymbol{X}_t^*\hat{\beta} + u_t^*.$$

A conventional resampling bootstrap sets $\boldsymbol{X}_t^* = \boldsymbol{X}_t$ and the $u_t^*$ as IID drawings from the empirical distribution of the residuals $y_t - \boldsymbol{X}_t\hat{\boldsymbol{\beta}}$. This bootstrap DGP thus destroys the pairing between the $\boldsymbol{X}_t$ and the corresponding disturbances. The conventional pairs bootstrap, or $(y, X)$ bootstrap, which resamples pairs $(y_t, \boldsymbol{X}_t)$, preserves the pairing, at the expense of abandoning the condition that the median of the bootstrap disturbance for observation $t$ is zero conditional on $\boldsymbol{X}_t$. The conventional wild bootstrap assigns a random sign to each residual, retaining the pairing, so that $\boldsymbol{X}_t^* = \boldsymbol{X}_t$ and $u_t^* = S_t^*|u_t|$, with the $S_t^*$ IID drawings from (19). This bootstrap DGP does maintain the condition that $u_t^*$ has zero median conditional on $\boldsymbol{X}_t$. The new wild bootstrap is as described above.

For simplicity, we now restrict attention to the case in which there is only one explanatory variable, $X_t$, and one parameter $\beta$, with LAD estimate $\hat{\beta}$, and true value $\beta_0$. For all of the different bootstrap procedures, the LAD estimate $\beta^*$ is computed using the bootstrap data, and the bootstrap statistic $\beta^* - \hat{\beta}$ computed. The bootstrap uses the distribution of $\beta^* - \hat{\beta}$ as an estimate of that of $\hat{\beta} - \beta_0$ under the null, and so the bootstrap $P$ value is the proportion of the $\beta^* - \hat{\beta}$ that are more extreme than $\hat{\beta} - \beta_0$. Both two-tailed and one-tailed tests are possible. The nominal distribution of the bootstrap $P$ value under the null is U(0,1).

First, we look at the case with the $u_t$ IID drawings from some distribution with zero median. In this case, even with skewed disturbances, all the bootstraps work as well as can reasonably be expected when working with a non-pivotal quantity that is not even asymptotically pivotal. In Table 5 are shown the maximum discrepancies between the bootstrap distribution, based on 1,000 replications with 199 bootstraps each, and the nominal U(0,1) distribution, for a one-tailed test that rejects to the left. The symmetric

| $n$ | 33 | 129 | 513 |
|---|---|---|---|
| resampling | 0.065 | 0.040 | 0.037 |
| | 0.078 | 0.052 | 0.046 |
| wild | 0.098 | 0.047 | 0.039 |
| | 0.102 | 0.044 | 0.043 |
| new wild | 0.112 | 0.055 | 0.038 |
| | 0.125 | 0.056 | 0.048 |

Table 5: Maximum bootstrap discrepancy; IID disturbances.
Upper numbers for symmetric distribution, lower for skewed.

distribution is the stable distribution $S(1.5, 0)$, and the skewed distribution $S(1.5, 0.5)$. Thus in neither case does the variance exist.

We see that the new wild bootstrap is in fact the worst of the three in most cases, although, since the figures in the table may have a standard deviation of up to 0.016, the differences in the discrepancies are not measured very accurately in this experiment.

Things are very different, however, if the disturbances are heteroskedastic. In this case, we do not expect the resampling bootstrap to be appropriate. The conventional wild bootstrap should be fine so long as the disturbances are symmetrically distributed, less so if they are skewed. The new wild bootstrap should work well enough even with skewed, heteroskedastic disturbances. The DGP used to study the effects of heteroskedasticity is as follows:

$$y_t = X_t(\beta_0 + u_t),$$

where the $u_t$ again have median zero, and are independent of the corresponding $X_t$. The median of $y_t$ conditional on $X_t$ is again $X_t\beta_0$.

Maximum discrepancies are shown in Table 6. We see that resampling is invalid; its discrepancy does not diminish for larger samples. Both wild bootstraps work acceptably well, the conventional one somewhat better than the new one.

| $n$ | 33 | 129 | 513 |
|---|---|---|---|
| resampling | 0.240 | 0.248 | 0.281 |
| | 0.254 | 0.253 | 0.294 |
| wild | 0.093 | 0.039 | 0.034 |
| | 0.090 | 0.041 | 0.040 |
| new wild | 0.112 | 0.046 | 0.043 |
| | 0.128 | 0.044 | 0.057 |

Table 6: Maximum discrepancy; heteroskedastic disturbances
Upper numbers for symmetric distribution, lower for skewed.

All the results so far presented are for median regression. For a general quantile $p$, $0 < p < 1$, and model (22), one minimises the function

$$\sum_{t=1}^{n} \rho_p(y_t - \boldsymbol{X}_t\boldsymbol{\beta})$$

with respect to $\boldsymbol{\beta}$, where $\rho_p(u) = u\big(p - \mathrm{I}(u < 0)\big)$; see Koenker (2005). Let $\hat{\boldsymbol{\beta}}_p$ be the estimator resulting from this minimisation. If the disturbances are IID, then a possible resampling bootstrap DGP is

$$y_t^* = \boldsymbol{X}_t\hat{\boldsymbol{\beta}}_p + u_t^*,$$

where the $u_t^*$ are IID drawings from the empirical distribution of the residuals $y_t - \boldsymbol{X}_t\hat{\boldsymbol{\beta}}_p$, shifted so that their $p$-quantile is zero. Thus the $p$-quantile of $y_t^*$ is $\boldsymbol{X}_t\hat{\boldsymbol{\beta}}_p$. If the disturbances $u_t$ are IID, then this resampling bootstrap should be valid asymptotically. However, if the $u_t$ are heteroskedastic, it should be no more valid than it is for the case we examined, with $p = 0.5$.

The conventional wild bootstrap may be modified as follows for the model with just one explanatory variable. In order to test the null hypothesis that the $p$-quantile of $y_t$ is $X_t\beta_0$, the estimate $\hat{\beta}_p$ is found, along with the residuals $\hat{u}_t = y_t - X_t\hat{\beta}_p$, and the estimation error $\hat{\beta}_p - \beta_0$ computed. The residuals are shifted so that their $p$-quantile is zero. The bootstrap DGP is then

$$y_t^* = X_t\hat{\beta}_p - S_t^*|\hat{u}_t| + (1 - S_t^*)|\hat{u}_t|,$$

where $S_t^* = 1$ with probability $p$, and -1 with probability $1 - p$. Thus the $p$-quantile of $y_t^*$ conditional on $X_t$ is $X_t\hat{\beta}_p$. If $\beta_p^*$ denotes the estimate from the $p$-quantile regression with the bootstrap data, then the distribution of $\beta_p^* - \hat{\beta}_p$ is the bootstrap estimate of the distribution of $\hat{\beta}_p - \beta_0$.

The new wild bootstrap can be modified using the following Lemma.

### Lemma 3

Let $F$ be an absolutely continuous CDF, and let $0 < p < 1$. The random variable $X$ given by

$$\begin{aligned} &X = SF^{-1}(pU) + (1 - S)F^{-1}\big(1 - (1 - p)U\big), \\ &U \sim \mathrm{U}(0,1), \qquad S = \begin{cases} 1 & \text{with probability } p \\ 0 & \text{with probability } 1 - p \end{cases}, \qquad U \perp\!\!\!\perp S \end{aligned} \tag{23}$$

follows the distribution with CDF $F$. Conversely, if $X$ has CDF $F$, and if $U$ and $S$ are defined by

$$S = \mathrm{I}\big(X \le F^{-1}(p)\big) \qquad \text{and} \qquad U = S\frac{F(X)}{p} + (1 - S)\frac{1 - F(X)}{1 - p},$$

then $X$, $U$, and $S$ satisfy (23). In addition, the unconditional $p$-quantile of $X$ is also a $p$-quantile conditional on $U$.

**Proof:** Similar to that of Lemma 2.

Since only very preliminary results are as yet available, I will not go into details of the implementation of the new wild bootstrap for general $p$. These preliminary results show that, with heteroskedastic disturbances, the resampling bootstrap fails completely, the modified conventional wild bootstrap performs badly, whereas the new wild bootstrap gives rise to distortions not significantly different from those presented above for median regression.

## 8. Conclusion

The result of Bahadur and Savage imposes severe restrictions on the sort of model that can allow for reliable inference based on moments. The reasoning that leads to their result is unrelated to the existence or otherwise of heavy tails, but imposing the boundedness of higher moments does avoid the impossibility result of their theorem. The bootstrap can sometimes provide reasonable moment-based inference with heavy-tailed distributions, but its performance degrades as higher moments cease to exist. Quantile-based methods offer an escape. In particular, if one is prepared to assume the symmetry of a distribution, unreliable inference on the mean can be replaced by reliable inference on the median. The use of the bootstrap with quantile-based methods is worthy of considerable further research, given the encouraging results obtained with the new wild bootstrap procedure.

## References


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.

Caers, J. and J. Van Dyck (1999). “Nonparametric tail estimation using a double bootstrap method”, *Computational Statistics & Data Analysis* **29**, 191-211.

Cavaliere, G., I. Georgiev, and A. M. Robert Taylor (2009). “Wild Bootstrap of the Mean in the Infinite Variance Case”, Working paper.

Coles, S. (2001). *An Introduction to Statistical Modeling of Extreme Values*. London: Springer.

Cornea, A. and R. Davidson (2008). “A Refined Bootstrap for Heavy Tailed Distributions”, Working paper, Department of Economics, McGill University.

Cowell, F. A. (1985). “Measures of Distributional Change: an Axiomatic Approach”, *Review of Economic Studies*, **52**, 135–151.

Cowell, F. A., E. Flachaire, and S. Bandyopadhyay, S. (2009). “Goodness-of-Fit: An Economic Approach”, Distributional Analysis Discussion Paper, 101, STICERD, LSE.

Davidson, R. (2009). “Reliable Inference for the Gini Index”, *Journal of Econometrics*, **150**, 30–40.

Davidson, R. and E. Flachaire (2007). “Asymptotic and Bootstrap Inference for Inequality and Poverty Measures”, *Journal of Econometrics*, **141**, 141–166.

Davidson, R. and E. Flachaire (2008). “The Wild Bootstrap, Tamed at Last”, *Journal of Econometrics*, **146**, 162–169.

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

Gilleland, E. and R. W. Katz (2005). *Extremes Toolkit: Weather and Climate Applications of Extreme Value Statistics*. R software and accompanying tutorial.

Hall, P. (1990. “Asymptotic properties of the bootstrap for heavy-tailed distributions”, *Annals of Probability* **18**, 1342–1360.

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.

Koenker, R. (2005). *Quantile Regression*, Econometric Society Monographs No 38, Cambridge University Press.

Koenker, R. and K. Hallock (2000). “Quantile Regression: An Introduction”, available at `http://www.econ.uiuc.edu/~roger/research/intro/intro.html`.

Koenker, R. and Z. Xiao (2002). “Inference on the Quantile Regression Process”, *Econometrica*, **70**, 1583-1612.

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-normalised sums”, *Annals of Probability* **11**, 788–809.

Mammen, E. (1993). “Bootstrap and wild bootstrap for high dimensional linear models”, *Annals of Statistics* **21**, 255–285.

Schluter, C. and M. Trede (2002). “Tails of Lorenz curves”, *Journal of Econometrics* **109**, 151–166.

Singh, S. K. and G. S. Maddala (1976). “A Function for Size Distribution of Incomes”, *Econometrica* **44**, 963–970.

## Appendix

### Lemma 1:

> If the class of distributions $\mathcal{F}$ is restricted so that there exists a finite $K$ and some $\theta > 0$ such that $\mathrm{E}|U|^{1+\theta} < K$ for all distributions in $\mathcal{F}$, the mapping from $\mathcal{F}$ to the real line that associates its expectation to each distribution is continuous.

### Proof:

Suppose that, contrary to what we wish to show, this restriction does not make the mapping continuous. Then, for the density $f$ of a distribution in $\mathcal{F}$ at which the mapping is not continuous, there exists a sequence of densities $g_m$ of distributions in $\mathcal{F}$ such that

$$\int |f(x) - g_m(x)|\,\mathrm{d}x \to 0 \text{ as } m \to \infty \tag{24}$$

with

$$\int x\, f(x)\,\mathrm{d}x = \mu_0, \quad \int x\, g_m(x)\,\mathrm{d}x \to \mu_1 \neq \mu_0 \text{ as } m \to \infty.$$

The uniform boundedness assumption means that there exist $K$ and $\delta > 0$ such that

$$\int |x|^{1+\theta} f(x)\,\mathrm{d}x < K \text{ and } \int |x|^{1+\theta} g_m(x)\,\mathrm{d}x < K \text{ for all } m.$$

Then

$$\int |x|^{1+\theta}|f(x) - g_m(x)|\,\mathrm{d}x \leq \int |x|^{1+\theta}\big(f(x) + g_m(x)\big)\,\mathrm{d}x < 2K. \tag{25}$$

For any $K_1 > 1$, we have for sufficiently large $m$ that

$$|\mu_0 - \mu_1| \leq \int_{|x|\leq K_1} |x|\,|f(x) - g_m(x)|\,\mathrm{d}x + \int_{|x|>K_1} |x||f(x) - g_m(x)|\,\mathrm{d}x. \tag{26}$$

But

$$\begin{aligned} \int |x|^{1+\theta}|f(x) - g_m(x)|\,\mathrm{d}x &\geq \int_{|x|>K_1} |x|^{1+\theta}|f(x) - g_m(x)|\,\mathrm{d}x \\ &\geq K_1^{\theta} \int_{|x|>K_1} |x|\,|f(x) - g_m(x)|\,\mathrm{d}x. \end{aligned} \tag{27}$$

Choose $M(K_1)$ such that

$$\int_{|x|\leq K_1} |x|\,|f(x) - g_m(x)|\,\mathrm{d}x \leq K_1 \int |f(x) - g_m(x)|\,\mathrm{d}x < |\mu_0 - \mu_1|/2$$

for all $m > M(K_1)$. This is possible by virtue of assumption (24). Then from (26), we see that

$$\int_{|x|>K_1} |x|\,|f(x) - g_m(x)|\,\mathrm{d}x > \frac{|\mu_0 - \mu_1|}{2},$$

and, combining this with (27), we obtain for all $m > M(K_1)$

$$\int |x|^{1+\theta}|f(x) - g_m(x)|\,\mathrm{d}x > \frac{|\mu_0 - \mu_1|}{2}K_1^\theta.$$

For $\theta > 0$, we can find a $K_1$ such that $K_1^\theta|\mu_0 - \mu_1|/2$ is greater than $2K$, which contradicts (25). Thus the uniform boundedness assumption restores the continuity of the mapping from $\mathcal{F}$ to the expectation.

It is not hard to check that the above proof does not require that the densities exist. We can systematically replace $f(x)\,\mathrm{d}x$ by $dF(x)$, where $F$ is a CDF, and similarly for $g_n(x)$. ■

## Proof of Lemma 2

Let $m = F^{-1}(0.5)$ be the median of the distribution $F$. We compute the CDF of $X$ given by (20). For $x \le m$,

$$\Pr(X \le x) = \Pr(S = 1)\Pr(F^{-1}(U) \le x) = 0.5\Pr(U \le F(x)) = F(x).$$

For $x > m$,

$$\begin{aligned}\Pr(X \le x) &= \Pr(S = 1) + \Pr(S = 0)\Pr(F^{-1}(1 - U) \le x) \\ &= \tfrac{1}{2}\big(1 + \Pr(1 - U \le F(x)\big) = \tfrac{1}{2}\big(1 + 2F(x) - 1\big) = F(x).\end{aligned}$$

This demonstrates the first part of the Lemma.

Let $U$ and $S$ be defined as in (21). Then obviously $S$ has the required distribution, and $U \in [0, 0.5]$. Further,

$$SF^{-1}(U) + (1 - S)F^{-1}(1 - U) = SX + (1 - S)F^{-1}\big(F(X)\big) = X.$$

We next show the independence of $U$ and $S$. We see that, for $u \in [0, 0.5]$,

$$\Pr(U \le u|S = 1) = \Pr\big(F(X) \le u|X \le F^{-1}(0.5)\big) = \frac{\Pr\big(X \le F^{-1}(u)\big)}{\Pr\big(X \le F^{-1}(0.5)\big)} = 2u,$$

while

$$\Pr(U \le u|S = 0) = \Pr\big(1 - F(X) \le u|X > F^{-1}(0.5)\big) = \frac{\Pr\big(X \ge F^{-1}(1 - u)\big)}{\Pr\big(X > F^{-1}(0.5)\big)} = 2u.$$

It follows that $S$ and $U$ are independent, and that $U \sim \mathrm{U}(0, 0.5)$.

Finally,

$$
\begin{aligned}
\Pr\big(X \leq m \mid U\big) &= \mathrm{E}\Big(\mathrm{I}\big(SF^{-1}(U) + (1-S)F^{-1}(1-U) \leq m\big) \mid U\Big) \\
&= \mathrm{E}\Big(S\,\mathrm{I}\big(F^{-1}(U) \leq m\big) \mid U\Big) + \mathrm{E}\Big((1-S)\,\mathrm{I}\big(F^{-1}(1-U) \leq m\big) \mid U\Big) \\
&= 0.5\,\mathrm{I}\big(U \leq F(m)\big) + 0.5\,\mathrm{I}\big(1-U \leq m\big) = 0.5,
\end{aligned}
$$

since $0 \leq U \leq 0.5$ and $F(m) = 0.5$, so that $U \leq F(m)$ with probability 1 and $1 - U \leq m$ with probability 0. This proves the last statement of the Lemma.

■