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

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. 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.

March 2010
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.
Outline of Talk

The Result of Bahadur and Savage

Illustration with the Gini Index

Measures of Inequality

A Parametric Bootstrap for the Domains of Stable Laws

Goodness of Fit

A Wild Bootstrap for Quantile Regression
The Result of Bahadur and Savage

The model for which the impossibility results of Bahadur and Savage 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 $F \in \mathcal{F}$, $\mu_F \equiv \int_{-\infty}^{\infty} x \, dF(x)$ 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 $\inf_{F \in \mathcal{F}_m} \mathbb{E}\phi$ and $\sup_{F \in \mathcal{F}_m} \mathbb{E}\phi$ 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.
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.
Each distribution in the family is characterised by a parameter $p \in [0, 1]$. A random
variable from the distribution can be written as

$$U = Y/p^2 + (1 - Y)W - 1/p$$

where $W \sim N(0, 1)$ 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 $E_p U = 0$.

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

$$\Pr(N = 0) = (1 - p)^n.$$ 

The $t$ statistic for a test of the hypothesis that $E U = 0$ 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 $N(-n^{1/2}/p, 1)$. Then, since $U_t - \hat{\mu} = W_t - \bar{W}$ if $N = 0$, the conditional distribution of $n\hat{\sigma}^2_\mu$ 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 $c > 0$:

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

$$
= \Pr(N = 0) \Pr(T < -c | N = 0).
$$

Now

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

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

Statistical Inference in the Presence of Heavy Tails

5
For fixed $c$ and $n$, let $p \to 0$. Then $\Pr(N = 0) \to 1$. Also $\Pr(T < -c \mid N = 0)$ also tends to 1, since the noncentrality parameter tends to $-\infty$. Thus the rejection probability tends to 1 whatever the critical value $c$, and so the test has size 1. 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 has all its moments finite for positive $p$, imposing conditions on the existence of moments does not prevent all the distributions of our one-parameter family from being present in the null model, with the unfortunate consequences predicted by the theorem of BS. In our example, see that for $p > 0$, the first moment is zero. In the limit, however, it is infinite.
We do not want to impose a bound on the first moment, since that is completely
counter to the spirit of a test for the expectation. Nevertheless, whatever restriction
we may impose in order to restore the possibility of valid inference, we know that
the restriction must exclude the distributions with small $p$. We are therefore led to
consider a *uniform* bound on some higher moment. 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 $\theta > 0$,
$E|U|^{1+\theta} < K$, for some specified $K$. Lemma 1 in the paper 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.
Illustration with the Gini Index

Most of this section is borrowed from one of my recent papers 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 $F(x) = 1 - e^{-x}$, $x \geq 0$. 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 $\hat{\sigma}_G$ I derived, for sample sizes $n = 10$ and 100. The graph of the standard normal CDF is also given as a benchmark.
Figure 1. Distribution of the standardised statistic as a function of sample size

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.
In Figure 2 are 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 2. Distribution of the standardised statistic for the Pareto distribution](image-url)
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.

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

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 $\sigma = 0, 0.5, 1.0, 1.5$.

Figure 3. Distribution of $\tau$ for the lognormal distribution
Table 2 gives 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.

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

Table 2. Coverage of percentile-t confidence intervals
Measures of Inequality

This section is 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:

$$\hat{\alpha} = H_{k,n}^{-1}, \quad H_{k,n} = k^{-1} \sum_{i=0}^{k-1} \log Y(n-i) - \log Y(n-k+1),$$

where $Y(j)$ is the $j^{\text{th}}$ order statistic of the sample. The estimator 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.
The choice of $k$ is a question of trade-off between bias and variance. 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. Experiments with this graphical method for samples of different sizes ranging from 100 to 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

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

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.

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

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

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 above Pareto distribution 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

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

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

$$W = (\hat{T} - T_0)/[\hat{V}(\hat{T})]^{1/2},$$

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 Hill’s estimator $\hat{\alpha}$. Compute the true value of the index, $T_0^*$, for the bootstrap distribution as $T(F_{bs})$.

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 tail distribution if $X_i = 1$, and a drawing from the empirical distribution of the $\bar{n}$ smallest order statistics $Y(j), j = 1, \ldots, \bar{n}$, if $X_i = 0$.

5. With the bootstrap sample, compute the index $\hat{T}^*$, its variance estimate $\hat{V}(\hat{T}^*)$, and the bootstrap statistic $W^* = (\hat{T}^* - T_0^*)/[\hat{V}(\hat{T}^*)]^{1/2}$.

6. Repeat steps 4 and 5 $B$ times, obtaining the bootstrap statistics $W_j^*, j = 1, \ldots, B$. The bootstrap $P$-value is computed as the proportion of $W_j^*, j = 1, \ldots, B$, that are smaller than $W$. 

Statistical Inference in the Presence of Heavy Tails 17
The index used for the above figure was Theil's index

\[ T(F) = \int \frac{y}{\mu_F} \log \frac{y}{\mu_F} \, dF(y). \]
Heavier tails

Although the bootstrap distribution of the statistic $W$ 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 $n \to \infty$.

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 $\alpha = 1.9$.

Table 3

<table>
<thead>
<tr>
<th></th>
<th>asymptotic</th>
<th>std bootstrap</th>
<th>$m$ out of $n$</th>
<th>semi-parametric</th>
</tr>
</thead>
<tbody>
<tr>
<td>$b = 2.1, c = 1$</td>
<td>0.41</td>
<td>0.24</td>
<td>0.15</td>
<td>0.16</td>
</tr>
<tr>
<td></td>
<td>-0.03</td>
<td>-0.04</td>
<td>-0.03</td>
<td>0.04</td>
</tr>
<tr>
<td>$b = 1.9, c = 1$</td>
<td>0.48</td>
<td>0.28</td>
<td>0.20</td>
<td>0.18</td>
</tr>
<tr>
<td></td>
<td>-0.03</td>
<td>-0.04</td>
<td>-0.02</td>
<td>0.06</td>
</tr>
</tbody>
</table>

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

Although the variance estimate in the denominator of the statistic is meaningless if the variance does not exist, we see from the Table that the ERPs seem to be continuous across the boundary at $\alpha = 2$. 

*Statistical Inference in the Presence of Heavy Tails*
A Parametric Bootstrap for the Domains of Stable Laws

This section is borrowed from a working paper, 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

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

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 $\delta$ can be consistently estimated by the sample mean.

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

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

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

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

where $\mathbf{Y}$ denotes the vector of the $Y_j$, and $E^*$ 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 $\mathbf{Y}$. As $n \to \infty$ this distribution tends to a nonrandom limit.

Consider the self-normalised sum

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

It has expectation 0 and variance 1 conditional on $\mathbf{Y}$, and so also unconditionally. Let $F^n_{\mathbf{Y}}$ denote the random CDF of $t(M)$. Then we have

$$p(\mathbf{Y}) = F^n_{\mathbf{Y}}\left(\frac{\sum_{j=1}^n Y_j}{(\sum_{j=1}^n (Y_j - \bar{Y})^2)^{1/2}}\right).$$
The principal question that asymptotic theory is called on to answer in the context of bootstrapping the mean is:

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

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

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

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 \( F^n_Y \) 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.
The asymptotic distribution turns out to be characterised by an integral in the complex plane 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 $p(Y)$.

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.
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 $S(\alpha, \beta)$.
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 is robust to heavy tails, and can be tailored so as 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 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:

\[
J_\alpha = \frac{1}{n\alpha(\alpha - 1)} \sum_{i=1}^{n} \left\{ \left( \frac{X_i}{\mu_1} \right)^\alpha \left( \frac{F^{-1}(i/(n+1))}{\mu_F} \right)^{1-\alpha} \right\},
\]

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 \( F^{-1}(i/(n+1)) \).

Statistical Inference in the Presence of Heavy Tails
For the purposes of inference, it is necessary to know the distribution of \( J_\alpha \) 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

\[
\frac{1}{2\mu_F} \left[ \int_0^1 \frac{B^2(t) \, dt}{F^{-1}(t)f^2(F^{-1}(t))} - \frac{1}{\mu_F} \left( \int_0^1 \frac{B(t) \, dt}{f(F^{-1}(t))} \right)^2 \right].
\]

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 above random variable 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, without the denominator of $n$,

$$J_\alpha = \frac{1}{\alpha(\alpha - 1)} \sum_{i=1}^{n} \left\{ \left( \frac{F(X(i))}{\hat{\mu}_U} \right)^\alpha \left( \frac{i/(n + 1)}{1/2} \right)^{1-\alpha} - 1 \right\},$$

where $\hat{\mu}_U = n^{-1} \sum_i F(X_i)$, and the $X_{(i)}$ are the order statistics. Since the $F(X_i)$ are IID drawings from $U(0,1)$, the distribution of the new $J_\alpha$ 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 this new $J_\alpha$ 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 $F(\theta)$ and a vector of consistent estimates $\hat{\theta}$. It can be shown that the distribution of $\hat{J}_\alpha$, in which $F$ is replaced by $F(\hat{\theta})$, is well defined, and the limiting distribution exists. In fact, under certain conditions on the family $F(\theta)$, $\hat{J}_\alpha$ 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 $X = \exp(\mu + \sigma W)$, $W \sim N(0,1)$. 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 $\Phi\left(\frac{\log x - \hat{\mu}}{\hat{\sigma}}\right)$, used to construct a realisation of the $\hat{J}_\alpha$ just described. Next, $B$ bootstrap samples of size $n$ were generated by the formula $X^* = \exp(\hat{\mu} + \hat{\sigma} W^*)$, and, for each bootstrap sample, estimates $\mu^*$ and $\sigma^*$ were obtained and used to compute a bootstrap statistic $J^*_\alpha$. 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 $\hat{J}_\alpha$ 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 the $N$ 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.

<table>
<thead>
<tr>
<th>$n$</th>
<th>16</th>
<th>32</th>
<th>64</th>
<th>128</th>
</tr>
</thead>
<tbody>
<tr>
<td>max discrepancy</td>
<td>0.0147</td>
<td>0.0048</td>
<td>0.0065</td>
<td>0.0049</td>
</tr>
</tbody>
</table>

Table 4: $P$ value discrepancies for $\hat{J}_\alpha$; lognormal distribution
A Wild Bootstrap for 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. 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:

\[ \varepsilon_t = \begin{cases} 
1 & \text{with probability } 1/2 \\
-1 & \text{with probability } 1/2.
\end{cases} \]

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 above wild bootstrap 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, but less well if they are skewed, which is why the original suggestion of Mammen (1993) was to use an asymmetric distribution instead of the Rademacher distribution.
When a distribution is symmetric about the origin, the absolute value of a drawing from the distribution is independent of its sign. With a skewed distribution with median zero, it is still possible to decompose a drawing into a sign and another variable.

**Lemma 2**

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

$$X = SF^{-1}(U) + (1 - S)F^{-1}(1 - U),$$

$$U \sim U(0, 0.5), \quad S = \begin{cases} 
1 & \text{with probability 0.5} \\
0 & \text{with probability 0.5}
\end{cases}, \quad U \parallel S,$$

(1)

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

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

then $X$, $U$, and $S$ satisfy (1).

A “new wild bootstrap” can be based on the result of Lemma 2. The result is applied to the empirical distribution of the sample, with the sample median subtracted. The new wild bootstrap DGP begins by sorting the quantities to be resampled, thereby obtaining the order statistics $X_{(i)}$, $i = 1, \ldots, n$. Next, it generates a sequence $S^*_t$ of IID drawings from the binary distribution, and then, for each $t$ such that $S^*_t = 1$, sets $X^*_t = X_t$, and, for each $t$ such that $S^*_t = 0$, finds the index $i$ such that $X_{(i)} = X_t$, and then sets $X^*_t = X_{(n-i)}$ for $n$ even, or $X_{(n+1-i)}$ for $n$ odd.
Suppose first that an IID sample is available, drawn from an unknown distribution, and that we wish to perform inference on the median of the distribution using some bootstrap procedure. If all we wish is a percentile bootstrap confidence interval, then an ordinary resampling bootstrap would seem to be perfectly adequate. If we feel happy with some asymptotic standard error, a percentile-$t$ interval is also available by resampling. In all cases, what is resampled should be the original observations minus the sample median, and the bootstrap distribution of the median used as an estimate of the estimation error. How do the wild and new bootstraps compare? We cannot expect a great difference, since in both cases, the bootstrap median depends on how many bootstrap observations are above and how many below the median of the original sample.

It must be noted that the new wild bootstrap is not appropriate for least-squares regression, since, although the median of the bootstrap distribution is zero, the mean is not, unless the underlying distribution is symmetric.
Table 5. Resampling and wild bootstrap of the median; symmetric distribution
The underlying distribution for Table 5 is the stable distribution with tail parameter \( \alpha = 1.5 \), so that the variance does not exist, and skewness parameter of zero, so that the true median is zero.

<table>
<thead>
<tr>
<th>( n )</th>
<th>17</th>
<th>33</th>
<th>65</th>
<th>129</th>
<th>257</th>
<th>513</th>
</tr>
</thead>
<tbody>
<tr>
<td>resampling</td>
<td>0.128</td>
<td>0.099</td>
<td>0.072</td>
<td>0.057</td>
<td>0.037</td>
<td>0.030</td>
</tr>
<tr>
<td>wild bootstrap</td>
<td>0.084</td>
<td>0.063</td>
<td>0.050</td>
<td>0.031</td>
<td>0.023</td>
<td>0.016</td>
</tr>
<tr>
<td>new wild bootstrap</td>
<td>0.154</td>
<td>0.120</td>
<td>0.081</td>
<td>0.070</td>
<td>0.049</td>
<td>0.039</td>
</tr>
</tbody>
</table>

Table 6. Resampling and wild bootstrap of the median; skewed distribution
Table 6 is based on \( S(1.5, 0.5) \). It shows that in fact the old wild bootstrap still outperforms anything else.

<table>
<thead>
<tr>
<th>( n )</th>
<th>17</th>
<th>33</th>
<th>65</th>
<th>129</th>
<th>257</th>
<th>513</th>
</tr>
</thead>
<tbody>
<tr>
<td>resampling</td>
<td>0.141</td>
<td>0.107</td>
<td>0.083</td>
<td>0.049</td>
<td>0.042</td>
<td>0.026</td>
</tr>
<tr>
<td>wild bootstrap</td>
<td>0.091</td>
<td>0.071</td>
<td>0.050</td>
<td>0.033</td>
<td>0.024</td>
<td>0.016</td>
</tr>
<tr>
<td>new wild bootstrap</td>
<td>0.148</td>
<td>0.114</td>
<td>0.078</td>
<td>0.074</td>
<td>0.051</td>
<td>0.040</td>
</tr>
</tbody>
</table>

Statistical Inference in the Presence of Heavy Tails