Abstract

This paper analyses the asymptotic distribution of the classical t-ratio from distributions with no finite moments and shows how classical testing is affected. Some surprising results are obtained in terms of bimodality vs. the usual unimodality of the standard studentized t-distribution prevailing in classical conditions. The paper develops a new distribution termed the “double Pareto,” which allows the thickness of the tails and the existence of moments to be determined parametrically. We also consider a Cauchy distribution truncated on a compact support to investigate the relative importance of tail thickness in case of finite moments. We find that the bimodality persists even in such cases. Simulation results are used to highlight the dangers of relying on naive testing in the face of thick-tailed distributions. Special cases analyzed include one- and two-sample statistical inference problems, as well as linear regression econometric problems.

JEL codes: C14, C15, H2, I32

1University of Milan
2London School of Economics
1 Introduction

Many economic phenomena are known to follow distributions with non-negligible probability of extreme events, termed thick tailed (TT) distributions. Top income and wealth distributions are often modelled with infinite variance Pareto distributions (see among others Cowell, 1995). The distribution of cities by size seems to fit a power law, i.e., the number of cities with population greater than a given threshold $C$ is proportional to $1/C$ (Zipf, 1949). Zipf’s law is a discrete form of a Pareto distribution with infinite variance (Gabaix, 1999). Also the distributions of firms by size appear to exhibit non-negligible probabilities of extreme events (Hart and Prais, 1956; Steindl, 1965; Gibrat, 1931). Furthermore, distributions with thick tails are of relevance in finance as it is well established that returns of financial data, as well as size of corporate bankruptcies, are non-normal with thick tails, causing peculiar problems for the regulation of markets where such extremes may be observed (Embrechts, 2001; Danielsson and de Vries, 1997; Loretan and Phillips, 1994).

Another example occurs in economics of information technology where it has been found that Web traffic often presents distributions of file sizes that decline with a power law (Arlitt and Williamson, 1996) and often with infinite variance (Crovella and Bestavros, 1996). For example the lengths of bursts in network traffic and the sizes of files in some systems appear well described by distributions with non-negligible probability of extremely large events. Additional evidence of power-law behavior is present in same data on transmission lengths of network transfers (Bodnarchuk and Bunt, 1991) and in data bytes in FTP bursts (Paxson and Floyd, 1995).

Although there is a large and increasing literature on robust estimation with data obeying thick tail distributions (see among others Victoria-Feser and Dupuis (2003); Hsieh (1999); Beirlant et al. (1996)), little is known about the consequences of performing classical inference using samples drawn from such distributions. This paper aims at this gap in the literature. It provides an analysis of the asymptotic distribution of the classical t-ratio for distributions with no finite variance both in statistical
as well as econometric inference, it discusses how classical testing is affected, and then proposes alternative ways to perform inference with such distributions. To aid the analysis, the paper develops a new distribution termed the “double Pareto,” which allows the thickness of the tails and the existence of moments to be determined parametrically. Section 2 briefly discusses the relevant literature and Section 3 describes the focus of the paper and introduces the double-Pareto distribution. Section 4.1 provides some simulation results for the distribution of the t-ratio with samples drawn from thick tailed distributions. Section 5 discusses the issue of naive testing with Pareto or double-Pareto distributions with finite mean and infinite variance: it discusses the error that a researcher would commit assuming that the standard t-ratio test is normally distributed when in fact it is not. Special cases analyzed include one- and two-sample statistical inference problems, as well as linear regression econometric problems. Solutions are suggested in Section 6. Section 7 concludes.

2 The relevant literature

Distribution functions with infinite first moments belong to the family of thick tail (TT) distributions. In the literature there is no universally accepted definition of a TT distribution. In general, a random variable from a TT distribution presents a non negligible probability of assuming very large values. In other words, TT distributions have more weight in the tails than some reference distribution. When it is assumed that the reference distribution is the normal whose tails decay as the square of an exponential, it implies that distributions with power-law decay, as well as with exponential decay, are considered to be TT distributions. Other, more complete definitions, consider as TT a distribution whose exponential moments are infinite, $E(e^{tx}) = \infty, \forall t \geq 0$, which implies that the moment-generating function does not exist. Since different distributions have different degrees of thick-tailedness, a number of quantitative indicators for evaluating the probability of extremal events have been developed, such as the extremal claim index to assign weights to the tails.
and thus the probability of extremal events (Embrechts et al., 1999). Finally, some cruder though widely used definitions consider as TT a distribution with an infinite variance, or kurtosis larger than 3 (leptokurtic) (Bryson, 1982).

The analysis that follows focuses on the t-ratio applied to data drawn from TT distributions. It is well known that the t-ratio of any sequence \( \{X_i\} \) of i.i.d. random variables with finite first two moments converges towards a standard normal distribution, where the t-ratio is defined as

\[
t = \frac{\bar{X}}{S} \tag{1}
\]

\( \bar{X} = N^{-1} \sum_{i}^{N} X_i \) and \( S = N^{-2} \sum_{i=1}^{N} X_i^2 \), \( i = 1, \cdots, N \). However, TT distributions typically fail the \( EX_i^2 < +\infty \) and possibly the \( EX_i < +\infty \) requirements for convergence.

It is well known that the ratio of two random variables gives rise to a random variable with a possibly bimodal distribution. Such a distribution is derived in Fieiller (1932) and its density is characterized in Phillips (1982). Marsaglia (1965) shows the conditions under which the ratio of two independent normal random variables with variance 1 and different means has a bimodal rather than a unimodal distribution. Phillips and Hajivassiliou (1987) show that the phenomenon of bimodality can also occur with the classical t-ratio test statistic for populations with undefined second moments. They showed that when \( X_1, X_2, \cdots, X_N \) is a random sample from a Cauchy (0,1) population, the numerator and the denominator of \( t_1 \) converge weakly to random variables, which are dependent, as \( n \to \infty \). Hence, asymptotically the t-statistic is a ratio of dependent random variables. In the classical case the numerator and the denominator statistics in the t-ratio are independent and, as \( n \to \infty \), the denominator, properly scaled, converges in probability to a constant. They argue that the dependence of the numerator and denominator in the t-statistic is the main factor that induces the bimodality in the distribution. The fact that the modes are at ±1 comes from simulation evidence that the numerator and denominator of the t-statistic are identical up to the sign.
They suggest studying the distribution of the t-statistic focusing on the dependence between the numerator and denominator statistics. Such dependency remains even in the limit. In fact they showed that $S^2$ converges weakly towards a stable random variate with exponent $\alpha = 1/2$ and that the numerator and the denominator of the t-statistic follow a jointly stable distribution\(^3\).

Our results underscore the point, made in Hajivassiliou (2005), that when data are generated from distributions with thick tails, orthogonality and zero correlation are not only very different properties compared to full statistical independence, but startlingly different results obtain when these properties are swapped. By construction, the random variables in the numerator of the t-ratio, $\bar{X}$, is orthogonal to the $S^2\bar{X}$ variable in the square root of the denominator. Under Gaussianity, this orthogonality implies full statistical independence between numerator and denominator, whether the underlying draws are dependent or independently drawn. But in the case of data drawn from the Cauchy distribution, independency of the numerator and denominator of the t-ratio rests crucially on whether or not the underlying data are independently drawn or not: if they are generated from a multivariate spherical Cauchy with a diagonal scale matrix (and hence they are non-linearly dependent), then the numerator and denominator in fact become independent and the usual unimodal t-distribution results (this is a result due to Zellner, 1976). If, on the other hand, they are drawn fully independently from one another, then $\bar{X}$ and $S^2\bar{X}$ turn

\(^3\)Stable distributions are not in general a subset of TT distribution as they also include the normal distribution. There are four different and equivalent ways to define stable distributions (Samorodnitsky and Taqqu, 1994; Focardi, 2001). A key characterization is that a random variable is said to have a stable distribution if it has the same distribution of the (normalized) independent sum of any number of identical replicas of the same variable. This property involves that the entire distribution be equal and not only the tails. Note that in this context “equal distributions” means that distributions have the same functional form but possibly with different parameters. Formally, a random variable $X$ is said to have a stable distribution if, for any positive numbers $a_i$ and $n$, there exist a positive number $c$ and a real number $d$ such that

$$\sum_{i=1}^{n} a_i X_i \sim cX + d$$  \hspace{1cm} (2)

where $X_i$ are independent copies of $X$. Stability in the limit implies

$$\sum_{i=1}^{n} a_i X_i \stackrel{d}{\to} cX + d$$  \hspace{1cm} (3)

where $\stackrel{d}{\to}$ denotes convergence in distribution.
out to be dependent and hence the distribution of the t-ratio exhibits the striking bimodality.

3 The focus of the analysis

What happens to the t-ratio statistics from random samples when the first moments do not exist? What implications does this behaviour have for hypothesis testing?

For distributions with infinite first moment let us consider the t-ratio statistic defined as:

$$t_1 = \frac{\bar{X}}{S_X} = \frac{\sum_1^N X_i/N}{\sqrt{(\sum_1^N (X_i - \bar{X})^2/N^2)}}$$

(4)

and for distributions with finite first moment, is defined as:

$$t_2 = \frac{\bar{X} - \mu}{S_X} = \frac{\sum_1^N X_i/N - \mu}{\sqrt{(\sum_1^N (X_i - \bar{X})^2/N^2)}}$$

(5)

where $X_1, X_2, \cdots, X_N$ is a random sample from some distribution and $\mu$ is the true mean. It can be proved that the difference between $S$ and $S_X$ is negligible as $N \to \infty$ as well as the difference between $t$ and $t_1$, hence asymptotically they give the same results$^4$: $t_2$ differs from $t_1$ only in the location factor, $-\mu/S_X$.

The first TT distribution considered is the standard Cauchy (0,1). It has density function (DF):

$$f(x) = \frac{1}{\pi(1 + x^2)}$$

(6)

and cumulative distribution function (CDF):

$$F(x) = \frac{1}{\pi} arctan(x) + \frac{1}{2}$$

(7)

All moments of the standard Cauchy (0,1) distribution are infinite.

The Pareto distribution (type I) has DF and CDF are respectively equal to:

$$f(x) = \alpha \beta^\alpha x^{-\alpha-1}$$

(8)

$^4$Formally, $S^2 - S_X^2 = O_p(N^{-1})$ and $t - t_1 = O_p(N^{-1})$ (Phillips and Hajivassiliou (1987), Lemma 1, p. 5.)
\[ F(x) = 1 - \left( \frac{\beta}{x} \right)^\alpha, \quad x \geq \beta, \alpha > 0, \beta > 0 \] (9)

The first moment, \( E(x) \), exists if \( \alpha > 1 \) and the second central moment, \( V(x) \), exists if \( \alpha > 2 \):

\[ E(x) = \frac{\alpha \beta}{\alpha - 1} \] (10)

\[ V(x) = \frac{\alpha \beta^2}{(\alpha - 1)^2(\alpha - 2)} \] (11)

The Pareto distribution belongs to the family of exponential distributions with PDF

\[ p_\alpha(x) = C(\alpha)e^{\sum_{i=1}^k Q_i(\alpha)t_i(x)h(x)} \] (12)

with \( C(\alpha) = \alpha \beta^\alpha \), \( Q_i(\alpha) = -(\alpha + 1) \), \( t_i(x) = \ln x \), \( h(x) = 1 \) (Silvey, 1975). Analogously to the double-exponential (see, Feller, 1971, p. 49), the CDF of the double-Pareto distribution, with \( |x| > \beta \), is:

\[
F(x) = \begin{cases} 
\frac{1}{2} \left( \frac{\beta}{x} \right)^\alpha & \text{iff } x \leq -\beta \\
\frac{1}{2} & \text{iff } -\beta < x < \beta \\
1 - \frac{1}{2} \left( \frac{\beta}{x} \right)^\alpha & \text{iff } x \geq \beta 
\end{cases}
\]

The PDF can be written as:

\[ f(x) = \frac{1}{2} \alpha \beta^\alpha |x|^{-\alpha-1}, \quad |x| > \beta, \beta > 0 \] (13)

and can be seen as the convolution of the Pareto density \( \alpha \beta^\alpha x^{-\alpha-1} \) \( (x \geq \beta, \alpha > 0, \beta > 0) \) with the mirrored density \( \alpha \beta^\alpha (-x)^{-\alpha-1} \) \( (x \leq -\beta, \alpha > 0, \beta > 0) \). In other words, the double-Pareto is the density of \( X_1 - X_2 \) when \( X_1 \) and \( X_2 \) are independent and have the common exponential density \( \alpha \beta^\alpha x^{-\alpha-1} \) \( (x > \beta, \beta > 0, \alpha > 0) \).
Its first two centered moments are (see Section 8):

\[
E(x) = 0, \alpha > 1 \\
V(x) = \frac{2\alpha \beta^2}{(\alpha - 1)^2(\alpha - 2)}, \alpha > 2
\]

4 Simulation results

4.1 Data generated from distributions with finite moments

The results that follow have been obtained via Monte Carlo simulations from random samples of dimension \(N\) using the method of inverted CDF, i.e., a random sample of dimension \(N\) is extracted from a unit rectangular variate, \(U(0, 1)\), and then it is mapped into the sample space using the inverse CDF. The number of simulations \(M\) has been set to 10,000. This study allows one to disentangle some differences about the asymptotic distribution of the t-ratio statistic when either one or both first two moments do not exist.\(^5\)

The Cauchy and the double-Pareto distribution with \(\alpha \leq 1\) are both symmetric and with infinite mean. For these distributions, as sample size increases, the statistic \(t_1\) converges towards a stable distribution which is symmetric and bimodal. The convergence is fairly rapid, even for samples as small as 10, and the two modes are located at \(\pm 1\). As for the double-Pareto, the t-ratio distribution does depend on \(\alpha\): the lower is \(\alpha\), the higher is the concentration around the two modes (Figure 1).

Figure 2 examines the case \(1 < \alpha < 2\). We now see that the t-ratio, \(t_2\), is not always clearly bimodally distributed. The more \(\alpha\) departs from 1 the less evident is the bimodal distribution of the t-ratio and the clearer the convergence towards a standard normal distribution (Figure 2). We set \(\beta = 3\) but these results apply for any value of \(\beta > 0\), since \(\beta\) is simply a threshold parameter that does not affect the \(t_1\) statistic behavior.

\(^5\)Using copulas, we could evaluate behaviour with \textit{correlated} double-Pareto draws. See (Hajivassiliou, 2005) for a development of this idea. See also (Ibragimov et al., 2003) for general results.
Figure 1: $t$-ratio of Cauchy and infinite-first-moment double-Pareto distributions

Figure 2: $t$-ratio of double-Pareto distributions with $1 < \alpha \leq 2$
Moving to the classic Pareto distribution defined on a positive support with \( \alpha \leq 1 \), the t-ratio \( t_1 \) is clearly non-normal. However, the convergence towards a unimodal distribution with mode located just above 1, is clearer the smaller is \( \alpha \). The closer \( \alpha \) gets to 1, the more dispersed the distribution becomes (Figure 3).

![Figure 3: \( t_1 \) Pareto distributions with 0 < \( \alpha \) ≤ 1](image)

When a sample is randomly drawn from a Pareto distribution with \( \alpha \) greater than 1 and less than 2, the t-ratio, \( t_2 \), is still clearly non-normally distributed. Due to the occurrence of large values, the \( t_2 \) distribution is asymmetric and biased towards negative values. The closer \( \alpha \) gets to 2, the clearer the convergence to a standard normal appears. With \( \alpha = 1.8 \), the distribution is still far from normal with strong skewness to the left (Figure 4).

![Figure 4: \( t_2 \) of Pareto distributions with 1 < \( \alpha \) ≤ 2.](image)

The regularity in the \( t_1 \) distribution leads us to investigate the relationship between the first and second centered moments, in the numerator and denominator of \( t_1 \) respectively. Phillips and Hajivassiliou (1987) noted that if the distrib-

\[ \text{\footnotesize Results for the negative Pareto distribution are symmetric to those for the positive Pareto and are not presented here.} \]
bution is Cauchy, the variance converges toward a unimodal distribution with the mode lying in the interval \((0, 1)\). However, if the distribution is either an infinite-mean Pareto or double-Pareto, the sample variance does not converge towards a stable distribution but becomes more dispersed as the sample size increases (Figure 5). This behaviour confirms the surprising results obtained elsewhere (Ibragimov (2004), Hajivassiliou (2005)) concerning inference with thick-tailed (TT) and extremely-thick-tailed (ETT) distributions\(^7\): in the first case, the dispersion of the distribution of sample averages remains invariant to the sample size \(N\), while in the second more observations actually hurt with the variance rising with \(N\). Furthermore, the usual asset diversification result that spreading a given amount of wealth of a larger number of assets reduces the variability of the portfolio no longer holds: with returns from a TT distribution the variability may remain invariant to the number of assets composing the portfolio, while in the ETT case portfolio variability actually rises with the number of assets. In such cases, all eggs should be placed in the same basket.

Moreover, it can be seen that there exists a strong parabolic relationship between the mean \((\overline{X})\) and the centered second moment \((S^2_X)\) when the first moment is infinite. All the distributions with infinite mean show a clear convex parabolic relationship between the \(S^2_X\) and \(\overline{X}\) (Figure 6).

A simple OLS estimate for the coefficient \(b\) of the parabolic relationship \((S^2_X = a + b\overline{X}^2)\) is always very close to 1 and highly significant using the Cauchy, the Pareto or the double-Pareto with \(\alpha \leq 2\). However, the coefficient \(a\) is not significantly different from zero for any value of the sample size\(^8\). In other words, the average of the squared deviation from the sample mean can be well approximated by the square of the sample mean. This property is a direct consequence of the fact that the Pareto distribution with infinite variance belongs to the class of subexponential distributions, characterized by two properties: the convolution closure property and

\(^7\)For the class of distributions for which \(Pr(|X| > c) = c^{-\alpha}\) a particular distribution is said to be extremely-thick-tailed (ETT) if \(\alpha < 1\). The Cauchy distribution corresponds to boundary case where \(\alpha = 1\).

\(^8\)Phillips and Hajivassiliou (1987) found a \(b\) coefficient between .570 and .376 for the Cauchy distribution and different sample sizes. However, in their regression the dependent variable was the uncentered second moment while the centered one is considered here.
Figure 5: Distribution of the variance of some distributions with infinite mean.

Figure 6: Relationship between the mean and the variance of some distributions with infinite mean.
the property of the sum (Embrechts et al., 1999). The first property states that the shape of the tail is preserved after the summation of a random sample from a given subexponential distribution. The second property states that in a sum of observations from a random sample, the largest value will be of the same order of magnitude as the sum itself\(^9\). The latter property implies that the deviation from the mean will be of the same order of magnitude as the mean, hence the ratio between the mean of the squared deviation from the mean and the squared mean will be of the same order of magnitude. The fact that the modes of the bimodal distribution for the \(t_1\) statistic are at ±1 comes from this property and the fact that the sample mean can be negative whereas its standard error cannot.

4.2 Data generated from distribution with thick tails but finite moments

In order to investigate the relative importance of tail thickness and non-existence of moments, we consider a Cauchy distribution truncated on a compact support, characterized as follows:

\[
Z = \begin{cases} 
X & \text{iff } |X| < c \\
NA & \text{otherwise}
\end{cases}
\]

where \(X\) is a standard \(\text{Cauchy}(0,1)\). The PDF of this truncated distribution is

\[
f(z) = \frac{1}{\pi (1 + z^2)^{\frac{3}{2}}} \arctan(c) \quad \text{over } |z| < c.
\]

The cutoff parameter \(c\) is a positive finite real number. Since the support of this distribution is by construction finite and compact, the moments of

\[Formally, for any sample size \(N\), if \(Z_N(x) = \sum_{i=1}^N X_i\) is the sum of i.i.d. random variables and \(M_N\) is their maxima, it is verified that

\[
\lim_{x \to \infty} \frac{P(Z_N > x)}{P(M_N > x)} = 1
\]
The r.v. \( Z \) are all finite.

The second trimmed distribution we consider is the double-Pareto truncated on a compact support as in 15 where \( X \) is a r.v. distributed according to the double-Pareto law introduced in Section 3. The PDF of the truncated double-Pareto distribution is:

\[
f_Z(z) = \begin{cases} 
\frac{1}{\beta} \frac{1}{\alpha |x|^{-\alpha-1}} & \text{iff } \beta < |x| < c \\
0 & \text{otherwise}
\end{cases}
\]

As with the trimmed Cauchy, the cutoff parameter \( c \) is a large finite real number.

Since the support of this distribution is by construction finite and compact, the moments of r.v. \( Z \) are all finite irrespective of the parameter \( \alpha \).

In the simulations below, we consider the following truncation points:

<table>
<thead>
<tr>
<th>Truncated Cauchy</th>
<th>( c )</th>
<th>100</th>
<th>1000</th>
<th>2000</th>
<th>5000</th>
</tr>
</thead>
<tbody>
<tr>
<td>prob(cutoff tails)</td>
<td>0.006</td>
<td>0.0006</td>
<td>0.0003</td>
<td>0.00012</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Truncated double-Pareto (( \beta = 3 ))</th>
<th>( c )</th>
<th>100</th>
<th>1000</th>
<th>2000</th>
<th>5000</th>
</tr>
</thead>
<tbody>
<tr>
<td>prob(cutoff tails), ( \alpha = 0.5 )</td>
<td>0.173</td>
<td>0.055</td>
<td>0.039</td>
<td>0.024</td>
<td></td>
</tr>
<tr>
<td>prob(cutoff tails), ( \alpha = 1.1 )</td>
<td>0.02</td>
<td>0.002</td>
<td>0.0008</td>
<td>0.0003</td>
<td></td>
</tr>
<tr>
<td>prob(cutoff tails), ( \alpha = 1.8 )</td>
<td>0.002</td>
<td>0.00002</td>
<td>0.000008</td>
<td>0.000002</td>
<td></td>
</tr>
</tbody>
</table>

The general conclusion we draw from this set of results is that the source of the bimodality is the rate of tail behaviour and \textit{not} unboundedness of support or non-existence of moments.

5 Testing with TT distribution

The preceding results are of relevance for hypothesis testing in regressions with error terms that are independent and identically distributed from a TT distribution. It
is of interest also for testing the hypothesis of difference in means or other statistics of two samples when either or both come from a TT distribution.

How serious are the mistakes in such cases if the classical t-ratio test statistic is compared with the critical values of a $N(0, 1)$ distribution? This problem is illustrated using the $p$–value discrepancy plot (Davidson and MacKinnon, 1998). The $p$–value discrepancy plot is based on the empirical distribution function (EDF) of the $p$–values of some test statistic $\tau$, generated via Monte Carlo simulation using a data-generating process (DGP) that is a special case of the null hypothesis. The simulation is usually carried out for a large number of $M$ replications obtaining simulated values $\tau_j, j = 1, 2, \cdots, M$. The $p$–value of the $\tau_j$ is the probability of observing a value of $\tau$ more extreme than $\tau_j$, according to some distribution $F(\tau)$. This distribution could be the asymptotic distribution of $\tau$, derived numerically or theoretically, as well as other distributions such as an approximation derived by bootstrapping. The $p$–value is a function of $\tau_j$, $p_j \equiv p(\tau_j)$. Assuming $\tau$ is asymptotically distributed as a standard normal with DF $\phi(z)$ and CDF $\Phi(z)$, then $p_j = 1 - \Phi(\tau_j)^{10}.$

The EDF of the $p_j$ is an estimate of the CDF of $p(\tau)$. At any point $x_i$ in the $(0, 1)$ interval, it is defined by

$$\hat{F}(x_i) \equiv \frac{1}{m} \sum_{j=1}^{m} I(p_j \leq x_i)$$

where $I(p_j \leq x_i)$ is a Boolean operator that takes the value 1 if the argument is true and 0 if not true. Although the function (16) can be evaluated at every data point, when $m$ is large it is only necessary to produce a reasonable picture of the $(0, 1)$ interval or one of its portions. In these applications 1000 equally spaced data points are considered, $x_i, i = 1, 2, \cdots, 1000$. The simplest graph that can be analyzed is the plot of $\hat{F}(x_i)$ against $x_i$. However, for dealing with test statistics that are well behaved, it is more revealing to plot the $p$–value discrepancy plot, namely $\hat{F}(x_i) - x_i$ against $x_i$.

\footnote{For a two-sided test, the $p$–value is $p_j \equiv p(|\tau_j|) = 2(1 - \Phi(\tau_j))$.}
5.1 Inference on Location Parameter from a Single Sample

The \( p \)-value discrepancy plot of the t-ratio statistic, for the Pareto and the double-Pareto with different values of \( \alpha \) was constructed as in (16), where the \( p \)-value is derived both using the standard normal and the distributions derived previously by simulation. The \( p \)-value discrepancy plot allows one to distinguish at a glance among test statistics that systematically over-reject, those that under-reject and test statistics that reject about the right proportion of times at each desired level of \( x_i \): in the first case the plot will be over, in the second below, in the third around the zero line.

Let us now assume that we have a random sample from a double Pareto distribution with \( 1 < \alpha \leq 2 \) and we run a test \( H_0 : \mu = \mu_0 \) against the alternative \( H_A : \mu \neq \mu_0 \), where \( \mu \) is the true mean and \( \mu_0 \) some value on the real line. The sample mean is used to estimate \( \mu \). Performing such a test using the standard normal rather than the correct distribution causes the null hypothesis to be under-rejected by quite a small amount, not larger than 5\% for tests of size 5\%, and even less for tests of size 1\% or 10\%. This conclusion would often lead us to ignore the caveat of having a systematic error in rejection probability using the standard normal for testing two-sided hypothesis with a double-Pareto distribution with \( 1 < \alpha \leq 2 \). However, two important points should be noted.

The first conclusion is that the policy of ignoring the true nature of the t-ratio distribution under this particular DGP may be an acceptable policy if the size of the test is smaller than 10\%. If the test has a larger size - for instance 40\% - the ERP can be larger than 10 and is obviously more difficult to tolerate\(^{11}\). Clearly, the former policy corresponds to minimizing the type II error as opposed to minimizing the type I error, as it is typically performed in economics and several other disciplines. In such cases it is common to find confidence intervals with about 60\% coverage probability (see for instance Karlen, 2002).

\(^{11}\) Although tests of size larger than 10\% are rather unusual in economics it is much less so in other disciplines, such as physics, where the main point is often to maximize the power of the test, rather than to minimize its size. Also in physics and other related sciences, it is common to consider the “probable error” of a test procedure, which corresponds to a significance level of 50\%. 

Secondly, the “ignore” policy leads to major errors in the case of the Pareto distribution. The ERP for a two sided test about the mean of a Pareto distribution with \(1 \leq \alpha < 2\) can be quite larger than 10%. For instance if \(\alpha = 1.1\), the test will over-reject \(H_0\) about 60% of times (Figure 7), even for tests of size 5%. This result clearly comes from the non standard distribution of \(t_2\) (Figure 4). The same concerns apply to one-sided tests: standard testing is highly unreliable. For instance, a test of the hypothesis \(H_0 : \mu = \mu_0\) vs. \(H_A : \mu > \mu_0\), for the Pareto distribution with \(1 < \alpha \leq 2\), assuming asymptotic normality, will seriously under-reject with an ERP that increases with the size up to the 40% level. For test of the hypothesis \(H_0 : \mu = \mu_0\) vs. \(H_A : \mu < \mu_0\) the test will dramatically over-reject with an ERP which can be larger than 60%, even for tests of size 5% (Figure 8).

![Figure 7: ERP for two-tail test with a sample from a symmetric or a positive definite Pareto with \(1 \leq \alpha \leq 2\).](image)

### 5.2 Two-Sample Inference Problems

Obviously, the non standard distribution of the t-ratio with infinite second moment does also affect the two-sample test of difference of means. Let us assume that we have two independent samples from two different distributions, one of which is a Pareto distribution with infinite first or second moment. Call the two distributions A and B. We want to test whether the mean of the first is different from the mean of the second using the t-ratio

\[
t_2^D = \frac{\mu_A - \mu_B}{\sqrt{(S_X^A/N_A) + (S_X^B/N_B)}}
\]

(17)
Figure 8: ERP for one-tail test with a sample from a symmetric or a positive definite Pareto with $1 \leq \alpha \leq 2$.

where $\mu_A, \mu_B$ are the true means, and $S^A_X, S^B_X$ are the sample variance of $A$ and $B$, respectively. The distribution of $t^D_2$ is again non-standard. Moreover, in many cases it does not look to converge to a stable distribution as the sample size increases. Figure 9 shows via Monte Carlo simulations the distributions of the t-ratio, $t^D_2$, for testing the difference in means of two Pareto distributions that may differ in $\alpha$ but are constrained to have the same location parameter, $\beta = 3$, on the assumption that a sample of the same size has been drawn from each. Clearly, there is no point in using the sample t-ratio and comparing it with the normal critical values. This result comes from the fact that $t^D_2$ is a convolution of two stable distributions with different tail-thickenedness parameters. In general, convolutions of stable distributions also have a stable distribution only if the stable distributions involved present the same thickenedness (rate of decay) of the tails (see for instance Samorodnitsky and Taqqu, 1994).
5.3 Inference in Linear Regression with TT Errors

Finally, let us consider what are the effects on inference in linear regression when errors are TT. For reasons of space, only results with Cauchy error terms will be presented here. Let us consider two simple regression models with an exogenous regressor:

\[ y = \alpha + \beta x + \epsilon \]  \hspace{1cm} (18)

\[ y = \beta x + \epsilon \]  \hspace{1cm} (19)

The t-ratio considered are:
\[
\begin{align*}
  t_{2}^{\alpha} &= \frac{\hat{\alpha} - \alpha}{\sqrt{V(\hat{\alpha})}} \\
  t_{2}^{\beta} &= \frac{\hat{\beta} - \beta}{\sqrt{V(\hat{\beta})}}
\end{align*}
\]

In our exercise we generate some i.i.d. random variables:

- \( \epsilon \): \( \epsilon \sim \text{Cauchy}(0, 1) \);
- \( x \): \( x \sim U(a, b) \) or \( x \sim N(\mu, \sigma^2) \), where \( U(a, b) \) is a rectangular distribution with support \([a, b] \) and \( N(\mu, \sigma^2) \) is a normal distribution with mean \( \mu \) and variance \( \sigma^2 \).

Hence, \( y \) has been generated setting \( \beta = 3 \) in models (18) and (19), and \( \alpha = 10 \) in model (18).

Experiments are carried out generating samples of different dimensions, \( N = 5, 10, 100, 500 \), and with a number of replications \( M = 10,000 \).

Figures 10-11 and 12-13 present t-ratio densities for \( \hat{\alpha}_{OLS} \) and \( \hat{\beta}_{OLS} \) of model (18). Figure 14 and 15 present those for model (19).

Results can be summarized as follows (see also Table 1):

1. when the regression errors are \( \text{Cauchy}(0, 1) \) distributed, the distribution of the t-ratio of the coefficients of a simple linear regression estimated using OLS depends on the distribution of the regressors.

2. If the intercept is included (Model (18)), there is no sign of bimodality in the t-ratio of the slope coefficient.

3. The relevant parameter for bimodality is the mean relative to the variance. In particular:
   - the smaller it is, the clearer is bimodality in the intercept t-ratio of Model (18);
• the larger it is, the clearer is bimodality in the slope t-ratio of Model (19)).

To summarize, in simple regressions with Cauchy(0, 1) errors:

• One should be aware of the risk for a careful inference on the slope coefficient(s) of not including the intercept.

• One should be aware of the fact that the t-ratio of the intercept in a simple regression might be seriously misleading.

• Cauchy processes are infinite moment such as I(1) processes (e.g. random walk with or without drift): could this similarity be useful to look for a theory?

\[
\begin{array}{|c|c|c|c|c|c|c|}
\hline
\text{graph label} & \text{distrib.} & \text{mean} & \text{var.} & \text{Mod. (18) bimod. in } t_2^\alpha & \text{Mod. (18) bimod. in } t_2^\beta & \text{Mod. (19) bimod. in } t_2^\beta \\
\hline
\text{Um1p1} & \text{U(-1,1)} & 0 & 0.3 & \text{yes} & \text{no} & \text{no} \\
\text{U02} & \text{U(0,2)} & 1 & 0.3 & \text{no} & \text{no} & \text{no} \\
\text{U9_11} & \text{U(9,11)} & 10 & 0.3 & \text{no} & \text{no} & \text{yes} \\
\text{U9999_10001} & \text{U(9999,10001)} & 10000 & 0.3 & \text{no} & \text{no} & \text{yes} \\
\text{Um40_p60} & \text{U(-40,60)} & 10 & 833.3 & \text{yes} & \text{no} & \text{no} \\
\text{Um490_p510} & \text{U(-490,510)} & 10 & 83333.3 & \text{yes} & \text{no} & \text{no} \\
\text{N0_1} & \text{N(0,1)} & 0 & 1 & \text{yes} & \text{no} & \text{no} \\
\text{N1_1} & \text{N(1,1)} & 1 & 1 & \text{no} & \text{no} & \text{no} \\
\text{N100_1} & \text{N(100,1)} & 100 & 1 & \text{no} & \text{no} & \text{yes} \\
\text{N1_100} & \text{N(1,100)} & 1 & 100 & \text{yes} & \text{no} & \text{no} \\
\hline
\end{array}
\]

Table 1: Summary of results for OLS estimation of models (18) and (19).
6 Suggested Solution: Use Medians, Quantiles, and LAD Estimation

The main problem with testing with TT distribution, such as the Pareto and double-Pareto with $1 < \alpha \leq 2$, comes from the fact that the t-ratio distribution converges towards a distribution which is clearly not the standard normal. It also depends on unknown parameters, such as the $\alpha$ in the Pareto and double-Pareto distributions. Provided we had full knowledge of the true $\alpha$ parameter of the Pareto distribution from which the sample has been drawn, the first solution would be to correct the critical values using simulation methods. This solution would imply deriving the t-ratio distribution via Monte Carlo simulations as performed in Section 4.1, and tabulating the relevant critical values.

There are two complications, however, that make this solution not very appealing. The first is that if the true $\alpha$ parameter were known a test of the mean would revert to a problem of estimation of the threshold parameter $\beta$: $\alpha$ and $\beta$ fully determine the mean in the Pareto distribution. Since $\beta$ in some cases is given, testing a hypothesis about the true mean would be meaningless. More interestingly, the second complication derives from the fact that the parameter $\alpha$ is rather difficult to estimate with confidence even in large samples: the Maximum Likelihood estimation of $\alpha$, being based on the sample mean, which is highly non-robust, is highly non-robust itself and presents a large variability (Rytgaard, 1990); the Hill estimator (Hill, 1975), which is based on ordered statistics and in its simplest form produces a plot to identify the $\alpha$ parameter, in many cases is totally unhelpful, producing what have been defined “Hill horror plots” (Embrechts et al., 1999). Hence, it is not possible to derive the true distribution of the t-ratio using Monte Carlo simulation since it changes quite significantly for different values of $\alpha$, as we saw in Section 4.1.

An alternative solution is then to consider a more robust statistic than the mean, such as the median (Amemiya, 1985, among others).

Let $X_1, \cdots, X_N$ be a sample from a continuous distribution $F$, defined on the real line, and $X_1 < X_2 < \cdots, < X_N$ be the order statistics, obtained arranging
the observations in increasing orders without ties. The \( p \)-th quantile of \( F \) is defined as \( x_p = F^{-1}(p) \), and the \( p \)-th sample quantile is defined as \( X_k \) where \( k = [pN] \) is the smallest integer greater than or equal to \( pN \). Provided the DF \( f(x) \) exists and is continuous and positive in a neighborhood of some quantile, then the joint distribution of the corresponding sample quantile is asymptotically normal. For the median, \( x.5 \), it can be proved (Ferguson, 1996, Ch. 13) that:

\[
\sqrt{N}(X_{.5N} - x.5) \xrightarrow{d} N\left(0, \frac{1}{4f(x.5)^2}\right)
\]  

(20)

and the \( t \)-ratio statistic is

\[
t_3 = \frac{X.5 - x.5}{S_X}
\]

(21)

where \( X.5 \) and \( x.5 \) are the sample and true median, respectively, and \( S_X = \frac{1}{2\sqrt{N}f(x.5)} \) where \( \hat{f}(x.5) \) is a consistent estimate of \( f(x.5) \). The asymptotic normality of \( t_3 \) can be also seen in Figure 16 and 17, where \( \hat{f}(x.5) \) has been estimated using a kernel density estimator with fixed bandwidth\(^{12}\).

Performing inference with \( t_3 \) implies testing the hypothesis \( H_0 : x.5 = \eta \) vs. \( H_A : x.5 \neq \eta \) for a two-tail test or, \( H_0 : x.5 = \eta \) vs. \( H_A : x.5 > \eta \) for a right-tail test and \( H_0 : x.5 = \eta \) vs. \( H_A : x.5 < \eta \) for a left-tail test, where \( \eta \) is some number on the real line. In these cases, using a sample size \( N = 100 \), the ERP is never larger than 4\% for tests of size less than 10\%. The ERP is always smaller than 5\% even with larger test sizes for two-tail and left-tail tests. It is negligible for left-tail tests (Figure 18). The convergence to a standard normal distribution of \( t_3 \) can be directly applied to linear regressions with TT errors and to two-sample tests of difference in means. In the former case the \( t \)-ratio statistics will be normally distributed because of (20); in the latter case the \( t \)-ratio statistic of the difference in means of two independent samples is equal to a convolution of two normal distributions which is

\(^{12}\)The bandwidth used here is the Silverman’s rule-of-thumb bandwidth:

\[
h_N = 0.9A(N)^{1/5}
\]

(22)

where \( A = min\{\text{standard deviation, interquantile range}/1.34\} \) and \( N \) is the sample size (Silverman, 1986, p. 48). This bandwidth was chosen as it copes well with a wide range of densities, it is trivial to compute and our results are substantially unchanged using alternative methods for bandwidth selection.
7 Conclusions

This paper has investigated the issue of performing inference with TT distributions, which are often found in various fields of economics. It has been shown that when the distribution is TT, and the first moment is finite while the second is not, the standard t-ratio does not asymptotically converge to a standard normal distribution. Hence, it was discussed when inference is invalidated and how relevant the ERP can be. A simple road map is suggested to the careful researcher: whenever she suspects that the sample could come from a TT distribution with infinite variance, she should perform a inference using the sample median or other quantiles for statistical problems and LAD estimation for regression models. In such cases it is possible to compute a t-ratio statistic based on such estimators, and contrast them to the critical values of the standard normal distribution.

The paper presents evidence that the solution of constructing t-ratios based on the median is superior to the classical t-ratio and should be preferred in many situations. However, the median-based t-ratio statistic could still present problems in small samples, where dead intervals - especially in the tails - are more frequent (for an analysis of the robustness properties of quantiles, see Cowell and Victoria-Feser, 2002). We interpret the main results of our paper as issuing powerful cautions for the need of adapting suitably estimation and inference procedures to the special problems induced by TT and ETT distributions: standard results and procedures may be very seriously misleading indeed.
8 Appendix A: Moments of the symmetric Pareto distribution

The first moment of the double-Pareto distribution can be obtained using convolution:

\[
E(X) = \int_{-\infty}^{\infty} t \left[ \int_{\beta}^{\infty} (\alpha \beta)^2 (x + t)^{-\alpha-1} x^{-\alpha-1} dx \right] dt = \\
= (\alpha \beta)^2 \int_{\beta}^{\infty} x^{-\alpha-1} \left[ \int_{-\infty}^{\infty} t(x + t)^{-\alpha-1} dt \right] dx = \\
= (\alpha \beta)^2 \int_{\beta}^{\infty} x^{-\alpha-1} \left[ \int_{\beta-x}^{\infty} t(x + t)^{-\alpha-1} dt \right] dx = \\
= (\alpha \beta)^2 \int_{\beta}^{\infty} x^{-\alpha-1} \left[ \int_{\beta}^{\infty} (z - x)(z)^{-\alpha-1} dz \right] dx = \\
= (\alpha \beta)^2 \int_{\beta}^{\infty} x^{-\alpha-1} \left[ \frac{1}{-\alpha + 1} z^{-\alpha+1} \left|_{\beta}^{\infty} \right| + x \frac{1}{\alpha} z^{-\alpha} \left|_{\beta}^{\infty} \right| \right] dx = \alpha > 1 \\
= (\alpha \beta)^2 \int_{\beta}^{\infty} x^{-\alpha-1} \left[ \frac{\beta^{-\alpha+1}}{\alpha - 1} - \frac{x \beta^{-\alpha}}{\alpha} \right] dx = \\
= (\alpha \beta)^2 \int_{\beta}^{\infty} \frac{\beta^{-\alpha+1}}{\alpha - 1} x^{-\alpha-1} - \frac{\beta^{-\alpha}}{\alpha} x^{-\alpha} dx = \\
= (\alpha \beta^{2\alpha}) \left[ \frac{\beta^{-\alpha+1}}{\alpha - 1} - \frac{\beta^{-\alpha}}{\alpha - \alpha + 1} x^{-\alpha+1} \left|_{\beta}^{\infty} \right| \right] = \\
= 0
\]
The second central moment of the double-Pareto distribution is:

\[ V(X) = (\alpha \beta)^2 \int_{-\infty}^{\infty} t^2 \int_{-\infty}^{\infty} x^{-\alpha-1}(x + t)^{-\alpha-1}dxdt = \]

\[ = (\alpha \beta)^2 \int_{\beta}^{\infty} x^{-\alpha-1} \int_{\beta}^{\infty} t^2(x + t)^{-\alpha-1}dtdx = \]

\[ = (\alpha \beta)^2 \int_{\beta}^{\infty} x^{-\alpha-1} \int_{\beta}^{\infty} (z - x)^2(z)^{-\alpha-1}dzdx = \]

\[ = (\alpha \beta)^2 \int_{\beta}^{\infty} x^{-\alpha-1} \left[ \int_{\beta}^{\infty} z^{-\alpha-1}dz - 2 \int_{\beta}^{\infty} xz^{-\alpha}dz + \int_{\beta}^{\infty} x^2z^{-\alpha-1}dz \right] dx = \quad \alpha > 2 \]

\[ = (\alpha \beta)^2 \int_{\beta}^{\infty} x^{-\alpha-1} \left[ \frac{-\beta^{-\alpha+2}}{-\alpha + 2} + \frac{2x\beta^{-\alpha+1}}{-\alpha + 1} - \frac{x^2\beta^{-\alpha}}{-\alpha} \right] dx = \]

\[ = (\alpha \beta)^2 \left[ \int_{\beta}^{\infty} \frac{\beta^{-\alpha+2}}{\alpha - 2} x^{-\alpha-1}dx - \int_{\beta}^{\infty} \frac{2\beta^{-\alpha+1}}{\alpha - 1} x^{-\alpha}dx + \int_{\beta}^{\infty} \frac{\beta^{-\alpha}}{\alpha} x^{-\alpha+1}dx \right] = \]

\[ = (\alpha \beta)^2 \left[ \frac{\beta^{-\alpha+2}}{\alpha - 2} \frac{\beta^{-\alpha}}{\alpha} - \frac{2\beta^{-\alpha+1}}{\alpha - 1} \frac{\beta^{-\alpha+1}}{\alpha - 1} + \frac{\beta^{-\alpha}}{\alpha} \frac{\beta^{-\alpha+2}}{\alpha - 2} \right] = \]

\[ = \alpha^2 \left[ \frac{\beta^2}{\alpha(\alpha - 2)} - \frac{2\beta^2}{(\alpha - 1)^2} + \frac{\beta^2}{\alpha(\alpha - 2)} \right] = \]

\[ = \frac{2 \alpha \beta^2 [((\alpha - 1)^2 - \alpha(\alpha - 2))]}{(\alpha - 2)(\alpha - 1)^2} = \]

\[ = 2 \frac{\alpha \beta^2}{(\alpha - 2)(\alpha - 1)^2} \]
References


28


Figure 10: t-ratio of OLS estimation of model (18), Part A.
Figure 11: t-ratio of OLS estimation of model (18), Part B.
Figure 12: t-ratio of OLS estimation of model (18), Part A.
Figure 13: t-ratio of OLS estimation of model (18), Part B.
Figure 14: t-ratio of OLS estimation of model (19)
Figure 15: t-ratio of OLS estimation of model (19)
Figure 16: $t_3$ with infinite first moment distributions

Figure 17: $t_3$ with Pareto distribution with $1 < \alpha \leq 2$

Figure 18: ERP with $t_3$ from a Pareto distribution with $1 < \alpha \leq 2$