Bimodal t-ratios:

The Impact of Thick Tails on Inference

Carlo V. Fiorio[1],
Vassilis A. Hajivassiliou[2]
and
Peter C.B. Phillips[3]

April 2008

Abstract

This paper studies the distribution of the classical t-ratio with data generated from distributions with no finite moments and shows how classical testing is affected by bimodality. A key condition in generating bimodality is independence of the observations in the underlying data generating process (DGP). The paper highlights the strikingly different implications of lack of correlation versus statistical independence in DGPs with infinite moments and shows how standard inference can be invalidated in such cases, thereby pointing to the need for adapting estimation and inference procedures to the special problems induced by thick-tailed (TT) distributions.

The paper presents theoretical results for the Cauchy case and develops a new distribution termed the "double Pareto," which allows the thickness of the tails and the existence of moments to be determined parametrically. It also investigates the relative importance of tail thickness in case of finite moments by using TT distributions truncated on a compact support, showing that bimodality can persist even in such cases. Simulation results highlight the dangers of relying on naive testing in the face of TT distributions.

JEL codes: C12, C15, C46

[1]University of Milan
[2]London School of Economics
[3]Yale University, University of Auckland, Singapore Management University

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) to address this feature of the data. The distribution of cities by size also 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). Another example is the size distribution of firms, which appears to exhibit a non-negligible probability of extreme events (Hart and Prais, 1956; Steindl, 1965; Gibrat, 1931). Further, TT distributions frequently arise in financial return data as well as data on financial phenomena such as corporate bankruptcies, which can cause difficulties in regulating markets where such extremes are observed (Embrechts, 2001; Danielsson and de Vries, 1997; Loretan and Phillips, 1994).

A different example arises in the economics of information technology where it has been found that Web traffic often presents distributions of file sizes that decline according to a power law (Arlitt and Williamson, 1996) and often with infinite variance (Crovella and Bestavros, 1996). For instance, the lengths of bursts in network traffic and the sizes of files in some systems appear to be 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 growing literature on robust estimation with data following 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. Important exceptions are Logan et al. (1973), which drew early attention to the possibility of bimodal distributions in self normalized sums of independent random variables, Marsaglia (1965) and Zellner (1976, 1978), who showed bimodality for certain ratios of normal variables, Phillips and Wickens (1978), who showed that the distribution of structural equation estimators was not always unimodal, and Phillips and Hajivassiliou (1988), who analyzed bimodality in t classical ratios. Nelson and Startz (1990) and Maddala and Jeong (1992) provided some further analysis of structural estimators with possibly weak instruments. More recent contributions on this latter topic include Woglom (2001), Hillier ((2005), Forchini (2005), and Phillips (2008), all of whom consider bimodality in structural equation distributions. Not much emphasis in this literature has been placed on the difference between orthogonal observations and full independence.

The present paper contributes to this literature in several ways. It provides an analysis of the asymptotic distribution of the classical t-ratio for distributions with no finite variance and discusses how classical testing is affected. 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.

The plan of the paper is as follows. We clarify the concept of TT distributions and provide a theoretical analysis of the bimodality of the t-ratio with data from an iid Cauchy distribution in Section 2. Some simulation analysis is given in Section 3 and extensions to the Stable family of distributions are discussed in Section 3.2.

Section 4 considers the different implications of lack of correlation and statistical independence. Section 5 illustrates some alternative results for a symmetric double Pareto distribution. Section 6 investigates inference in the context of t-ratios with TT distributions, considering single- and double-sample hypothesis testing. Section 7 shows that bimodality can arise even with TT distributions trimmed to have finite support. Other complications with TT distributions are discussed in Section 8, while Section 9 concludes.

2 Cauchy DGPs and Bimodality of the t-statistic

While, there is no universally accepted definition of a TT distribution, random variables drawn from a TT distribution have a non negligible probability of assuming very large values. Distribution functions with infinite first moments certainly belong to the family of thick tail (TT) distributions. Different TT distributions have differing degrees of thick-tailedness and, accordingly, quantitative indicators have been developed to evaluate the probability of extremal events, such as the *extremal claim index* to assign weights to the tails and thus the probability of extremal events (Embrechts et al., 1999). A crude though widely used definition describes any distribution with infinite variance as a TT distribution. Other weaker definitions require the kurtosis coefficient to larger than 3 (leptokurtic) (Bryson, 1982).

In this paper we say that a distribution is thick-tailed (TT) if it belongs to the class of distributions for which $Pr(|X| > c) = c^{-\alpha}$ and $\alpha \leq 1$. The Cauchy distribution corresponds to boundary case where $\alpha = 1$. Such distributions are sometimes called very heavy tailed.

It is well known that ratios of random variables frequently give rise to bimodal distributions. Perhaps the simplest example is the ratio

$$R = \frac{a+x}{b+y} \tag{1}$$

where x and y are independent $N(0,1)$ variates and a and b are constants. The distribution of R was found by Fieller (1932) and its density may be represented in series form in terms of a confluent hypergeometric function [see (Phillips, 1982)][equation (3.35)]. It turns out, however, that the mathematical form of the density of R is not the most helpful instrument in analyzing or explaining the bimodality of the distribution that occurs for various combinations of the parameters (a, b) Instead, the joint normal distribution of the numerator and denominator statistics, $(a + x, b + y)$ provides the most convenient and direct source of information about the bimodality. An interesting numerical analysis of situations where bimodality arises in this example shows that the density of R is unimodal or bimodal according to the region of the plane in which the mean (a, b) of the joint distribution lies. Thus, when (a, b) lies in the positive quadrant the distribution is bimodal whenever a is large (essentially $a > 2.257$).

Similar examples arise with simple posterior densities in Bayesian analysis and certain structural equation estimators in econometric models of simultaneous equations. Zellner (1978) provides an interesting example of the former, involving the posterior density of the reciprocal of a mean with a diffuse prior. An important example of the latter is the simple indirect least squares estimator in just identified structural equations as studied, for instance, by Bergstrom (1962) and recently

Forchini (2006), Hillier (2006) and Phillips (2006).

The present paper shows that the phenomenon of bimodality can also occur with the classical t-ratio test statistic for populations with undefined second moments. The case of primary interest to us in this paper is the standard Cauchy (0,1) with density

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

When the t-ratio test statistic is constructed from a random sample of n draws from this population the distribution is bimodal, even in the limit as $n \to \infty$. This case of a Cauchy (0,1) population is especially important because it highlights the effects of statistical dependence in multivariate spherical populations. To explain why this is so, suppose $(X_1, \cdots, X_n)$ is multivariate Cauchy with density

$$pdf(x) = \frac{\Gamma\left(\frac{n+1}{2}\right)}{\pi^{(n+1)/2}(1+x'^{(n+1)/2})} \tag{3}$$

This distribution belongs to the multivariate spherical family and may be written in terms of a variance mixture of a multivariate $N(0, \sigma^2 I_n)$ as

$$\int_0^\infty N(0, \sigma^2 I_n) dG(\sigma^2) \tag{4}$$

where $1/\sigma^2$ is distributed as χ_1^2 and $G(\sigma^2)$ is the distribution function of σ^2. Note that the marginal distributions of (3) are all Cauchy. In particular, the distribution of X_i is univariate Cauchy with density as in (2) for each i. However, the components of $(X_1, \cdots, X_n)$ are statistically dependent, in contrast to the case of a random sample from a Cauchy (0,1) population. The effect of this dependence, which is what distinguishes (3) from the random sample Cauchy case, is dramatically illustrated by the distribution of the classical t-statistic:

$$t_X = \frac{\overline{X}}{S_X} = \frac{n^{-1}\Sigma_1^n X_i}{\{n^{-2}\Sigma_1^n (X_i - \overline{X}^2)\}^{1/2}} \tag{5}$$

Under (3), t_X is distributed as t with $n-1$ degrees of freedom, just as in the classical case of a random sample from a $N(0, \sigma^2)$ population. This was pointed out by Zellner (1976) and is an immediate consequence of (4) and the fact that t_X is scale invariant. However, the spherical assumption that underlies (3) and (4) and the dependence that it induces in the sample $(X_1, \cdots, X_n)$ is very restrictive. When it is removed and $(X_1, \cdots, X_n)$ comprise a random sample from a Cauchy $(0, 1)$ population, the distribution of t_X is very different. The new distribution is symmetric about the origin but it has distinct modes around ± 1. This bimodality persists even in the limiting distribution of t_X so that both asymptotic and small sample theory are quite different from the classical case.

We know that the numerator and denominator statistics in the classical t-ratio are independent. Moreover, as $n \to \infty$ the denominator, upon suitable scaling, converges in probability to a constant. By contrast, in the i.i.d. Cauchy case, the numerator and denominator statistics of t_X converge weakly to non-degenerate random variables which are (non-linearly) dependent, so that as $n \to \infty$ the t-statistic is a ratio of random variables. Moreover, it is the dependence between the numerator and denominator statistics (even in the limit) which induces the bimodality in the distribution. These differences are important and, as we shall show, they explain the contrasting shapes of the distributions in the two cases.

Let $(X_1, \cdots, X_n)$ be a random sample from a Cauchy (0,1) distribution with density (2). Define

$$S^2 = n^{-2}\Sigma_1^n X_i^2, S_X^2 = n^{-2}\Sigma_1^n (X_i - \overline{X})^2 \tag{6}$$

$$t = \frac{\overline{X}}{S}, t_X = \frac{\overline{X}}{S_X} \tag{7}$$

Throughout the paper, we will use the symbol "$\Rightarrow$" to signify weak convergence as $n \to \infty$ and the symbol "$\equiv$" to signify equality in distribution.

As is well known $X \equiv$Cauchy (0,1) for all n and, of course, $\overline{X} \to X \equiv$Cauchy (0,1) as $n \to \infty$. Our attention will concentrate on the joint distribution of (X, S^2) and the associated statistic t given in (7). In fact, the distributions of t and t_X are asymptotically equivalent. More specifically, we have:

LEMMA 1

$$S^2 - S_X^2 = O_p(n^{-1})$$

$$t - t_X = O_p(n^{-1})$$

Note that X_i^2 has density

$$pdf(y) = \frac{1}{\pi y^{1/2}(1+y)}, y > 0 \tag{8}$$

In fact, X_i^2 belongs to the domain of attraction of a stable law with exponent $\alpha = 1/2$. To see this we need only verify (Feller, 1971, p. 313) that if $F(y)$ is the distribution function of X_i^2 then

$$1 - F(y) + F(-y) \sim 2/\pi y^{1/2}, y \to \infty$$

which is immediate from (8); and that the tails are well balanced. Here we have:

$$\frac{1 - F(y)}{1 - F(y) + F(-y)} \to 1, \frac{F(-y)}{1 - F(y) + F(-y)} \to 0$$

LEMMA 2

$$S^2 \Rightarrow Y$$

where Y is a stable random variate with exponent $\alpha = 1/2$ and characteristic function given by

$$cf_Y(v) = E(e^{ivY}) = exp\left\{-\frac{2}{\pi^{1/2}}cos\left(\frac{\pi}{4}\right)|v|^{1/2}\left[1 - isgn(v)tan\left(\frac{\pi}{4}\right)\right]\right\} \tag{9}$$

Note that the characteristic function of the limiting variate Y given by (9) belongs to the general stable family, whose characteristic function (see Ibragimov and

Linnik (1971)][p.43]) has the following form:

$$\varphi(v) = exp\left\{i\gamma v - c|v|^{\alpha}\left[1 - i\beta sgn(v)tan\left(\frac{\pi\alpha}{2}\right)\right]\right\} \tag{10}$$

In the case of (10) the exponent parameter $\alpha = 1/2$, the location parameter $\gamma = 0$, the scale parameter $c = 2\pi^{-1/2}cos(\pi/4)$ and the symmetry parameter $\beta = 1$. Lemma 2 shows that the denominator of the t ratio (7) is the square root of a stable random variate in the limit as $n \to \infty$. This is to be contrasted with the classical case where $nS_X^2 \underset{p}{\to} \sigma^2 = E(X_i^2)$ under general conditions.

The density of S^2 is graphed for various values of n in Figure 1. We see that S^2 is unimodal with mode lying in the interval (0,1) for all n. The distribution is very well approximated by the asymptotic even for small values of n ($n \geq 10$) .

Note that when $n = 1$, the numerator and denominator of t are identical up to sign. In this case we have $t = \pm 1$ and the distribution assigns probability mass of 1/2 at +1 and -1. When $n > 1$ the numerator and denominator statistics of t continue to be statistically dependent. This dependence persists as $n \to \infty$. We have:

LEMMA 3

$$(\overline{X}, S^2) \Rightarrow (X, Y)$$

where (X, Y) are jointly stable variates with characteristic function given

$$cf_{X,Y} = exp\left\{-2\pi^{-1/2}(-iv)^{-1/2}{}_1F_1\left(-\frac{1}{2}, \frac{1}{2}; u^2/4iv\right)\right\} \tag{11}$$

where ${}_1F_1$ denotes the confluent hypergeometric function. An equivalent form is

$$cf_{X,Y}(u, v) = exp\left\{-|u| - \pi^{-1/2}e^{-iu^2/4v}\Psi(3/2, 3/2; iu^2/4v)\right\} \tag{12}$$

where Ψ denotes the confluent hypergeometric function of the second kind.

For the definition of the hypergeometric functions that appear in (11) and (12) see Lebedev (1972, Ch. 9). Note that when $u = 0$ (11) reduces to

$$exp\left\{-2\pi^{-1/2}(-iv)^{1/2}\right\} \tag{13}$$

We now write $-iv$ in polar form as

$$-iv = |v|e^{-isgn(v)\pi/2}$$

so that

$$(-iv)^{1/2} = |v|^{1/2}e^{-isgn(v)\pi/4} = |v|^{1/2}cos(\pi/4)\left(1 - isgn(v)tan(\pi/4)\right)$$

from which it is apparent that (11) reduces to the marginal characteristic function of the stable variate Y given earlier in (9). When $v = 0$ the representation (12) reduces immediately to the marginal characteristic function, $exp(-|u|)$, of the Cauchy variate X . In the general case the joint characteristic function $cf_{XY}(u, v)$ does not factorize and X and Y are dependent stable variates.

Figures 2a-d show Monte Carlo estimates (by smoothed kernel methods) of the joint probability surface of $(\overline{X}, S^2)$ for various values of n. As is apparent from the

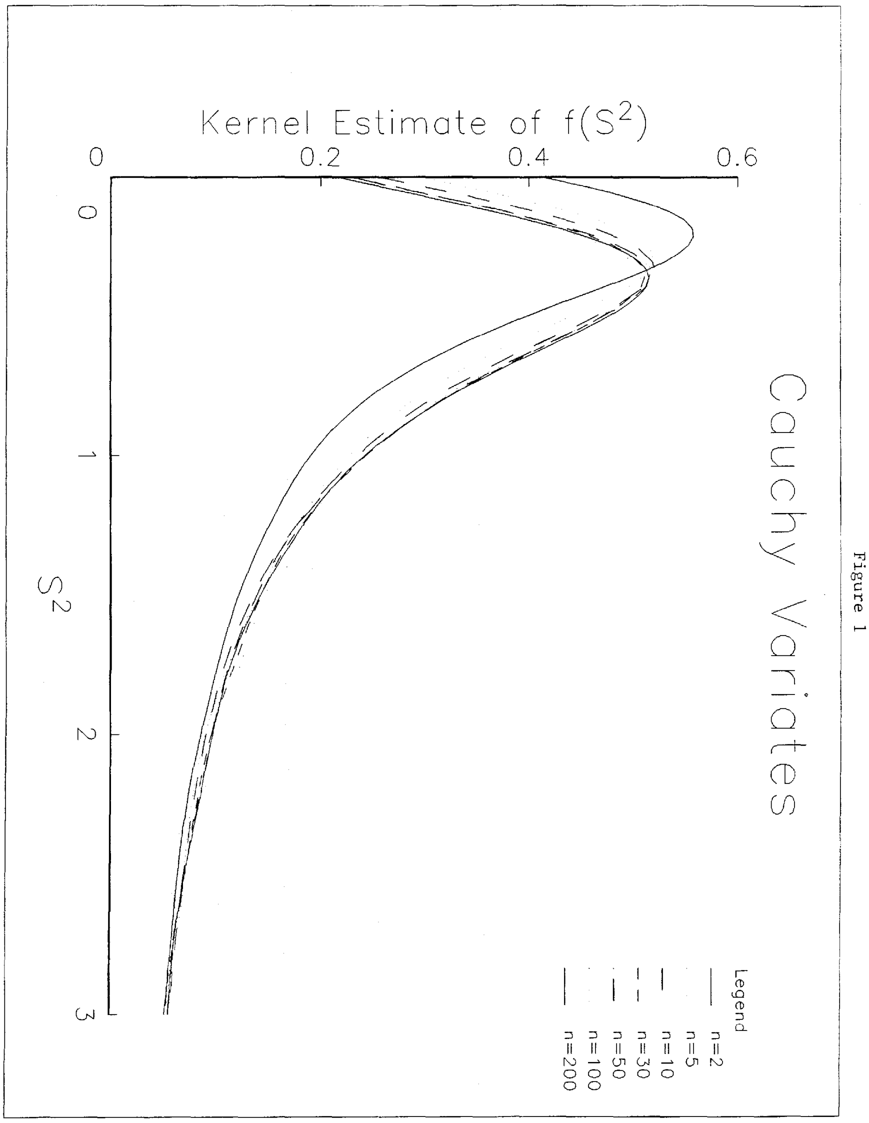


Figure 1: Density Function Estimates of S^2 for the iid Cauchy DGPs

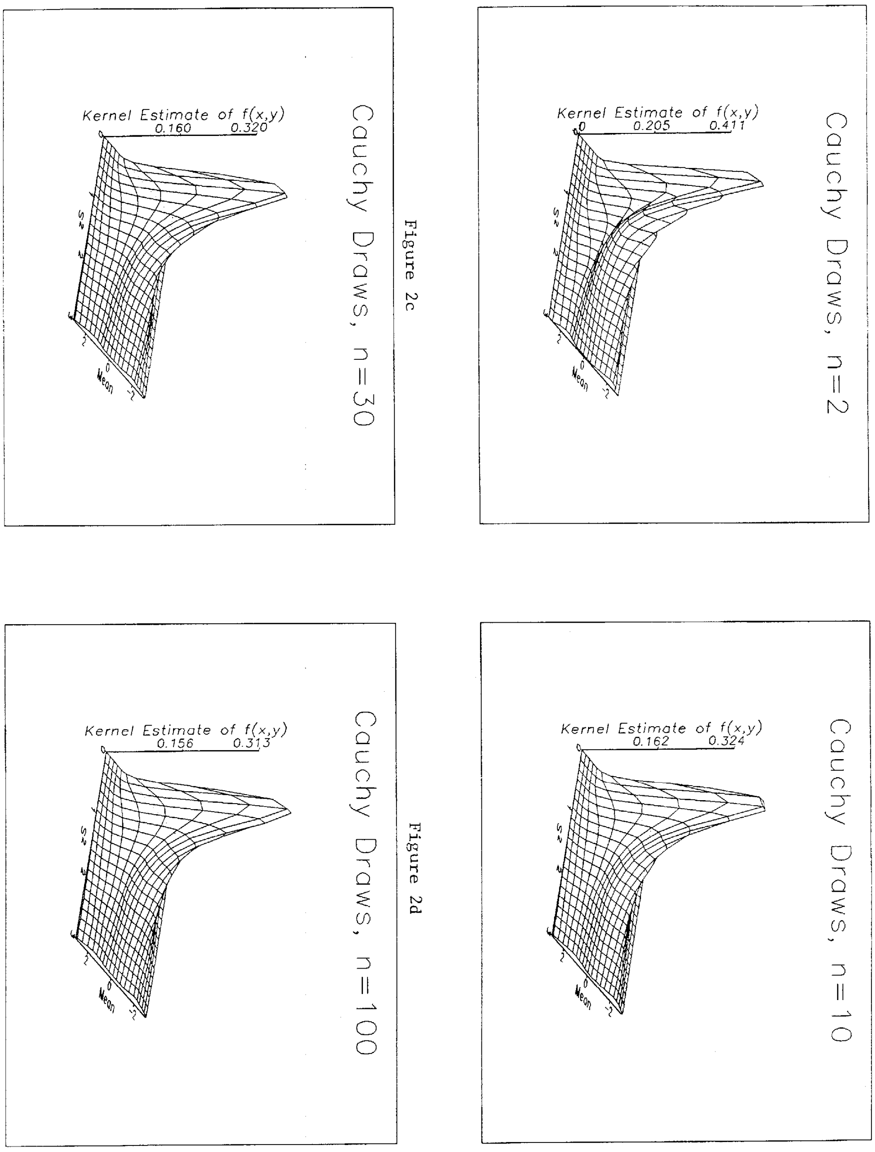


Figure 2: Joint Density Function Estimates of $\bar{X}$ and S^2 for the iid Cauchy DGPs

pictures the density involves a long curving ridge that follows roughly a parabolic shape

$$Y = a + bX^2, a \geq 0, b > 0$$

in the $(\overline{X}, S^2)$ plane. Simple estimates by ordinary least squares of such quadratic relations are presented in Figure 3 for seven values of n. The point estimates obtained are given in Table 1, part A. The prominent ridge in the joint density of tween the two statistics $(\overline{X}, S^2)$ is a manifestation of the dependence between the two statistics $\overline{X}$ and S^2. As is also apparent from these figures the joint distribution of $(\overline{X}, S^2)$ seems to be well approximated by the asymptotic distribution of (X, Y). Indeed the probability surfaces (X, S^2) appear to stabilize quite rapidly (from $n \geq 10$).

TABLE 1

OLS estimates of the Ridge in $pdf(\overline{X}, S^2)$

Estimated Relation: $\hat{y} = \hat{a} + \hat{b}x^2$

A'. Cauchy Draws

Value of n	$\hat{a}$ estimate	$\hat{b}$ estimate
2	.169	.570
5	.288	.451
10	.330	.411
30	.364	.390
50	.364	.398
100	.376	.382
200	.375	.376

B'. Stable Density Draws (with exponent parameter α and $n = 10$)

Value of α	$\hat{a}$ estimate	$\hat{b}$ estimate
1/3	.484	.485
2/3	.475	.420
1	.329	.406
4/3	.252	.322
5/3	.218	.171

Note that the ridge in the joint density surfaces of Figures 2a-d is symmetric about the Y axis. The ridge is associated with clusters of probability mass for various values of Y on either side of the Y axis and equidistant from it. These clusters of mass along the ridge produce a clear bimodality in the conditional distribution of X given S^2 for all moderate to large S^2 . For small S^2 the probability mass is concentrated in the vicinity of the origin in view of the dependence between $\overline{X}$ and S^2. The clusters of probability mass along the ridge in the (X, Y) plane are also responsible for the bimodality in the distribution of certain ratios of the statistics $(\overline{X}, S^2)$ such as the t ratio statistics $t = \overline{X}/S$ and $t_X = \overline{X}/S_X$. These distributions are investigated by simulation in the following section.

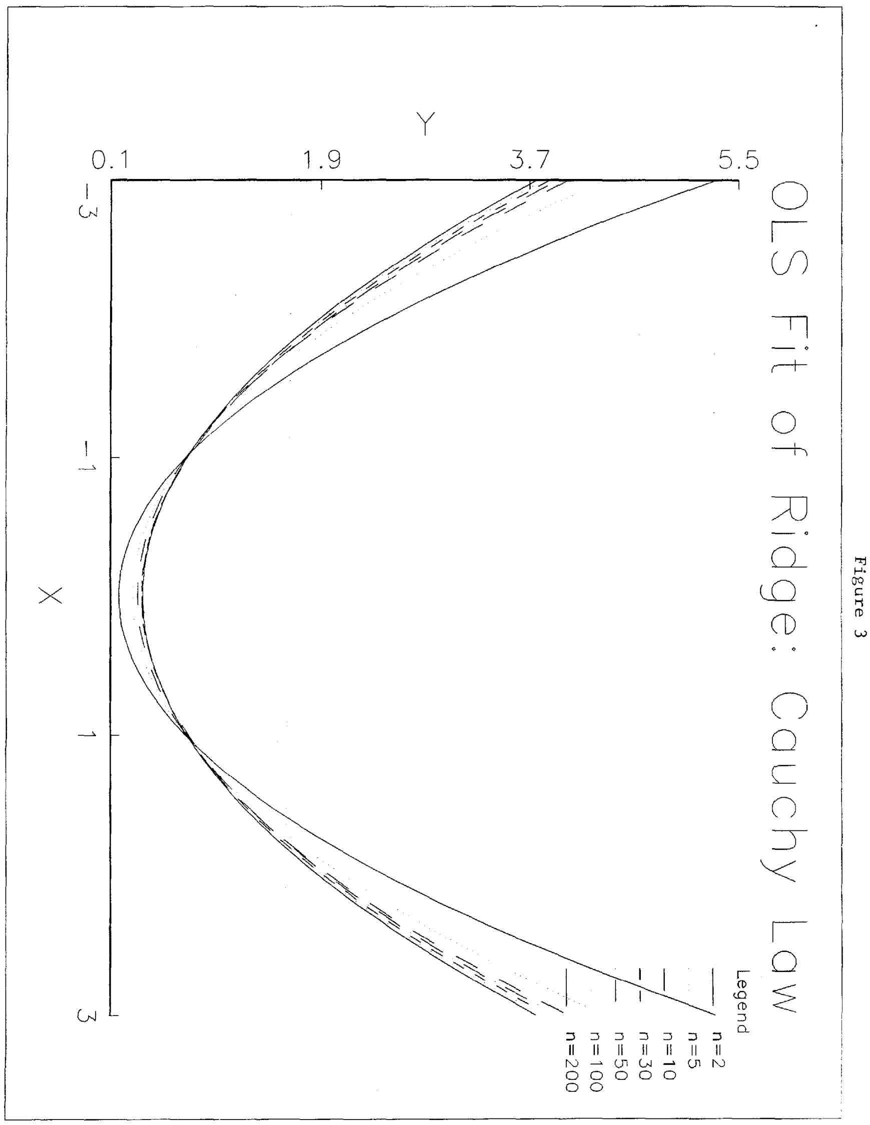


Figure 3: OLS Fit of Ridge in $(\bar{X}, S^2)$ distribution for iid Cauchy DGPs

3 Simulation Evidence

3.1 The Cauchy Case

The empirical distributions reported here were obtained as follows: For a given value of n, $m = 10,000$ random samples of size n were drawn from the standard Cauchy distribution with density given by (2) and corresponding cumulative distribution function

$$F(x) = \frac{1}{\pi} arctan(x), -\infty < x < \infty. \tag{14}$$

Since (14) has a closed form inverse, the probability integral transform method was used in generating the draws. To estimate the probability density functions, the kernel method was employed (see (Tapia and Thompson, 1978)). For the univariate distributions (Figures 1 and 4 above and 6 below) the kernel estimate at point x is

$$\hat{f}(x) = \frac{1}{m} \sum_{r=1}^{m} \phi\left(\frac{x - x_r}{h}\right) / h, \qquad m = 10,000,$$

where $\phi(\cdot)$ is the standard $N(0,1)$ density and the window width h was chosen to be equal to 0.2. For the bivariate distributions in Figures 2 and 4 above and 5 below, the estimate at point (x, y) is

$$\hat{f}(x, y) = \frac{1}{m^2} \sum_{r=1}^{m} \sum_{s=1}^{m} \phi\left(\frac{x - x_r}{h_x}\right) \phi\left(\frac{y - y_s}{h_y}\right) / h_x \cdot h_y$$

with $h_x = h_y = 0.2$.

We now investigate the sampling behavior of the t-ratio statistics t and t_X. These are shown in Figures 4a and 4b. Note that the bimodality is quite striking and persists for all sample sizes.

3.2 Extensions to the Stable Family

Our attention has concentrated on the sampling and asymptotic behavior of statistics based on a random sample from an underlying Cauchy (0,1) population. This has helped to achieve a sharp contrast between our results and those that are known to apply with Cauchy (0,1) populations under the special type of dependence implied by spherical symmetry. However, many of qualitative results given here, such as the bimodality of the t ratios, continue to apply for a much wider class of underlying populations. In particular, if $(X_1, \cdots, X_n)$ is a random sample from a symmetric stable population with characteristic function

$$cf(s) = e^{-|s|^\alpha} \tag{15}$$

and exponent parameter $\alpha < 2$ then the t-ratios t and t_X have bimodal distributions similar in form to those shown in Figures 4a and 4b above for the special case $\alpha = 1$. To generate random variates characterized by (15) a procedure described in Section 1 of (Kanter and Steiger, 1974) was used. We show some examples of

Figure 4a

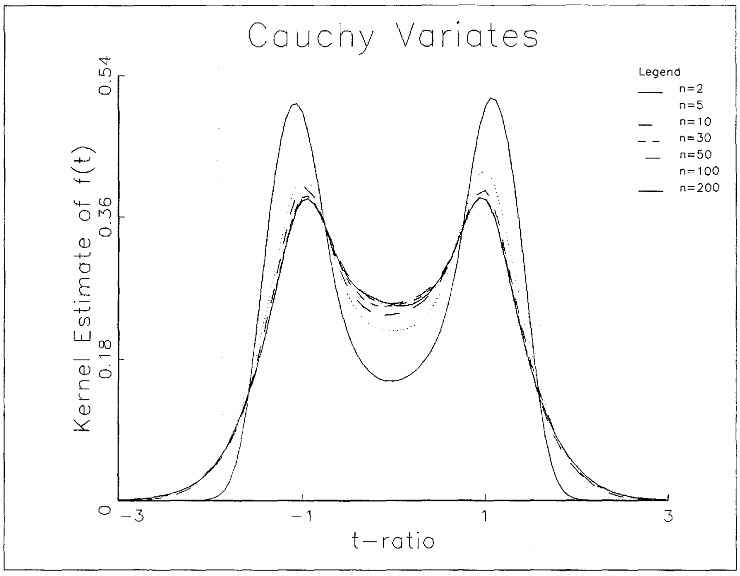


Figure 4b

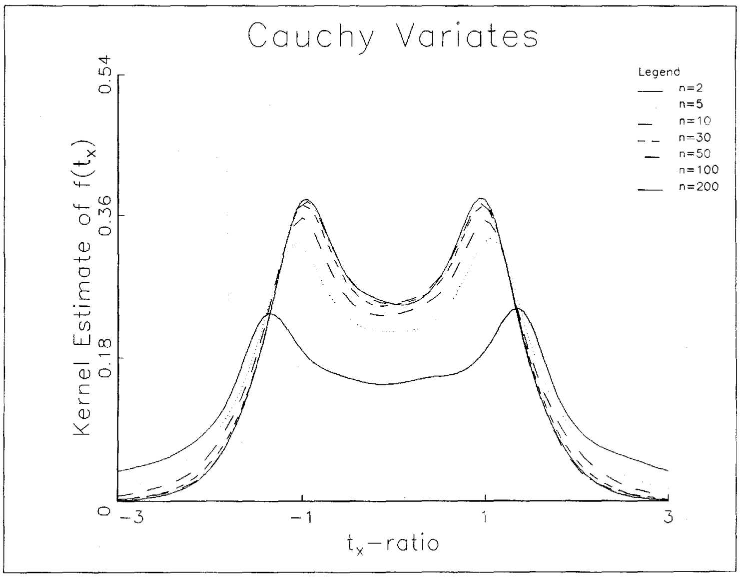


Figure 4: Density Functions of the t-ratio for iid Cauchy DGPs

the new distributions which apply in these cases for various values of a in Figures 5 through 7. Note how in Figures 5a and 5b the bimodality is accentuated for $\alpha < 1$ and attenuated as $\alpha \to 2$. When $\alpha = 2$, of course, the distribution is classical t with $n-1$ degrees of freedom. These effects are also evident from Figures 6a-d and 7, which show that the ridge in the joint distribution is most pronounced for $\alpha = 1/3$ but withers as α rises to 5/3. See also Table 1, part B.

4 Lack of Correlation versus Independence

Data from an n dimensional spherical population with finite second moments have zero correlation, but are independent only when normally distributed. The standard multivariate Cauchy (with density given by (3)) has no finite integer moments but its spherical characteristic may be interpreted as the natural analogue of uncorrelated components in multivariate families with thicker tails. When there is only "lack of correlation" as in the spherical Cauchy case, it is well known (e.g., (King, 1980)) that the distribution of inferential statistics such as the t-ratio reproduce the behavior that they have under independent normal draws. When there are independent draws from a Cauchy population, the statistical behavior of the t-ratio is very different. It no longer mimics behavior under a normal population but has characteristics, such as a random denominator in the limit, which distinguish its distribution from the classical t-ratio and induce the bimodality studied in this article. Examples of this type highlight the statistical implications of the differences between lack of correlation and independence in nonnormal populations.

We include three figures, (8)-(10), to highlight these differences for the bivariate Cauchy case. Figure 8 plots the iso-pdf contours of the bivariate *spherical* Cauchy (implying that the two observations are non-linearly *dependent*), while Figure 9 gives the contours for the bivariate *independent* Cauchy case (implying that the distribution is *non-spherical*). In view of the thick tails, we see the striking divergence between sphericality and statistical independence: whereas for normal Gaussian distribution, sphericality (=uncorrelatedness) and full statistical independence coincide, we now see that for non-Gaussianity, sphericality is neither necessary nor sufficient for independence.

Figure 10 considers 6 representative squares on the domain of the bivariate Cauchy distributions, and for each calculates various measures of deviation from independence for the spherical, dependent version. The last column in the tabulated values in Figure 10 shows that the measure $P_{dependent}$(in given square)$/P_{independent}$(in given square) $- 1 = P_X|Y/P_X - 1$ can get close to 0 (for square E) and sometimes exceeed 50% (for square C).

Hajivassiliou (2005) emphasized that when data are generated from distributions with thick tails independence and zero correlation are very different properties and can have startlingly different outcomes. By construction, the random variables in the numerator of the t-ratio, $\bar{X}$, is linearly orthogonal to the S_X^2 variable in the square root of the denominator. Under Gaussianity, this orthogonality implies full statistical independence between numerator and denominator. But in the case of data drawn from the Cauchy distribution, statistical independence 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,

Figure 5a

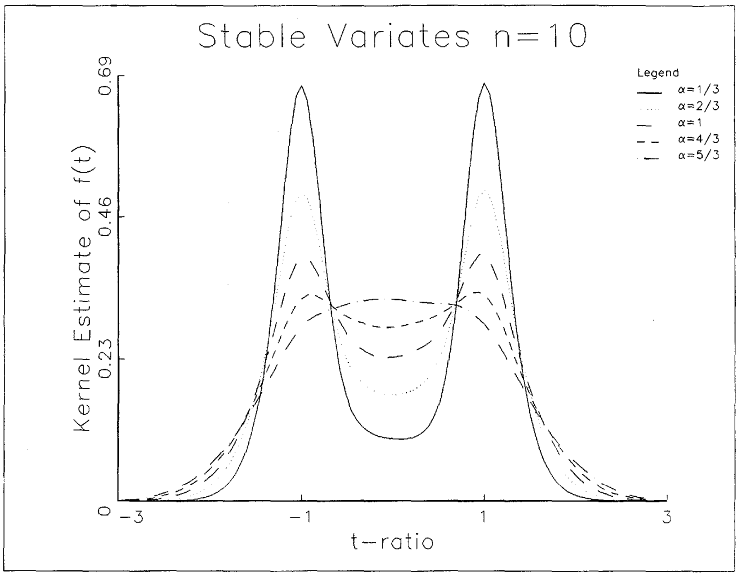


Figure 5b

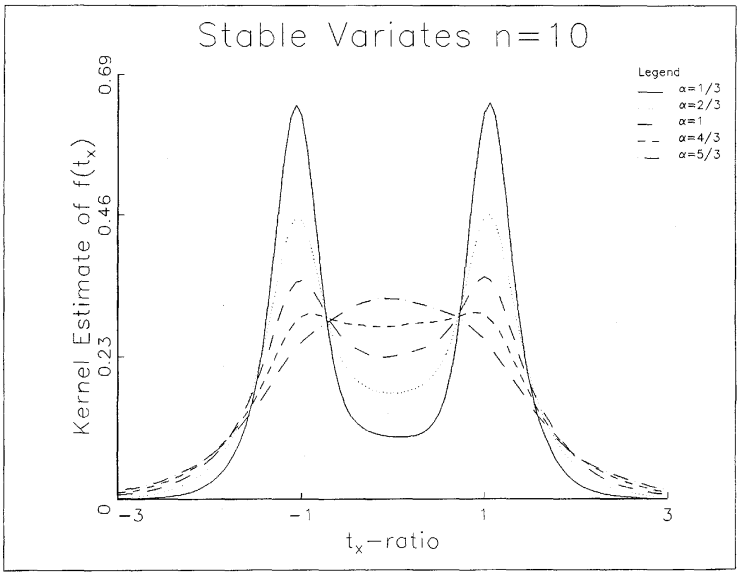


Figure 5: Density Functions of the t-ratio for iid Stable DGPs

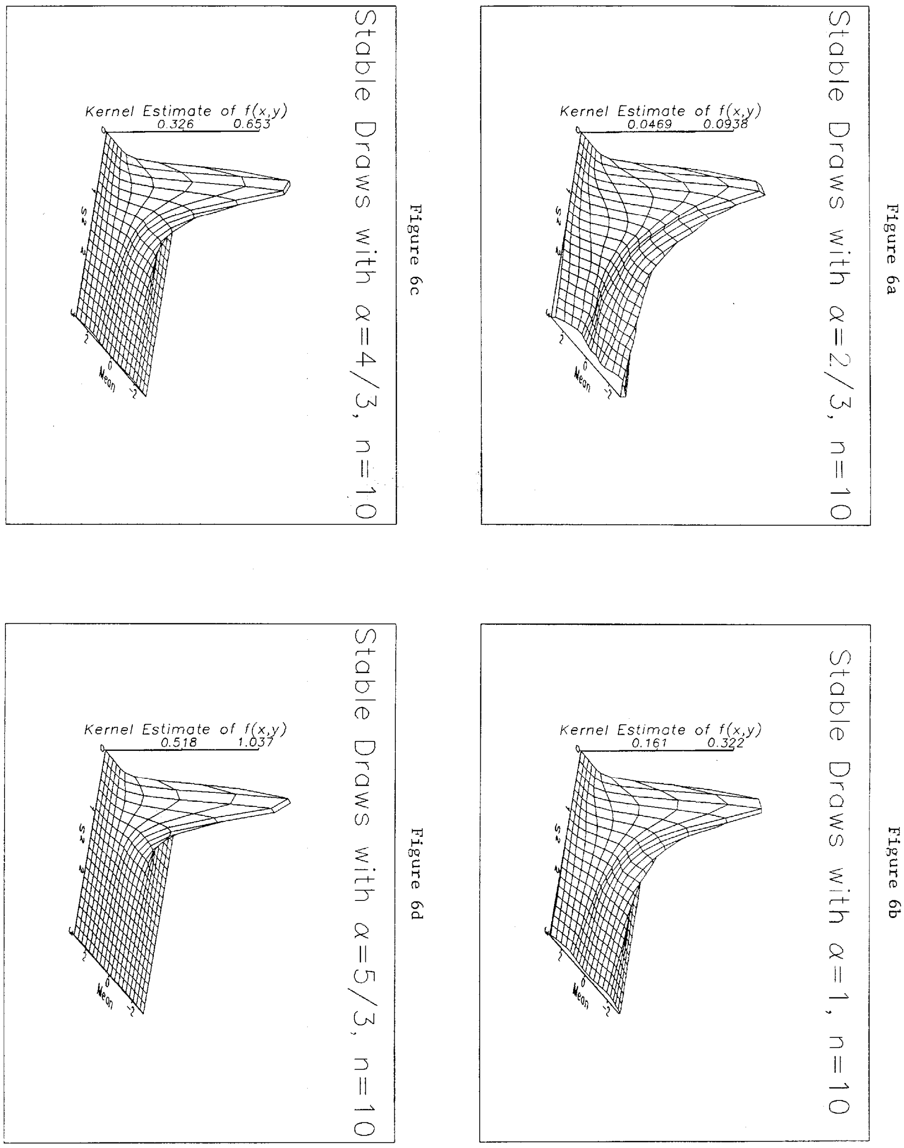


Figure 6: Joint Density Function Estimates of $\bar{X}$ and S^2 for iid Stable DGPs

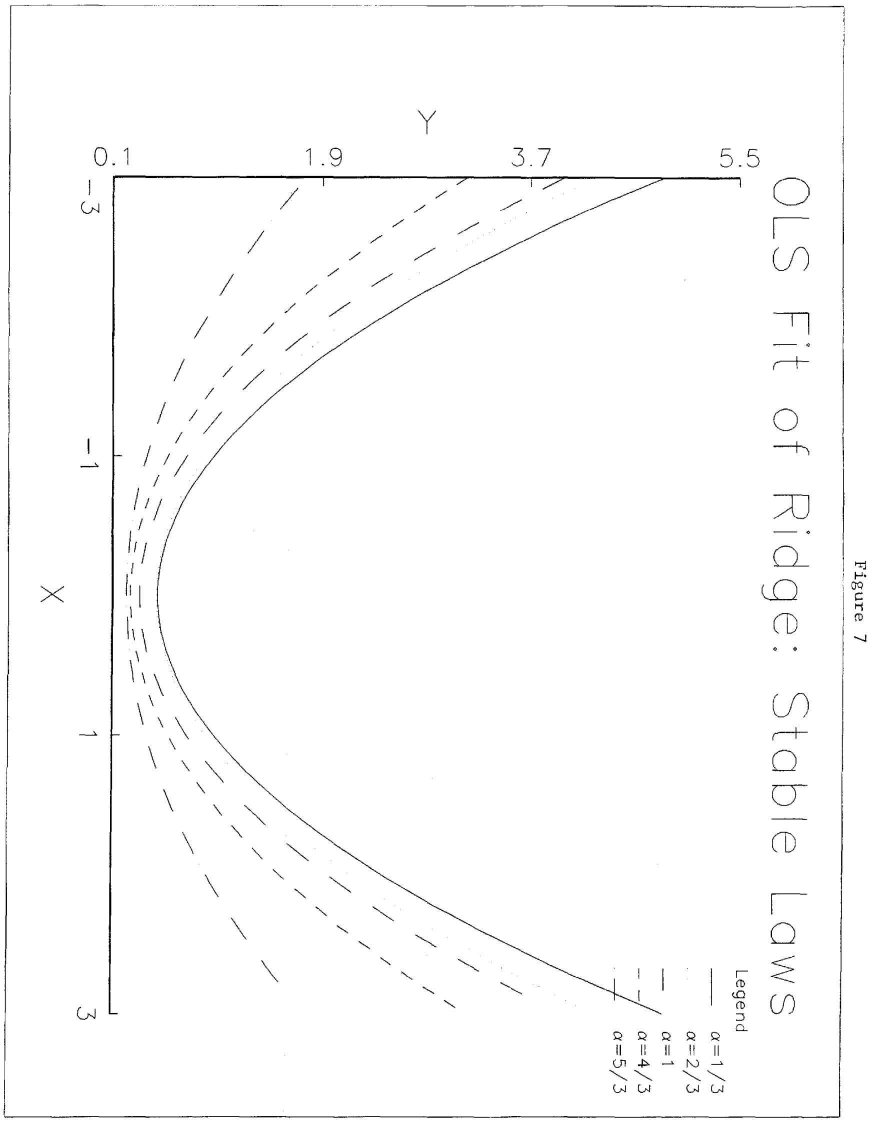


Figure 7: OLS Fit of Ridge in $(\bar{X}, S^2)$ distribution for iid Stable DGPs

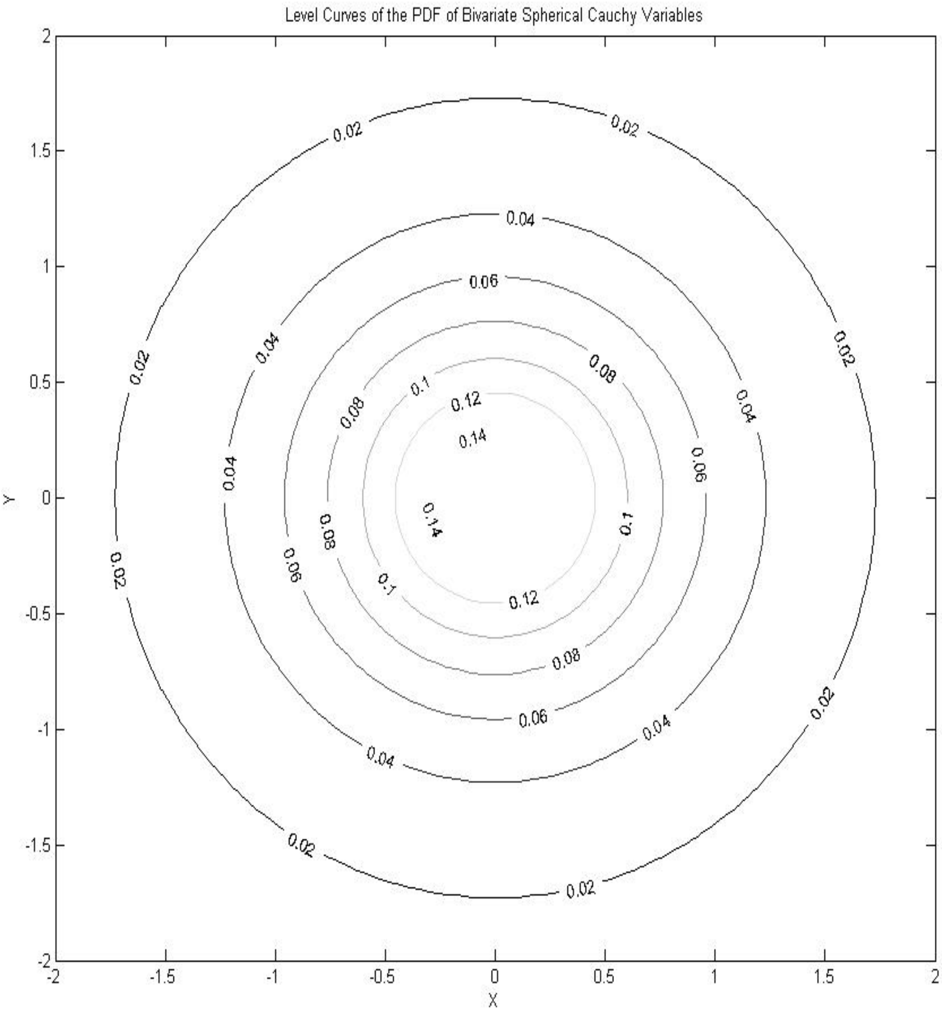


Figure 8: Bivariate Spherical Cauchy (Dependent)

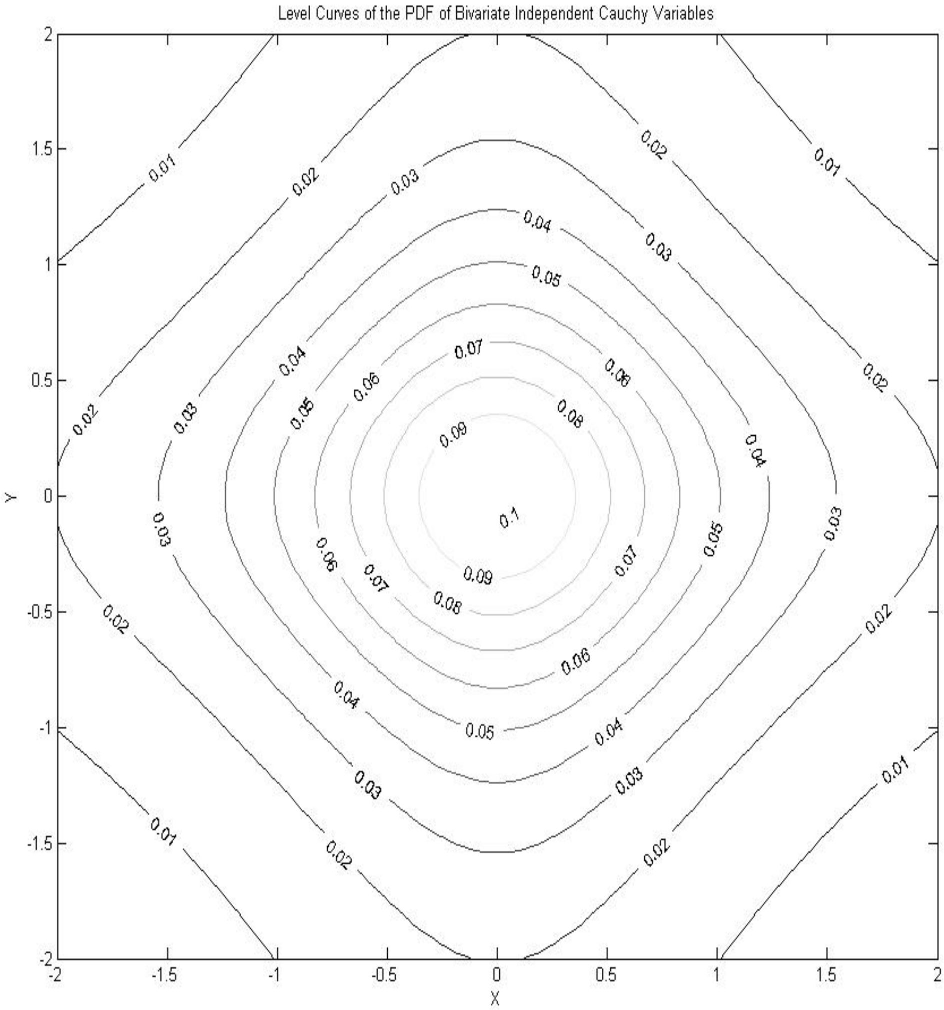
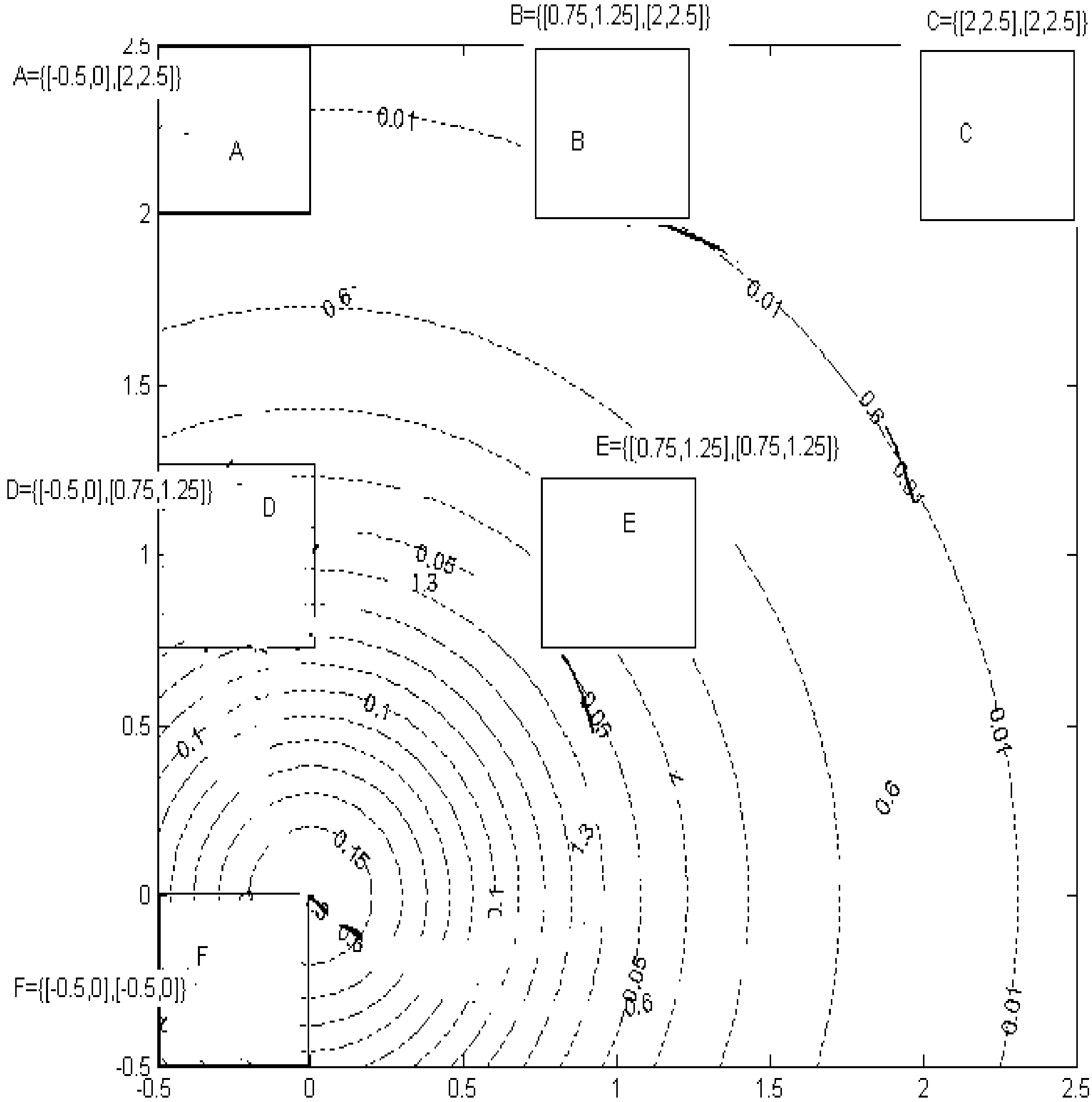


Figure 9: Bivariate Independent Cauchy (Nonspherical)



Square	Pdep	Pindep	Px	Px|y	Py	Py|x	Pdep/Pindep-1=Px|y/Px-1=Py|x/Py-1
A	0.003	0.004	0.148	0.100	0.027	0.018	-32.05%
B	0.008	0.006	0.08	0.097	0.080	0.097	20.15%
C	0.001	0.0007	0.026	0.041	0.027	0.041	54.21%
D	0.014	0.012	0.148	0.168	0.080	0.092	14.13%
E	0.002	0.002	0.08	0.081	0.027	0.027	0.66%
F	0.032	0.022	0.148	0.217	0.148	0.217	47.13%

Figure 10: Measures of Dependency of the Spherical vs. Independent Bivariate Cauchy

then the numerator and denominator in fact become *independent* and the usual unimodal t-distribution obtains (this result is given by Zellner, 1976, but apparently may be traced back to original geometric proofs by Fisher). If, on the other hand, they are drawn fully independently from one another, then $\bar{X}$ and S_X^2 turn out to be *dependent* and hence the distribution of the t-ratio exhibits the striking bimodality documented here.

5 Is the Cauchy DGP Necessary for Bimodality?

Analogous to the double-exponential (see, Feller, 1971, p. 49), we define the double Pareto distribution as the convolution of two independent Pareto (type I) distributed random variables, $X_1 - X_2$, where X_1 and X_2 have density $\alpha_1\beta_1^{\alpha_1}x^{-\alpha_1-1}$ ($x \geq \beta_1, \alpha_1 > 0, \beta_1 > 0$) and $\alpha_2\beta_2^{\alpha_2}(x)^{-\alpha_2-1}$ ($x \geq \beta_2, \alpha_2 > 0, \beta_2 > 0$), respectively. [4] Its density is

$$\int_{-\infty}^{\infty}(\alpha_1\beta_1^{\alpha_1})(\alpha_2\beta_2^{\alpha_2})(x_2+t)^{-\alpha_1-1}(x_2)^{-\alpha_2-1}dx_2$$

and first two moments are (see Section 11):

$$\begin{aligned}
E(x) &= \frac{\alpha_1\beta_1(\alpha_2-1)-\alpha_2\beta_2(\alpha_1-1)}{(\alpha_1-1)(\alpha_2-1)} \quad \text{with } \alpha_1>1, \alpha_2>1 \\
V(x) &= \frac{\alpha_1\beta_1^2}{\alpha_1-2}-\frac{2\alpha_1\alpha_2\beta_1\beta_2}{(\alpha_1-1)(\alpha_2-1)}+\frac{\alpha_2\beta_2^2}{\alpha_2-2} \quad \text{with } \alpha_1>2, \alpha_2>2
\end{aligned}$$

The results that follow were obtained via Monte Carlo simulations from random samples of dimension n using the method of inverted CDFs, 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 replications m was 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 $\alpha_1 = \alpha_2 \leq 1$ are both symmetric and have infinite mean. For these distributions, as the 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 ± 1. As for the symmetric double Pareto, the t-ratio distribution does depend on $\alpha_i, i = 1, 2$: the lower is α_i, the higher is the concentration around the two modes (Figure 11).

Figure 12 examines the case $1 < \alpha < 2$. We now see that the t-ratio, t_2, is not always clearly bimodally distributed. The more α 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 12). We set $\beta = 3$ but these results apply for

[4] The name double Pareto was also used by Reed and Jorgensen (2003) for the distribution of a random variable that is obtained as the ratio of two Pareto random variables and is only defined over a positive support.

[5] Using copulas, we could evaluate behaviour with *correlated* double Pareto draws. See (Hajivassiliou, 2005) for a development of this idea. See also (Ibragimov et al., 2003) for some general results.

any value of $\beta > 0$, since β is simply a threshold parameter that does not affect the t_1 statistic behavior.

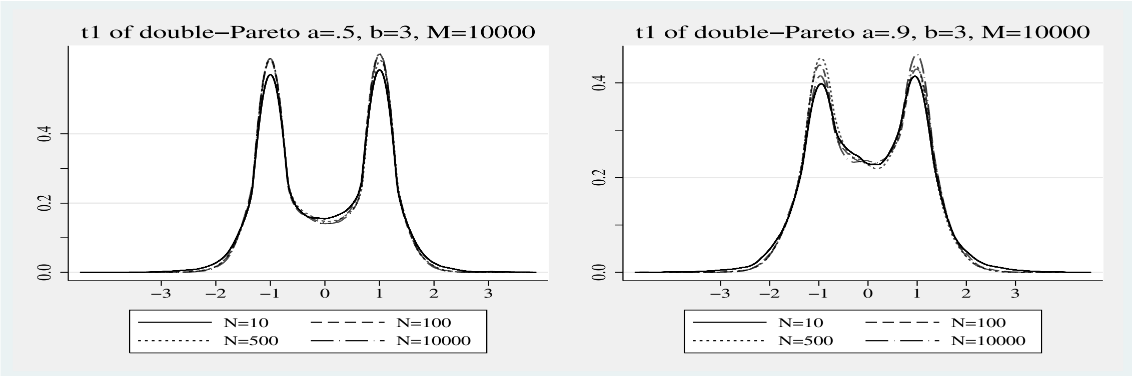


Figure 11: t-ratio of infinite-first-moment double Pareto distributions

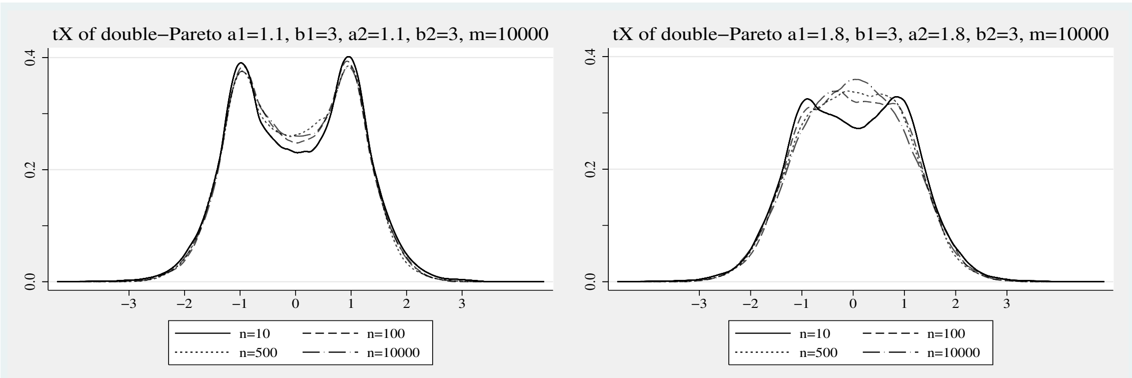


Figure 12: t-ratio of double Pareto distributions with $1 < \alpha \leq 2$

If $\alpha_1 \neq \alpha_2$ it suffices to have either $\alpha_1 \leq 1$ or $\alpha_2 \leq 1$ for the double Pareto to have infinite mean. However, in this case the t-ratio distribution is not bimodal nor stable (Figure 13).

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. In Section 2 above, we showed that if the distribution is Cauchy, the variance converges toward a unimodal distribution with the mode lying in the interval $(0, 1)$. However, if the distribution is double Pareto, the sample variance does not converge towards a stable distribution but becomes more dispersed as the sample size increases (Figure 14). This behaviour confirms the surprising results obtained elsewhere (Ibragimov (2004), Hajivassiliou (2005)) concerning inference with thick-tailed (TT) distributions depending on the tail thickness parameter, α: for $\alpha = 1$, the dispersion of the distribution of sample averages remains invariant to the sample size n, for $\alpha < 1$ 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 if $\alpha = 1$, while portfolio

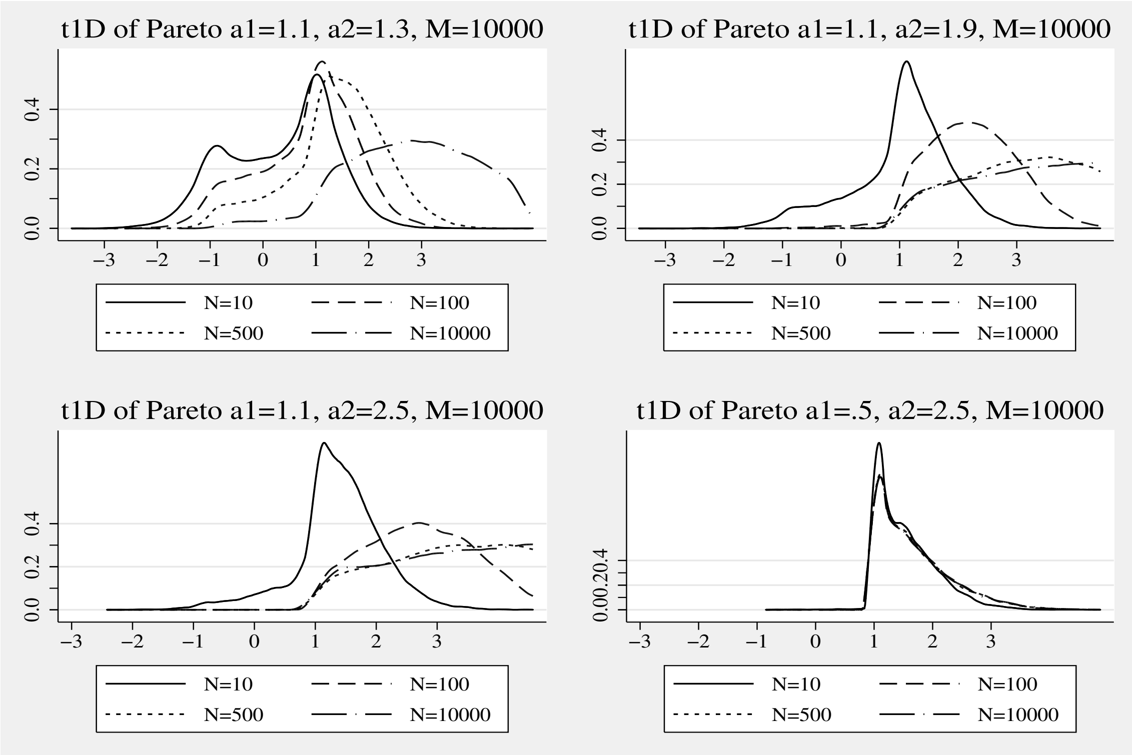


Figure 13: The t-ratio of non-symmetric double Pareto distributions

variability actually *rises* with the number of assets if $\alpha < 1$. 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_X^2) when the first moment is infinite. All the distributions with infinite mean show a clear convex parabolic relationship between the S_X^2 and $\overline{X}$ (Figure 15).

A simple OLS estimate for the coefficient b of the parabolic relationship ($S_X^2 = a + b\overline{X}^2$) is always very close to 1 and highly significant using the Cauchy, 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[6]. 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 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[7]. The latter property implies that the deviation from the

[6]In the Cauchy section above, we found a b coefficient between .570 and .376 for the Cauchy distribution and different sample sizes. However, in those regressions the dependent variable was the uncentered second moment while the centered one is considered here.

[7]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, then it follows that

$$\lim_{x\to\infty} \frac{P(Z_n > x)}{P(M_n > x)} = 1 \tag{16}$$

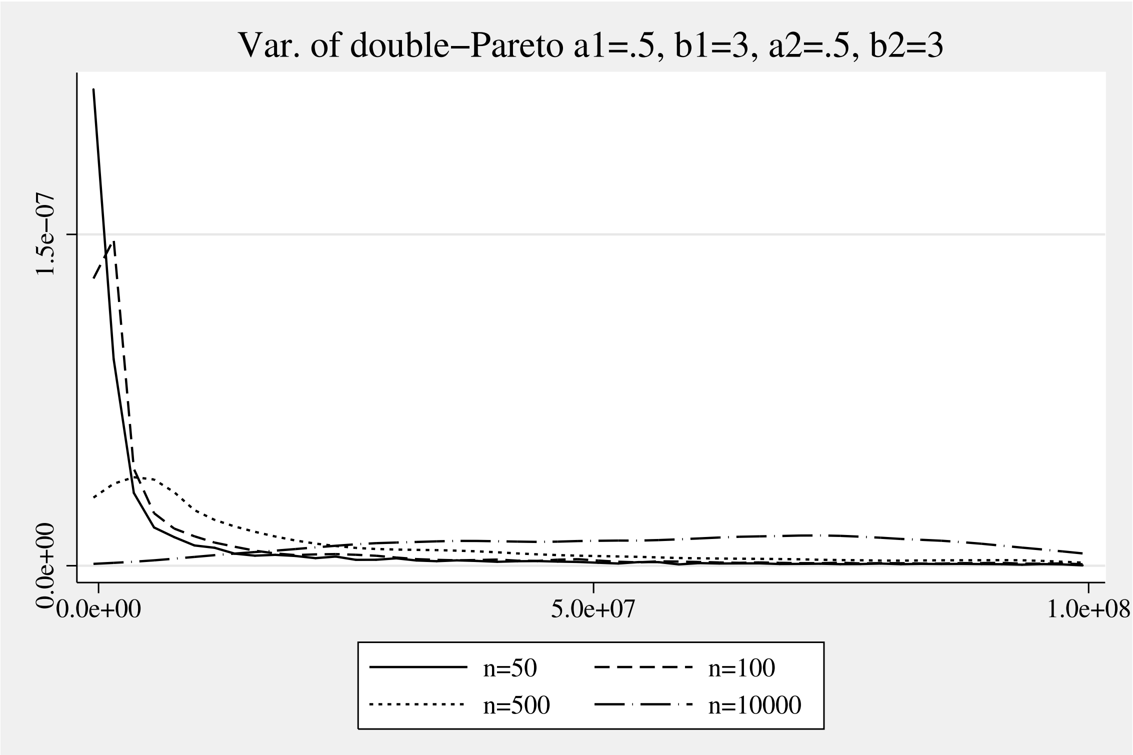


Figure 14: Distribution of the variance of double Pareto distributions with infinite mean.

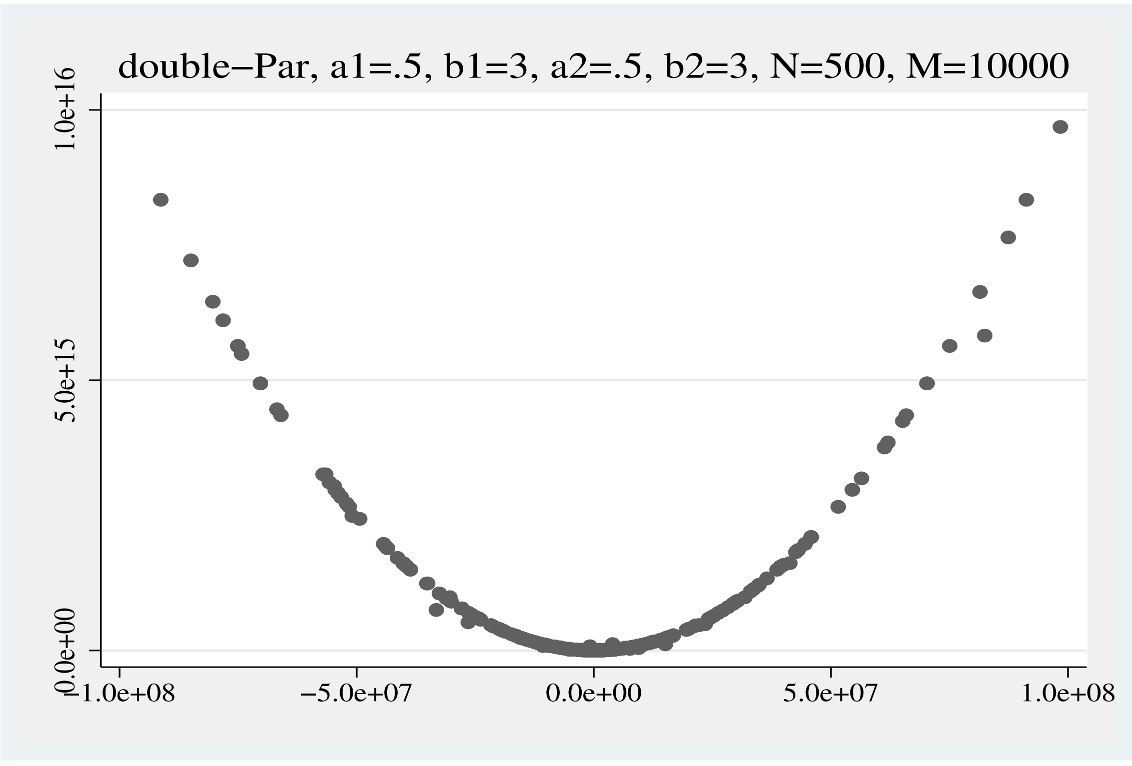


Figure 15: Relationship between the mean and the variance of a double Pareto with infinite mean.

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.

6 How Misleading is the Use of Standard t-ratios with TT distributions?

The preceding results are relevant for hypothesis testing in regressions with errors that are independent and identically distributed from a TT distribution. They are also relevant 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 critical values of a $N(0,1)$ distribution are used in classical t-ratio testing. The issue is well 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 τ, generated via Monte Carlo simulation using a data-generating process (DGP) under 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 τ_j is the probability of observing a value of τ more extreme than τ_j, according to some distribution $F(\tau)$. This distribution could be the asymptotic distribution of τ, derived numerically or theoretically, as well as other distributions such as an approximation derived by bootstrapping. The $p-$value is a function of τ_j, $p_j \equiv p(\tau_j)$. Assuming τ is asymptotically distributed as a standard normal with DF $\phi(z)$ and CDF $\Phi(z)$, then $p_j = 1 - \Phi(\tau_j)$[8].

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

$$\widehat{F}(x_i) \equiv \frac{1}{m} \sum_{j=1}^{m} I(p_j \leq x_i) \tag{17}$$

where $I(p_j \leq x_i)$ is the indicator function. Although the function (17) 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 $\widehat{F}(x_i)$ 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 $\widehat{F}(x_i) - x_i$ against x_i.

6.1 Inference on Location from a Single Sample

The $p-$value discrepancy plot of the t-ratio statistic for the double Pareto with different values of α was constructed as in (17), where the $p-$value is derived both

[8] For a two-sided test, the $p-$value is $p_j \equiv p(|\tau_j|) = 2(1 - \Phi(\tau_j))$.

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.

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 μ is the true mean and μ_0 some value on the real line. The sample mean is used to estimate μ. 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 (ERP) using the standard normal for testing two-sided hypothesis with a symmetric double Pareto distribution with $1 < \alpha \leq 2$. However, three 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 acceptable 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[9]. 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).

Secondly, the "ignore" policy leads to major errors in the case of the symmetric double Pareto distribution. The ERP for a two sided test about the mean this distribution can be as large as 10%. This result clearly comes from the non standard distribution of the t-ratio and is even more dramatic in the single Pareto results of (Fiorio and Hajivassiliou, 2006). Finally, if the non-symmetric double Pareto distribution is considered, then the t-ratio statistic is not even stable.

7 Are Infinite Moments Necessary for Bimodality?

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:

$$Z = \begin{cases} X & iff \quad |X| < c \\ NA & otherwise \end{cases}$$

[9]Although tests with nominal 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%.

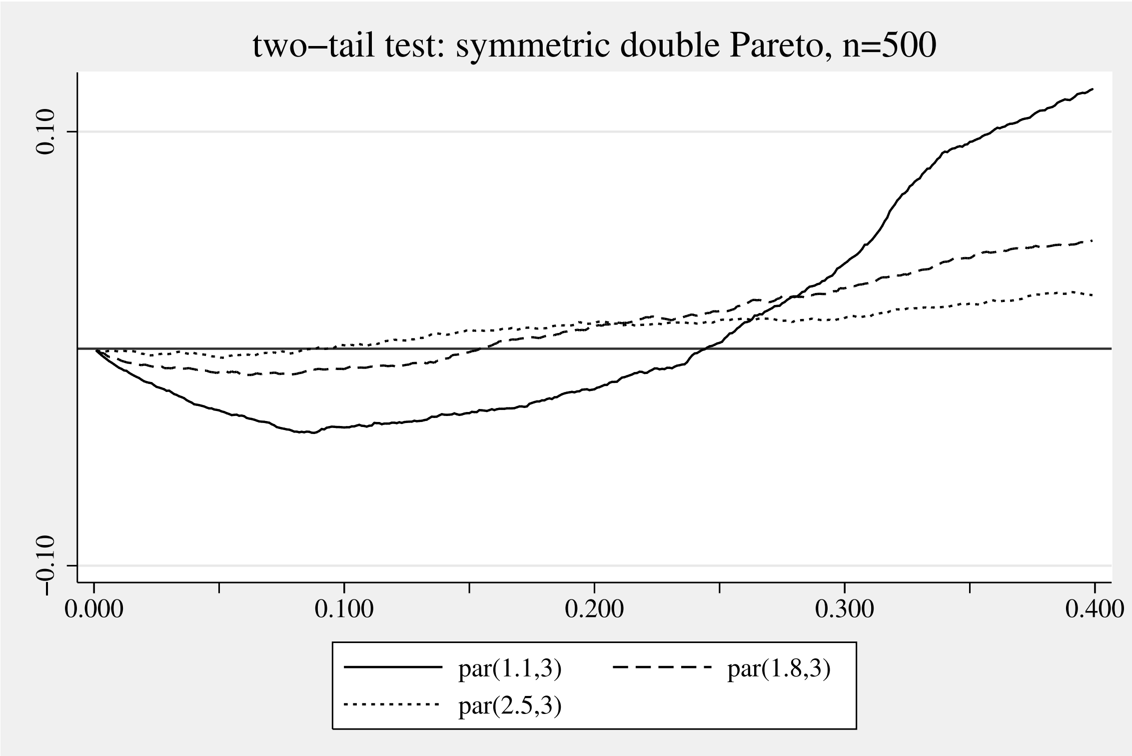


Figure 16: ERP for two-tail test with a sample from a double Pareto with $\alpha = 1.1, 1.8, 2.5$.

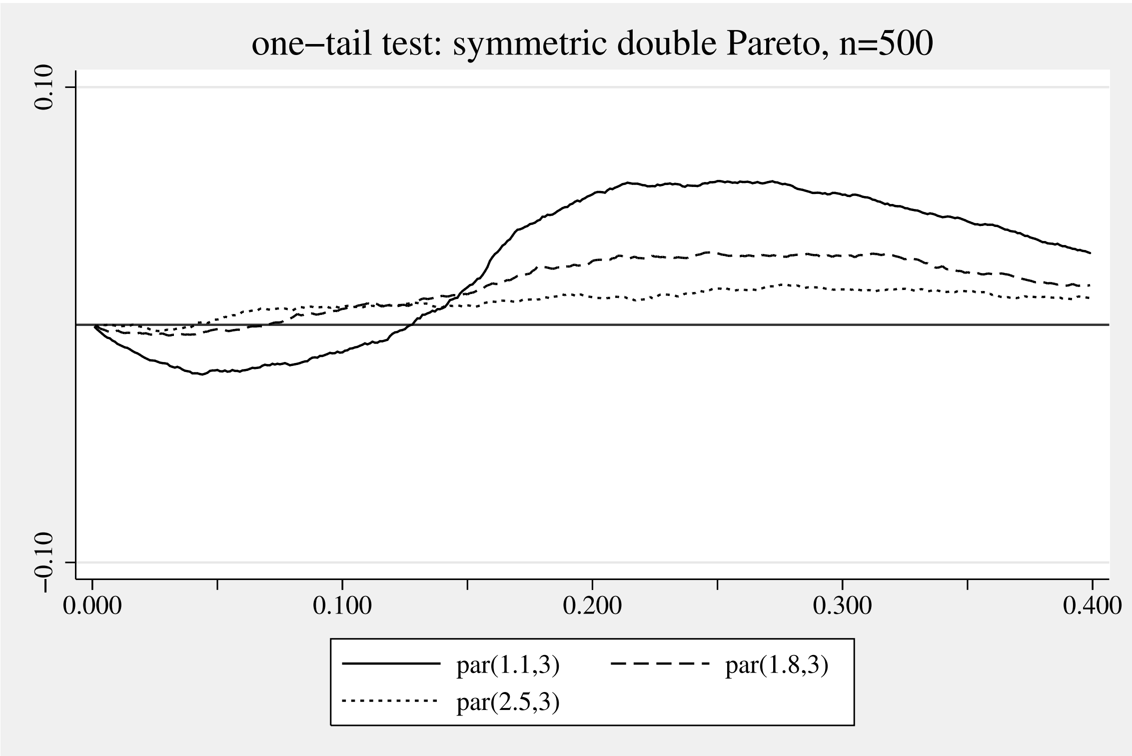


Figure 17: ERP for one-tail test with a sample from a double Pareto with $\alpha = 1.1, 1.8, 2.5$.

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

$$f(z) = \frac{\frac{1}{\pi(1+z^2)}}{\frac{2}{\pi}arctan(c)}, \qquad |z| < c. \tag{18}$$

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 the r.v. Z are all finite.

The second trimmed distribution we consider is the double Pareto truncated on a compact support as in (18) where X is a r.v. distributed according to the double Pareto law introduced in Section 5. 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 α.

By considering truncated versions of distributions whose untruncated counterparts do not have finite moments, we can control the relative importance of the tails while working with distributions with all moments finite. In the simulations below, we consider the following truncation points:

Truncated Cauchy				
c	500	1,000	3,000	5,000
prob(cutoff tails)	0.0012	0.0006	0.0002	0.0001

Truncated Double Pareto				
c	5,000	100,000	250,000	500,000
prob(cutoff tails), $\alpha = 0.5$	0.049	0.011	0.0069	0.0048

The higher the absolute value of c is, the less attenuated the impact of tail behaviour will be. In contrast, low absolute values of c imply cutting out most of the (thick) tails of the distribution.

The general conclusion is that the bimodality can appear also when moments are finite and the sample size is finite, but reasonably large for many empirical applications. Our results with $N = 500$ show that the source of the bimodality is the rate of tail behaviour and *not* unboundedness of support or non-existence of moments (Figure 18), the non-normal behavior being more evident the larger the truncation point c.

The heuristic explanation for these results is that any large draw in a finite sample from the underlying TT distribution will tend to dominate both the numerator and denominator of a t ratio statistic, even if the DGP distribution has bounded support. Especially when there is a single extremely large draw that dominates all others, then the t will be approximately $\pm$, therefore leading to a distribution that has modal activity in the neighbourhood of these two points. Clearly, it is not necessary for the distribution to have infinite moments or unbounded support for this phenomenon to occur.

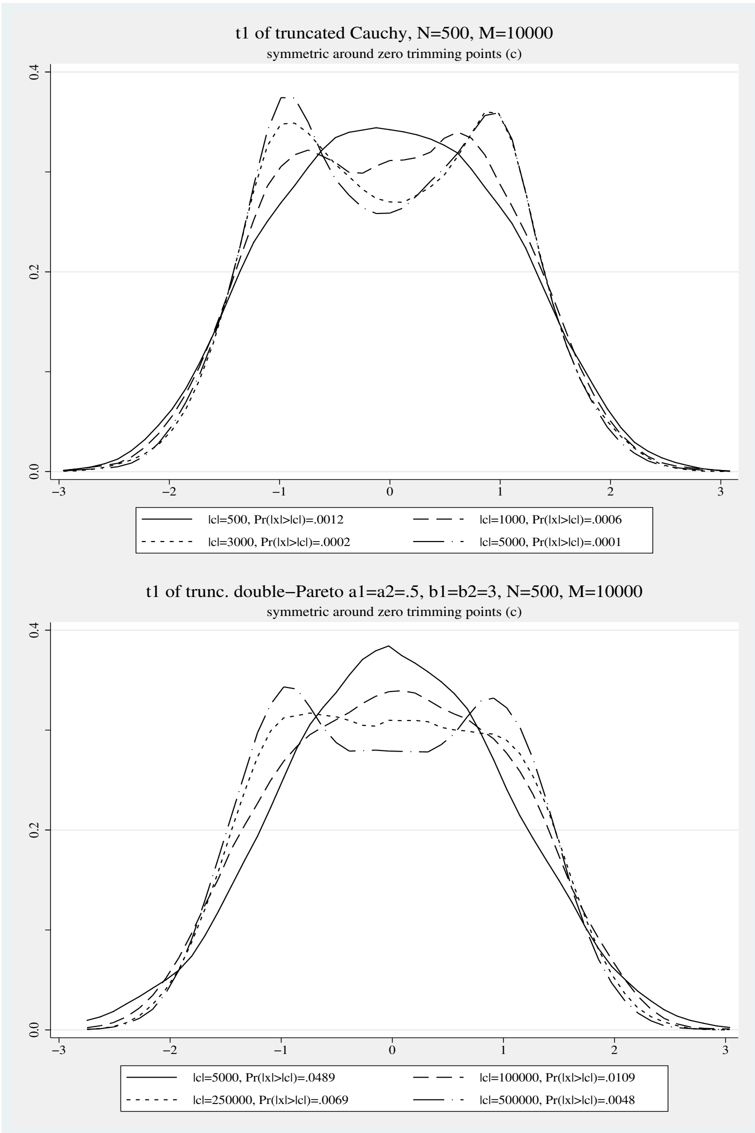


Figure 18: t-ratio of Cauchy and double Pareto on compact support.

8 Other Complications with TT Distributions

Obviously, the non standard distribution of the t-ratio with infinite second moment also affects two-sample tests 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 (μ_A) is different from the mean of the second (μ_B) using the t-ratio, under $H_0 = \mu_A - \mu_B = 0$:

$$t_2^D = \frac{\overline{X}_A - \overline{X}_B}{\sqrt{(S_{\overline{X}}^A/n^A) + (S_{\overline{X}}^B/n^B)}} \tag{19}$$

where $\overline{X}_A, \overline{X}_B$ are the sample means, and $S_{\overline{X}}^A, S_{\overline{X}}^B$ are the sample variance of A and B, respectively. The distribution of t_2^D is again non-standard. Moreover, in many cases it does converge to a stable distribution as the sample size increases. Figure 13 shows via Monte Carlo simulations the distributions of the t-ratio, t_2^D, for testing the difference in means of two Pareto distributions that may differ in α 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, employing the standard normal critical values for the sample t-ratio in this case would give extremely misleading results. These results are driven by the fact that t_2^D is a convolution of two stable distributions with different tail-thickness parameters. In general, convolutions of stable distributions also have a stable distribution only if the stable distributions involved present the same thickedness (rate of decay) of the tails (see for instance Samorodnitsky and Taqqu, 1994). Simulation results are not provided here, but can be obtained from the authors upon request.

9 Conclusions

This paper has investigated issues of inference from data based on independent draws from TT distributions. When the distribution is TT with infinite second moment, the standard t-ratio formed from a random sample does not converge to a standard normal distribution and the limit distribution is generally bimodal. Conventional inference is invalidated in such cases and errors in the rejection probability in testing can be serious. The results provide some cautionary evidence and show the need to adapt estimation and inference procedures to account for the special features induced by heavy tailed data. Bimodality in the finite sample distribution of the t-ratio arises even in cases of trimmed TT distributions, showing that non-existence of moments is not necessary for the phenomenon to occur.

10 Appendix A: Proofs of Lemmas of Section (2)

Proof of Lemma 1

$$S_X^2 = S^2 - n^{-1}\overline{X}^2 = S^2 + O_p(n^{-1})$$

since $\overline{X} \Rightarrow$ Cauchy (0,1). Similarly,

$$t_X = \overline{X}\left[S^2 + O_p(n^{-1})\right]^{-1/2} = t + Op(n^{-1})$$

as required.

Proof of Lemma 2. We start by finding the characteristic function of X_i^2. This is

$$\begin{aligned} E\left(e^{ivX_i^2}\right) &= \int_{-\infty}^{\infty} \frac{e^{ivx^2}dx}{\pi(1+x^2)} \\ &= \int_0^{\infty} \frac{e^{ivr}dr}{\pi r^{1/2}(1+r)} \\ &= \left(\Gamma\left(\frac{1}{2}\right)\right)^{-1} \Psi\left(\frac{1}{2},\frac{1}{2};-iv\right) \end{aligned}$$

where Ψ is a confluent hypergeometric function of the second kind. It follows that the characteristic function of $S^2 = n^{-2}\Sigma_1^n X_i^2$ is:

$$\begin{aligned} E\left(e^{ivS^2}\right) &= \Pi_{i=1}^n E\left(e^{ivX_i^2/n^2}\right) \\ &= \left[\left(\Gamma\left(\frac{1}{2}\right)\right)^{-1} \Psi\left(\frac{1}{2},\frac{1}{2};-iv/n^2\right)\right]^n \end{aligned} \tag{20}$$

We now use the following asymptotic expansion of the Ψ function (se Erdélyi, 1953, p. 262)

$$\Psi\left(\frac{1}{2},\frac{1}{2};\frac{-iv}{n^2}\right) = \Gamma\left(\frac{1}{2}\right) + \frac{\Gamma\left(-\frac{1}{2}\right)}{\Gamma\left(\frac{1}{2}\right)}\left(\frac{-iv}{n^2}\right)^{1/2} + o(1/n)$$

so that (20) tends as $n \to \infty$ to:

$$exp\left\{\frac{\Gamma\left(-\frac{1}{2}\right)}{\Gamma\left(\frac{1}{2}\right)^2}(-iv)^{1/2}\right\} = exp\left\{\frac{-2}{\pi^{1/2}}(-iv)^{1/2}\right\}.$$

Using the argument given in the text from equations (13) to (14) we deduce (9) as stated.

Proof of Lemma 3. We take the joint Laplace transform

$$L(z,w) = \int_{-\infty}^{\infty} \frac{e^{zx+wx^2}}{\pi(1+x^2)}dx$$

and transform $x \to (r,h)$ according to the decomposition $x = r^{1/2}h$ where $r = x^2$ and $h = sgn(x) = \pm 1$. Using the Bassel function integral

$$\int_h e^{zrh/2} dh = {}_0F_1\left(\frac{1}{2}, \frac{1}{4}z^2 r\right) = \Sigma_{k=0}^{\infty} \frac{(z^2 4)^k r^k}{k! \left(\frac{1}{2}\right)_k}$$

we obtain

$$\begin{aligned} L(z,w) &= \frac{1}{\pi}\sum_{k=0}^{\infty} \frac{(z^2/4)^k}{k!\left(\frac{1}{2}\right)_k} \int_0^{\infty} \frac{e^{wr} r^{k-1/2}}{(1+r)} dr \\ &= \frac{1}{\pi}\sum_{k=0}^{\infty} \frac{(z^2/4)^k \Gamma\left(k+\frac{1}{2}\right)}{k!\left(\frac{1}{2}\right)_k} \Psi\left(k+\frac{1}{2}, k+\frac{1}{2}, -w\right) \qquad (21) \end{aligned}$$

from the integral representation of the Ψ funciton (Erdélyi, 1953, p. 255). We now use the fact that

$$\begin{aligned} \Psi\left(k+\frac{1}{2}, k+\frac{1}{2}; -w\right) &= \Gamma\left(\frac{1}{2}-k\right) {}_1F_1\left(k+\frac{1}{2}, k+\frac{1}{2}; -w\right) \qquad (22) \\ &+ \frac{\Gamma\left(k-\frac{1}{2}\right)}{\Gamma\left(k+\frac{1}{2}\right)} (-w)^{1/2-k} {}_1F_1\left(1, \frac{3}{2}-k; -w\right) \end{aligned}$$

(see Erdélyi, 1953, p. 257)

$$\Gamma\left(\frac{1}{2}-k\right) = \frac{\pi}{(-1^k)\Gamma\left(k+\frac{1}{2}\right)}$$

and

$${}_1F_1\left(k+\frac{1}{2}, k+\frac{1}{2}; -w\right) = e^{-w}$$

Combining (21) and (22) we have:

$$\begin{aligned} L(z,w) &= \sum_{k=0}^{\infty} \frac{(-z^2/4)^k}{k!\left(\frac{1}{2}\right)_k} e^{-w} \qquad (23) \\ &+ \frac{1}{\pi}\sum_{k=0}^{\infty} \frac{(z^2/4)^k \Gamma\left(k-\frac{1}{2}\right)}{k!\left(\frac{1}{2}\right)_k} (-w)^{1/2-k} {}_1F_1\left(1, \frac{3}{2}-k; -w\right) \end{aligned}$$

Let

$$z = \frac{iu}{T} \quad , \quad w = \frac{iv}{T^2}$$

It follows from (23) that

$$L\left(\frac{iu}{T},\frac{iv}{T^2}\right)=1+\left(\frac{\Gamma\left(-\frac{1}{2}\right)}{\pi}\sum_{k=0}^{\infty}\frac{\left(-\frac{1}{2}\right)_k(u/4iv)^k}{k!\left(\frac{1}{2}\right)_k}\right)\left(\frac{-iv}{T^2}\right)^{1/2}+o\left(\frac{1}{T}\right)$$

and thus

$$\left[L\left(\frac{iu}{T},\frac{iv}{T^2}\right)\right]^T\to exp\left\{\frac{\Gamma\left(-\frac{1}{2}\right)}{\pi}{}_1F_1\left(-\frac{1}{2},\frac{1}{2};\frac{u^2}{4iv}\right)(-iv)^{1/2}\right\}$$

Since

$$cf_{\overline{X},S^2}(u,v)=\left[L\left(\frac{iu}{T},\frac{iv}{T^2}\right)\right]^T$$

and

$$\Gamma\left(-\frac{1}{2}\right)=-2\pi^{1/2}$$

we deduce that

$$cf_{X,Y}(u,v)=exp\left\{-\frac{2}{\pi^{1/2}}{}_1F_1\left(-\frac{1}{2},\frac{1}{2};\frac{u^2}{4iv}\right)(-iv)^{(1/2)}\right\}\tag{24}$$

as required for (11).

The second representation in the Lemma is obtained by noting that

$$a^{-1}x^a{}_1F_1(a,a+1;-x)=\Gamma(a)-e^{-x}\Psi(1-a,1-a,x)$$

(Erdélyi, 1953, p. 266). Using this result we find

$$\left(-\frac{1}{2}\right)^{-1}(-iv)^{1/2}{}_1F_1\left(-\frac{1}{2},\frac{1}{2};\frac{u^2}{4iv}\right)=\frac{1}{2}|u|\left\{\Gamma\left(-\frac{1}{2}\right)-e^{u^2/4iv}\Psi\left(\frac{3}{2},\frac{3}{2};\frac{-u^2}{4iv}\right)\right\}.$$

Using (10) in (10) we obtain (12) as stated.

11 Appendix B: Moments of the symmetric Pareto distribution

The first moment of the double Pareto distribution introduced in Section 5 is:

$$
\begin{aligned}
E(X) &= \int_{-\infty}^{\infty} t \left(\int_{-\infty}^{\infty} (\alpha_1 \beta_1^{\alpha_1})(\alpha_2 \beta_2^{\alpha_2})(x_2+t)^{-\alpha_1-1} x_2^{-\alpha_2-1} dx_2 \right) dt = \\
&= (\alpha_1 \beta_1^{\alpha_1})(\alpha_2 \beta_2^{\alpha_2}) \int_{\beta_2}^{\infty} x_2^{-\alpha_2-1} \left(\int_{\beta_1-x_2}^{\infty} t(x_2+t)^{-\alpha_1-1} dt \right) dx_2 = \qquad x_1 > \beta_1, x_2 > \beta_2 \\
&= (\alpha_1 \beta_1^{\alpha_1})(\alpha_2 \beta_2^{\alpha_2}) \int_{\beta_2}^{\infty} x_2^{-\alpha_2-1} \left(\int_{\beta_1}^{\infty} (z-x_2)(z)^{-\alpha_1-1} dz \right) dx_2 = \\
&= (\alpha_1 \beta_1^{\alpha_1})(\alpha_2 \beta_2^{\alpha_2}) \int_{\beta_2}^{\infty} x_2^{-\alpha_2-1} \left(\int_{\beta_1}^{\infty} z^{-\alpha_1} dz - x_2 \int_{\beta_1}^{\infty} z^{-\alpha_1-1} dz \right) dx_2 = \\
&= (\alpha_1 \beta_1^{\alpha_1})(\alpha_2 \beta_2^{\alpha_2}) \int_{\beta_2}^{\infty} x_2^{-\alpha_2-1} \left(\frac{1}{-\alpha_1+1} z^{-\alpha_1+1} |_{\beta_1}^{\infty} - x_2 \frac{1}{-\alpha_1} z^{-\alpha_1} |_{\beta_1}^{\infty} \right) dx_2 = \qquad \alpha_1 > 1 \\
&= (\alpha_1 \beta_1^{\alpha_1})(\alpha_2 \beta_2^{\alpha_2}) \int_{\beta_2}^{\infty} x_2^{-\alpha_2-1} \left(\frac{\beta_1^{-\alpha_1+1}}{\alpha_1-1} - \frac{x_2 \beta_1^{-\alpha_1}}{\alpha_1} \right) dx_2 = \\
&= (\alpha_1 \beta_1^{\alpha_1})(\alpha_2 \beta_2^{\alpha_2}) \left(\frac{\beta_1^{-\alpha_1+1}}{\alpha_1-1} \frac{1}{-\alpha_2} x_2^{-\alpha_2} |_{\beta_2}^{\infty} - \frac{\beta_1^{-\alpha_1}}{\alpha_1} \frac{1}{-\alpha_2+1} x_2^{-\alpha_2+1} |_{\beta_2}^{\infty} \right) = \qquad \alpha_2 > 1 \\
&= \frac{\alpha_1 \beta_1 (\alpha_2-1) - \alpha_2 \beta_2 (\alpha_1-1)}{(\alpha_1-1)(\alpha_2-1)}
\end{aligned}
$$

The second central moment of the double Pareto distribution introduced in Section 5 is:

$$
\begin{aligned}
V(X) &= (\alpha_1 \beta_1^{\alpha_1})(\alpha_2 \beta_2^{\alpha_2}) \int_{-\infty}^{\infty} t^2 \int_{-\infty}^{\infty} x_2^{-\alpha_2-1} (x_2+t)^{-\alpha_1-1} dx_2 dt = \\
&= (\alpha_1 \beta_1^{\alpha_1})(\alpha_2 \beta_2^{\alpha_2}) \int_{\beta_2}^{\infty} x_2^{-\alpha_2-1} \left(\int_{\beta_1-x_2}^{\infty} t^2 (x_2+t)^{-\alpha_1-1} dt \right) dx_2 = \qquad x_1 > \beta_1, x_2 > \beta_2 \\
&= (\alpha_1 \beta_1^{\alpha_1})(\alpha_2 \beta_2^{\alpha_2}) \int_{\beta_2}^{\infty} x_2^{-\alpha_2-1} \left(\int_{\beta_1}^{\infty} (z-x_2)^2 (z)^{-\alpha_1-1} dz \right) dx_2 = \\
&= (\alpha_1 \beta_1^{\alpha_1})(\alpha_2 \beta_2^{\alpha_2}) \int_{\beta_2}^{\infty} x_2^{-\alpha_2-1} \left(\frac{-\beta_1^{-\alpha_1+2}}{-\alpha_1+2} + \frac{2 x_2 \beta_1^{-\alpha_1+1}}{-\alpha_1+1} - \frac{x_2^2 \beta_1^{-\alpha_1}}{-\alpha_1} \right) dx_2 = \qquad \alpha_1 > 2 \\
&= (\alpha_1 \beta_1^{\alpha_1})(\alpha_2 \beta_2^{\alpha_2}) \int_{\beta_2}^{\infty} \left(\frac{\beta_1^{-\alpha_1+2}}{\alpha_1-2} x_2^{-\alpha_2-1} - \frac{2\beta_1^{-\alpha_1+1}}{\alpha_1-1} x_2^{-\alpha_2} + \frac{\beta_1^{-\alpha_1}}{\alpha_1} x_2^{-\alpha_2+1} \right) dx_2 = \\
&= (\alpha_1 \beta_1^{\alpha_1})(\alpha_2 \beta_2^{\alpha_2}) \left(\frac{\beta_1^{-\alpha_1+2}}{\alpha_1-2} \frac{\beta_2^{-\alpha_2}}{\alpha_2} - \frac{2\beta_1^{-\alpha_1+1}}{\alpha_1-1} \frac{\beta_2^{-\alpha_2+1}}{\alpha_2-1} + \frac{\beta_1^{-\alpha_1}}{\alpha_1} \frac{\beta_2^{-\alpha_2+2}}{\alpha_2-2} \right) = \qquad \alpha_2 > 2 \\
&= \frac{\alpha_1 \beta_1^2}{\alpha_1-2} - \frac{2\alpha_1 \alpha_2 \beta_1 \beta_2}{(\alpha_1-1)(\alpha_2-1)} + \frac{\alpha_2 \beta_2^2}{\alpha_2-2}
\end{aligned}
$$

References

Arlitt, M. F. and Williamson, C. L. (1996). Web server workload characterization: The search for invariants. In *Measurement and Modeling of Computer Systems*, pages 126–137.

Beirlant, J., Vynckier, P., and Teugels, J. L. (1996). Tail Index Estimation, Pareto Quantile Plots, and Regression Diagnostics. *Journal of the American Statistical Association*, 91(436):1659–1667.

Bergstrom, A. (1962). The exact sampling distributions of least squares and maximum likelihood estimators of the marginal propensity to consume. *Econometrica*, 30:480–490.

Bodnarchuk, R. R. and Bunt, R. B. (1991). A synthetic workload model for a distributed sustem file server. In *Proceedings of the 1991 SIGMETRICS Conference on Measurement and Modeling of computer Systems*, pages 50–59.

Bryson, M. (1982). Heavy-tailed distributions. In Kotz, N. and Read, S., editors, *Encyclopedia of Statistical Sciences*, volume 3, New York.

Cowell, F. A. (1995). *Measuring Inequality.* Harvester Wheatsheaf, Hemel Hempstead, second edition.

Crovella, M. and Bestavros, A. (1996). Self-Similarity in World Wide Web Traffic: Evidence and Possible Causes. In *Proceedings of SIGMETRICS'96: The ACM International Conference on Measurement and Modeling of Computer Systems.*, Philadelphia, Pennsylvania. Also, in Performance evaluation review, May 1996, 24(1):160-169.

Danielsson, J. and de Vries, C. (1997). Extreme Returns, Tail Estimation, and Value-at-Risk. *FMG Discussion Paper*, 273.

Davidson, R. and MacKinnon, J. (1998). Graphical Methods for Investigating the Size and Power of Hypothesis Tests. *The Mancester School*, 66(1):1–26.

Embrechts, P. (2001). Extremes in economics and the economics of extremes. *paper presented at SemStat meeting on Extreme Value Theory and Applications.*

Embrechts, P., Klupperlberg, C., and Mikosch, T. (1999). *Modelling Extremal Events for Insurance and Finance.* Springer Verlag, Berlin.

Erdélyi, A. (1953). *Higher Trancendental Functions, Vol.1.* McGraw-Hill, New York.

Feller, W. (1971). *An Introduction to Probability Theory and its Applications*, volume Vol.II. Wiley Press, New York, second edition.

Fieller, E. (1932). The distribution of the index in a noraml bivariate population. *Biometrika*, (24):428–440.

Fiorio, C. and Hajivassiliou, V. (2006). Inference with thick tailed distributions. mimeo.

Gabaix, X. (1999). Zipf's law for cities: an explanation. *The Quarterly Journal of Economics*, 114(3):739–767.

Gibrat, R. (1931). *Les inegalites economiques; applications: aux inegalites des richesses, a la concentration des entreprises, aux populations des villes, aux statistiques des familles, etc., d'une loi nouvelle, la loi de l'effet proportionnel*. Librairie du Recueil Sirey, Paris, France.

Hajivassiliou, V. (2005). Correlation versus statistical dependence and thick tail distributions: Some surprising results. *LSE Department of Economics Working Paper*.

Hart, P. E. and Prais, S. J. (1956). An analysis of business concentration. *Journal of the Royal Statistical Society, A*, 119:150–181.

Hsieh, P.-H. (1999). Robustness of Tail Index Estimation. *Journal of Computational and Graphical Statistics*, 8(2):318–332.

Ibragimov, I. and Linnik, V. (1971). *Independent and Stationary Sequences of Random Variables*. Noordhorf, Groninger-Wolter.

Ibragimov, R. (2004). On the robustness of economic models to heavy-tailedness assumptions. *Cowles Foundation Discussion Paper*.

Ibragimov, R., de la Peña, V., and Sharkhmetov, S. (2003). Characterizations of joint distributions, copulas, information, dependence and decoupling, with applications to econometrics. *Cowles Foundation Discussion Paper*.

Kanter, M. and Steiger, W. (1974). Regression and autoregression with infinite variance. *Advances in Applied Probability*, 6:768–783.

Karlen, D. (2002). Credibility of Confidence Intervals. In *Advanced Statistical Techniques in Particle Physics, Proceedings*, pages 53–57. Grey College, Durham.

King, M. (1980). Robust tests for spherical symmetry and their application to least squares regression. *Annals of Statistics*, 8:1265–1271.

Lebedev, N. (1972). *Special Functions and Their Applications*. Prentice-Hall, Englewood Cliffs.

Loretan, M. and Phillips, P. C. B. (1994). Testing the covariance stationarity of heavy-tailed time series. *Journal of Empirical Finance*, (1):211–248.

Paxson, V. and Floyd, S. (1995). Wide area traffic: the failure of Poisson modeling. *IEEE/ACM Transactions on Networking*, 3(3):226–244.

Phillips, P. C. B. (1982). Exact small sample theory in the simultaneous equation model. In Intrilligator, M. and Griliches, Z., editors, *Handbook of Econometrics*, Amsterdam. North-Holland.

Reed, W. J. and Jorgensen, M. (2003). The double pareto-lognormal distribution a new parametric model for size distributions. mimeo.

Samorodnitsky, G. and Taqqu, M. S. (1994). *Stable Non-Gaussian Random Processes. Stochastic models with infinte variance*. Chapman and Hall, New York.

Steindl, J. (1965). *Random Processes and The Growth of Firms*. Griffin, London.

Tapia, R. and Thompson, J. (1978). *Non Parametric Probability Density Estimation*. The Johns Hopkins University Press, Baltimore.

Victoria-Feser, M. P. and Dupuis, D. J. (2003). A Prediction Error Criterion for Choosing the Lower Quantile in Pareto Index Estimation. *mimeo*.

Zellner, A. (1976). Bayesian and non-bayesian analysis of the regression model with multivariate student-t error terms. *Journal of the American Statistical Association*, 71:400–405.

Zellner, A. (1978). Estimation of functions of population means and regression coefficients including structural coefficients: A minimum expected loss (melo) approach. *Journal of Econometrics*, 8:127–158.

Zipf, G. (1949). *Human behavior and the principle of the last effort*. Addison-Wesley, Cambridge, MA.