Econometrics Journal (2010), volume **11**, pp. 1–19.
Article No. ectj??????

Bimodal t-ratios: The Impact of Thick Tails on Inference

Carlo V. Fiorio[⋆], Vassilis A. Hajivassiliou[†] and Peter C.B. Phillips[‡]

[⋆] *University of Milan and Econpubblica*
E-mail: carlo.fiorio@unimi.it
[†] *London School of Economics and Financial Markets Group*
E-mail: vassilis@econ.lse.ac.uk
[‡] *Yale University, University of Auckland, Singapore Management University, and University of York*
E-mail: peter.phillips@yale.edu
Received: April 2008

Summary 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. Novel density estimation kernel methods are employed, given that our theoretical results yield cases that exhibit density discontinuities.

Keywords: *t-ratio, Bimodality, Thick Tails, Cauchy, Double Pareto*

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 Zipf's law, a discrete form of a Pareto distribution with infinite variance (Gabaix, 1999). Another example is the size distribution of firms Hart and Prais (1956); Steindl (1965). Further, TT distributions frequently arise in financial return data and data on corporate bankruptcies, which can cause difficulties in regulating markets where such extremes are observed Embrechts (2001); Loretan and Phillips (1994). A final example arises in the economics of information technology where Web traffic file sizes follow distributions that decline according to a power law Arlitt and Williamson (1996), often with infinite variance Crovella and Bestavros (1997).

Although there is a large and growing literature on robust estimation with data following thick tail distributions (e.g., Dupuis and Victoria-Feser (2006); 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. (1972), 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 (1987), who analyzed bimodality in classical t-ratios. Nelson and Startz (1990) and Maddala and Jeong (1992) provided some further analysis of structural estimators with possibly weak instruments. More recent contributions include Woglom (2001), Hillier (2006), Forchini (2006), and Phillips (2006), who all consider bimodality in structural equation distributions. Not much emphasis in this literature has been placed on the difference between orthogonal and fully independent observations.

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. In Section 2 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. A simulation analysis of this case is given in Section 3. Novel density estimation kernel methods are employed, given that our theoretical results yield cases that exhibit density discontinuities. Section 4 considers the different implications of lack of correlation and statistical independence. Section 5 illustrates extensions to other distributions with heavy tails: the Stable family of distributions (subsection 5.1) and a symmetric double Pareto distribution (subsection 5.2), which allows tail thickness and existence of moments to be determined parametrically. Section 6 investigates inference in the context of t-ratios with TT distributions. Section 7 shows that bimodality can arise even with TT distributions trimmed to have finite support. Section 8 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 the 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 $R = \frac{a+x}{b+y}$ 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 is given by Marsaglia (1965), who 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 by Hillier (2006), Forchini (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

$$pdf(x) = \frac{1}{\pi(1+x^2)} \tag{2.1}$$

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

$$pdf(x) = \frac{\Gamma\left(\frac{n+1}{2}\right)}{\pi^{(n+1)/2}(1+x'x)^{(n+1)/2}} \tag{2.2}$$

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

$$\int_0^\infty N(0, \sigma^2 I_n) dG(\sigma^2) \tag{2.3}$$

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 (2.2) are all Cauchy. In particular, the distribution of X_i is univariate Cauchy with density as in (2.1) 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 (2.2) from the random sample Cauchy case, is dramatically illustrated by the distribution of the classical t-statistic:

$$t_X = \frac{\overline{X}}{S_X} = \frac{n^{-1}\Sigma_1^n X_i}{\{n^{-2}\Sigma_1^n (X_i - \overline{X})^2\}^{1/2}} \tag{2.4}$$

Under (2.2), 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 (2.3) and the fact that t_X is scale invariant.[1] However, the spherical assumption that underlies (2.2) and (2.3) 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 has symmetric density about the origin but with 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.

In the classical t-ratio the numerator and denominator statistics 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 will prove below, they explain the contrasting shapes of the distributions in the two cases.

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

Recalling that for an i.i.d. sample from a Cauchy $(0, 1)$ distribution, the sample mean $\bar{X} \equiv$Cauchy (0,1) for all n, and, of course, $\overline{X} \to X \equiv$Cauchy (0,1) as $n \to \infty$, the following theorem will focus on the distribution of $(\bar{X}, S_X)$ and that of the associated t-ratio statistic.

Theorem 2.1. *Let $(X_1, \cdots, X_n)$ be a random sample from a Cauchy (0,1) distribution with density (2.2). Define*

$$S^2 = n^{-2}\Sigma_1^n X_i^2 \tag{2.5}$$

$$t = \frac{\overline{X}}{S} \tag{2.6}$$

Then:

(a)

$$S^2 \Rightarrow Y$$

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

$$cf_Y(v) = E(e^{ivY}) = exp\left\{-\frac{2}{\pi^{1/2}}cos\left(\frac{\pi}{4}\right)|v|^{1/2}\left[1 - isgn(v)tan\left(\frac{\pi}{4}\right)\right]\right\} \tag{2.7}$$

(b)

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

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

$$cf_{X,Y} = exp\left\{-2\pi^{-1/2}(-iv)^{-1/2}{}_1F_1\left(-\frac{1}{2}, \frac{1}{2}; u^2/4iv\right)\right\} \tag{2.8}$$

[1]This fact may be traced back to original geometric proofs by Fieller (1932).

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

$$cf_{X,Y}(u,v) = exp\left\{-|u| - \pi^{-1/2}e^{-iu^2/4v}\Psi(3/2,3/2;iu^2/4v)\right\} \tag{2.9}$$

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

(c)

$$S^2 - S_X^2 = O_p(n^{-1}) \tag{2.10}$$

$$t - t_X = O_p(n^{-1}) \tag{2.11}$$

(d) The probability density of the t-ratio (2.6) is bimodal, with infinite poles at ± 1.

Proof. See Appendix A.

Theorem 1 establishes the joint distribution of $(\bar{X}, S^2)$ and shows that the distributions of t and t_X, and of S and S_X are respectively asymptotically equivalent.[2]

Note that X_i^2 has density

$$pdf(y) = \frac{1}{\pi y^{1/2}(1+y)}, y > 0 \tag{2.14}$$

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 (2.14); and that the tails are well balanced. Here we have:

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

Note also that the characteristic function of the limiting variate Y given by (2.7) belongs to the general stable family, whose characteristic function (see Ibragimov and Linnik (1971, p. 43)) has the following form:

$$\varphi(v) = exp\left\{i\gamma v - c|v|^\alpha\left[1 - i\beta sgn(v)tan\left(\frac{\pi\alpha}{2}\right)\right]\right\} \tag{2.15}$$

In the case of (2.7) 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 $\beta = 1$. Part (a) of Theorem 1 shows that the denominator of the t ratio (2.6) 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 $nS_X^2 \xrightarrow{p} \sigma^2 = E(X_i^2)$ under general conditions.

[2]For the definition of the hypergeometric functions that appear in (2.8) and (2.9) see Lebedev (1972, Ch. 9). Note that when $u = 0$ (2.8) reduces to

$$exp\left\{-2\pi^{-1/2}(-iv)^{1/2}\right\} \tag{2.12}$$

We now write $-iv$ in polar form as

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

so that

$$(-iv)^{1/2} = |v|^{1/2}e^{-isgn(v)\pi/4} = |v|^{1/2}cos(\pi/4)\left(1 - isgn(v)tan(\pi/4)\right) \tag{2.13}$$

from which it is apparent that (2.8) reduces to the marginal characteristic function of the stable variate Y given earlier in (2.7). When $v = 0$ the representation (2.9) 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.

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 $n \to \infty$.

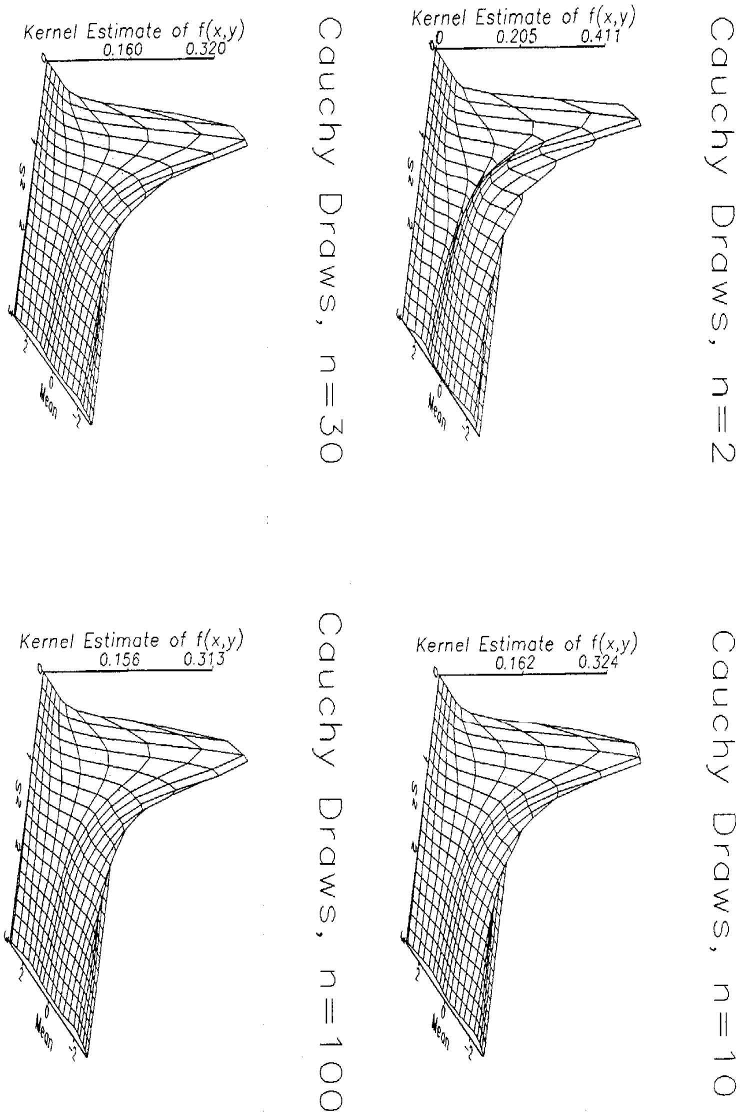


Figure 1. Joint Density Function Estimates of $\bar{X}$ and S^2 for the iid Cauchy DGPs

Figures 1a-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 pictures the density involves a long curving ridge that follows roughly a parabolic shape in the

$(\overline{X}, S^2)$ plane. OLS estimates of the ridge in the joint *pdf* stabilize quickly as a function of n and confirm the dependence between $\overline{X}$ and S^2 for the Cauchy DGP.

Further note that the ridge in the joint density is symmetric about the S^2 axis. The ridge is associated with clusters of probability mass for various values of S^2 on either side of the S^2 axis and equidistant from it. These clusters of mass along the ridge produce a clear bimodality in the conditional distribution of $\bar{X}$ 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 $(\bar{X}, S^2)$ 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.

3. SIMULATION EVIDENCE FOR THE CAUCHY CASE

The empirical densities 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 (2.1) and corresponding cumulative distribution function

$$F(x) = \frac{1}{\pi} arctan(x) + \frac{1}{2}, -\infty < x < \infty. \tag{3.16}$$

Since (3.16) has a closed form inverse, the probability integral transform method was used to generate the draws.

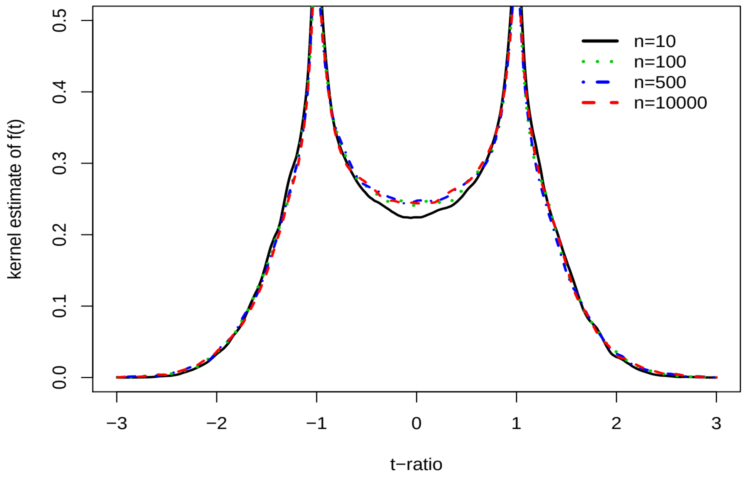
To estimate the probability density functions, conventional kernel methods, e.g., Tapia and Thompson (1978), would not provide consistent estimates of the true density in a neighbourhood of ± 1 in view of the infinite singularities (poles) there. An extensive literature considers how to correct the so-called boundary effect, although there is no single dominating solution that is appropriate for all shapes of density.[3] The method adopted here follows Zhang et al. (1999), which is a combination of methods of pseudodata, transformation and reflection, is nonnegative anywhere, and performs well compared to the existing methods for almost all shapes of densities and especially for densities with substantial mass near the boundary. For the univariate densities (Figures 2, 4, and 5) a bandwidth of $h = 0.2$ was used, while for the bivariate densities in Figure 1, we employed equal bandwidths $h_x = h_y = 0.2$.

We investigated the sampling behavior of the t-ratio statistics t and t_X, by combining four kernel densities, two estimating the density on the left of ± 1 and two estimating the density on the right of ± 1 using the fact that for $x > 1 + h$, $x < -1 - h$ and $-1 + h < x < 1 - h$ the densities estimated with and without boundary correction coincide (Zhang et al. (1999, p. 1234)). These are shown in Figure 2. Note that the bimodality is quite striking and persists for all sample sizes.

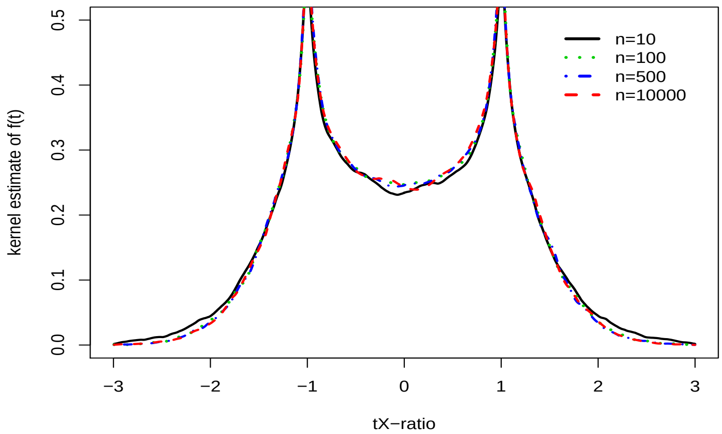
4. LACK OF CORRELATION VERSUS INDEPENDENCE

Data from an n dimensional spherical population with finite second moments have zero correlation, but are independent only when normally distributed. The standard multivari-

[3] For an introductory discussion of density estimation on bounded support, cf. Silverman (1986, p. 29). Methods to correct for the boundary problem include the reflection method Cline and Hart (1991); Silverman (1986), the boundary kernel method Cheng et al. (1997); Jones (1993); Zhang and Karunamuni (1998), the transformation method Marron and Ruppert (1994) and the pseudodata method Cowling and Hall (1996).



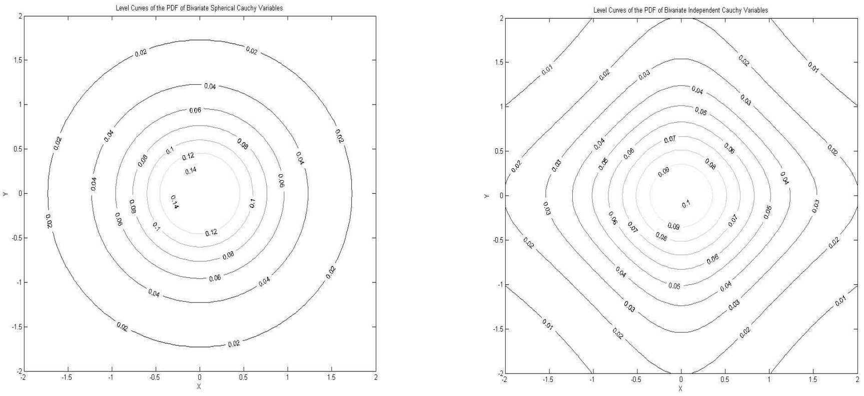
(a) Density Function of the t-ratio



(b) Density Function of the t_x-ratio

Figure 2. Density Functions for iid Cauchy DGPs

ate Cauchy (with density given by (2.2) 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. Examples of this type highlight the statistical implications of the differences between lack of correlation and independence in non-normal populations.



(a) Spherical (Dependent) (b) Independent (Nonspherical)

Figure 3. Bivariate Cauchy: Spherical (Dependent) vs. Independent (Nonspherical)

Figure 3 highlights these differences for the bivariate Cauchy case. The left panel plots the iso-pdf contours of the bivariate *spherical* Cauchy (with the two observations being non-linearly *dependent*), while the right panel gives the contours for the bivariate *independent* Cauchy case (where 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 confirm that for non-Gaussianity, sphericality is neither necessary nor sufficient for independence.[4]

These results confirm the findings of Hajivassiliou (2008), who 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: 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 obtains. 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 density of the t-ratio exhibits the striking bimodality documented here.

[4]Figure 10 of the extended version of this paper, Fiorio et al. (2008) considers 6 representative squares on the domain of the bivariate Cauchy distributions, and calculates various measures of deviation from independence for the spherical, dependent version.

5. IS THE CAUCHY DGP NECESSARY FOR BIMODALITY?

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 the qualitative results given here, such as the bimodality of the t ratios, continue to apply for a much wider class of underlying populations. In this Section we show that the bimodality of the t-ratio persists for other heavy-tailed distributions. Two cases are illustrated: (a) draws from the Stable family of distributions and (b) draws from the "Double-Pareto" distribution.

5.1. Draws from the Stable Family of Distributions

Let $(X_1, \cdots, X_n)$ denote a random sample from a symmetric stable population with characteristic function

$$cf(s) = e^{-|s|^\alpha} \tag{5.17}$$

and exponent parameter $\alpha < 2$ then the t-ratios t and t_X have bimodal densities similar in form to those shown in Figure 2 above for the special case $\alpha = 1$. To generate random variates characterized by (5.17) a procedure described in Section 1 of Kanter and Steiger (1974) was used. In our experiments we considered several examples of stable distributions for various values of a. We found that 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. In a similar vein to the Cauchy case, we found the ridge in the joint density to be most pronounced for $\alpha = 1/3$ but withers as α rises to $5/3$. For extended simulation results see Fiorio et al. (2008).

5.2. Draws from the Double-Pareto Distribution

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.[5] 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 its first two moments are:[6]

$$E(x) = \frac{\alpha_1\beta_1(\alpha_2 - 1) - \alpha_2\beta_2(\alpha_1 - 1)}{(\alpha_1 - 1)(\alpha_2 - 1)} \text{ with } \alpha_1 > 1, \alpha_2 > 1$$

$$V(x) = \frac{\alpha_1\beta_1^2}{\alpha_1 - 2} - \frac{2\alpha_1\alpha_2\beta_1\beta_2}{(\alpha_1 - 1)(\alpha_2 - 1)} + \frac{\alpha_2\beta_2^2}{\alpha_2 - 2} \text{ with } \alpha_1 > 2, \alpha_2 > 2$$

[5] 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.

[6] For derivations, see Appendix B of the extended version of this paper, Fiorio et al. (2008).

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.[7]

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_X 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. For the double Pareto distribution we find that 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 4(a)).

We also examined the case $1 < \alpha < 2$ and found that the t-ratio, t_X, is not always clearly bimodally distributed. The more α departs from 1 the less evident is the bimodality of the t-ratio density and the clearer the convergence towards a standard normal distribution (Figure 4(b)). We set $\beta = 3$ but these results apply for any value of $\beta > 0$, since β is simply a threshold parameter that does not affect the t_X statistic behavior.[8]

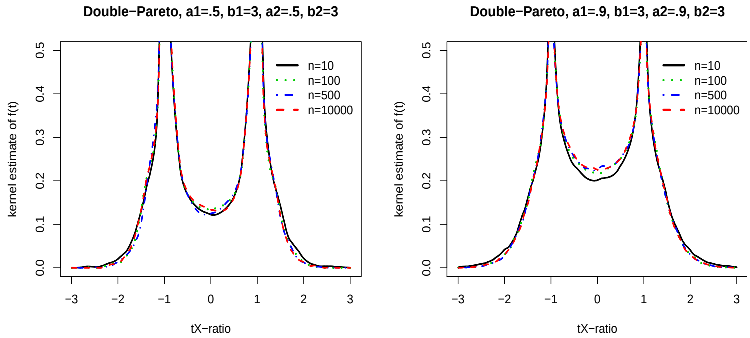
If $\alpha_1 \neq \alpha_2$ it suffices to have either $\alpha_1 \leq 1$ or $\alpha_2 \leq 1$ for the double Pareto to have infinite mean. However, in this case the t-ratio distribution does not have a bimodal density, nor is it stable (see Figure 6 of the extended online version of this paper, Fiorio et al. (2008)).

The regularity in the t_X distribution for the symmetric double Pareto case leads us to investigate the relationship between the first and second centered moments, in the numerator and denominator of t_X 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. This behaviour confirms the surprising results obtained elsewhere Ibragimov (2004); Hajivassiliou (2008) 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 variability actually *rises* with the number of assets if $\alpha < 1$. In such cases, all eggs *should* be placed in the same basket. [9]

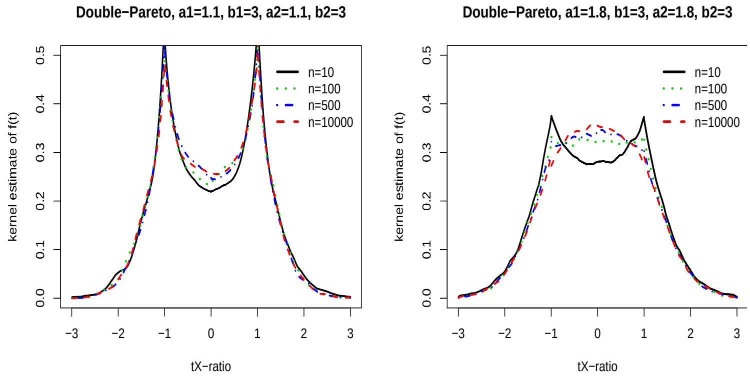
[7]Using copulas, we could evaluate behaviour with *correlated* double Pareto draws. See Hajivassiliou (2008) for a development of this idea. See also Peña et al. (2006) for some general results.

[8]These findings can be proved theoretically along the lines of Appendix A: The theory behind the Double-Pareto Figures 4(a)-4(b) corresponds to the Logan et al. (1972) case of $p = 2$ and $Prob(t < -q) \sim rq^{-\alpha} = Prob(t > q) \sim \ell q^{-\alpha}$ with $r = \ell$. When $0 < \alpha < 1$ as in (4(a)), the density of t_X has infinite singularities at ± 1, while for $1 < \alpha < 2$ as in (4(b)) the density is continuous throughout with modes at ± 1.

[9]For specific analysis of the distribution of the variance of double Pareto distributions with infinite mean, and of the relationship between the sample mean and variance in this case, the interested reader is referred to the extended online version of this paper, Fiorio et al. (2008).



(a) t-ratio of infinite-first-moment double Pareto distributions ($\alpha < 1$)



(b) t-ratio of finite-first-moment double Pareto distributions ($1 < \alpha \leq 2$)

Figure 4. t-ratios for Double Pareto Distributions

6. REJECTION PROBABILITY ERRORS OF T-RATIOS

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).

Let us now summarize results, which are extensively described in Fiorio and Hajivassiliou (2006). 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 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.[10] Second, if the non-symmetric double Pareto distribution is considered, then the t-ratio statistic is not even stable. Finally, although the "ignore" policy leads to minor errors (below $\pm 5\%$) for one sided tests in the case of the double Pareto distribution, the ERP might be much larger for other TT processes.

7. BIMODALITY WITHOUT INFINITE MOMENTS?

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

$$Z = \begin{cases} X & iff \quad |X| < c \\ NA & otherwise \end{cases} \tag{7.18}$$

where X is a standard Cauchy(0,1). 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 first trimmed distribution truncated on a compact support as in (7.18) that we consider is the Cauchy $X \sim Cauchy(0,1)$, while the second is the double Pareto law introduced introduced in subsection 5.2.

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

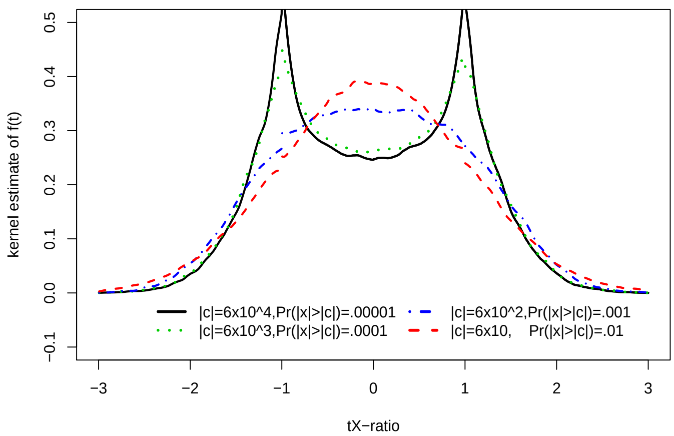
The higher the absolute value of c is, the less attenuated the impact of tail behaviour will be. In contrast, low absolute values of c imply cutting out most of the (thick) tails of the distribution.

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.

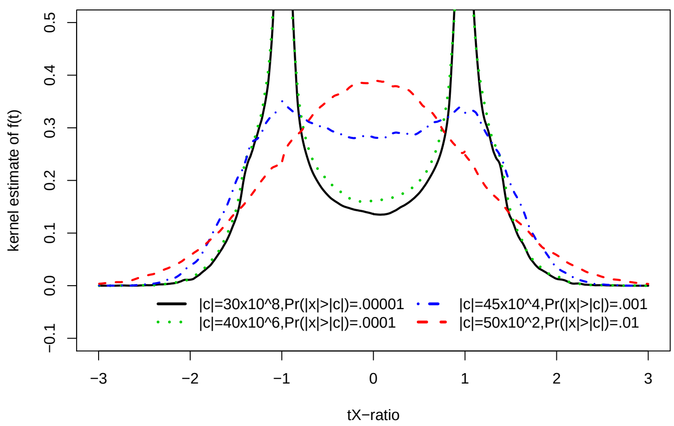
[10] Although tests of size larger than 10% are rather unusual in economics it is much less so in other disciplines, such as physics, where the main point is often to maximize the power of the test, rather than to minimize its size. Also in physics and other related sciences, it is common to consider the "probable error" of a test procedure, which corresponds to a significance level of 50%. In such cases it is common to find confidence intervals with about 60% coverage probability (see for instance Karlen, 2002).

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 5), 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 ± 1, 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.



(a) Truncated Cauchy, $n = 500$.



(b) Truncated Double Pareto, $a1 = a2 = .5$, $b1 = b2 = 3$, $n = 500$.

Figure 5. t-ratio of Cauchy and double Pareto on compact support. Symmetric around zero. Trimming Points (c)

8. CONCLUSIONS

This paper has investigated issues of inference from data based on independent draws from TT distributions. When the distribution is TT with infinite moments, the standard t-ratio formed from a random sample does not converge to a standard normal distribution and the limit distribution is bimodal. Conventional inference is invalidated in such cases and errors in the rejection probability in testing can be serious. 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.

9. APPENDIX A: PROOF OF THEOREM 1

Proof Part (a):

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

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

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:

$$E\left(e^{ivS^2}\right) = \Pi_{i=1}^n E\left(e^{ivX_i^2/n^2}\right) = \left[\left(\Gamma\left(\frac{1}{2}\right)\right)^{-1} \Psi\left(\frac{1}{2},\frac{1}{2};-iv/n^2\right)\right]^n \tag{9.19}$$

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 (9.19) 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 (2.12) to (2.13) we deduce (2.7) as stated.

Part (b):

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

$$\int_h e^{zrh/2}dh = {}_0F_1\left(\frac{1}{2},\frac{1}{4}z^2r\right) = \Sigma_{k=0}^{\infty}\frac{(z^24)^k r^k}{k!\left(\frac{1}{2}\right)_k}$$

we obtain

$$L(z,w) = \frac{1}{\pi}\sum_{k=0}^{\infty}\frac{(z^2/4)^k}{k!\left(\frac{1}{2}\right)_k}\int_0^{\infty}\frac{e^{wr}r^{k-1/2}}{(1+r)}dr = \frac{1}{\pi}\sum_{k=0}^{\infty}\frac{(z^2/4)^k\Gamma\left(k+\frac{1}{2}\right)}{k!\left(\frac{1}{2}\right)_k}\Psi\left(k+\frac{1}{2},k+\frac{1}{2},-w\right) \tag{9.20}$$

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

$$\Psi\left(k+\frac{1}{2},k+\frac{1}{2};-w\right) = \Gamma\left(\frac{1}{2}-k\right){}_1F_1\left(k+\frac{1}{2},k+\frac{1}{2};-w\right) \tag{9.21}$$

$$+\frac{\Gamma\left(k-\frac{1}{2}\right)}{\Gamma\left(k+\frac{1}{2}\right)}(-w)^{1/2-k}{}_1F_1\left(1,\frac{3}{2}-k;-w\right)$$

(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 (9.20) and (9.21) we have:

$$L(z,w)=\sum_{k=0}^{\infty}\frac{(-z^2/4)^k}{k!\left(\frac{1}{2}\right)_k}e^{-w} \tag{9.22}$$
$$+\frac{1}{\pi}\sum_{k=0}^{\infty}\frac{(z^2/4)^k\Gamma\left(k-\frac{1}{2}\right)}{k!\left(\frac{1}{2}\right)_k}(-w)^{1/2-k}{}_1F_1\left(1,\frac{3}{2}-k;-w\right)$$

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

It follows from (9.22) 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
$\left[L\left(\frac{iu}{T},\frac{iv}{T^2}\right)\right]^T\rightarrow exp\left\{\frac{\Gamma\left(-\frac{1}{2}\right)}{\pi}{}_1F_1\left(-\frac{1}{2},\frac{1}{2};\frac{u^2}{4iv}\right)(-iv)^{1/2}\right\}$
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

$$cf_{X,Y}(u,v)=exp\left\{-\frac{2}{\pi^{1/2}}{}_1F_1\left(-\frac{1}{2},\frac{1}{2};\frac{u^2}{4iv}\right)(-iv)^{(1/2)}\right\} \tag{9.23}$$

as required for (2.8).

The second representation in this part of the Theorem 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

$$\left(-\frac{1}{2}\right)^{-1}(-iv)^{1/2}{}_1F_1\left(-\frac{1}{2},\frac{1}{2};\frac{u^2}{4iv}\right)=\frac{1}{2}|u|\left\{\Gamma\left(-\frac{1}{2}\right)-e^{u^2/4iv}\Psi\left(\frac{3}{2},\frac{3}{2};\frac{-u^2}{4iv}\right)\right\}. \tag{9.24}$$

Using (9.24) in (9.23) we obtain (2.9) as stated.

Part (c):

To prove equations (2.10) and (2.11), note that
$S_X^2=S^2-n^{-1}\overline{X}^2=S^2+O_p(n^{-1})$ 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.

Part (d):

To prove that the density of the t-ratio has singularities with infinite poles at ± 1, it suffices to note that in the notation of Logan et al. (1972), the case of the t-ratio (2.6) based on i.i.d. Cauchy draws corresponds to their parameters: $p=2$, $\alpha=1$, and $r/l=1$. Then their equations (5.1) and (5.2) and Lemmas A and B guarantee the result.

ACKNOWLEDGEMENTS

The authors would like to thank the Coeditor and anonymous referees for constructive comments and suggestions. The usual disclaimer applies.

REFERENCES

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

Beirlant, J., P. Vynckier, and J. L. Teugels (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.

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

Cheng, M., J. Fan, and J. Marron (1997). On automatic boundary corrections. *The Annals of Statistics 25*, 1691–1708.

Cline, D. and J. Hart (1991). Kernel estimation of densities of discontinousus derivatives. *Statistics 22*, 69–84.

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

Cowling, A. and P. Hall (1996). On pseudodata methods for removing boundary effects in kernel density estimation. *Journal of the Royal Statistical Society, Ser. B 551–563*, 58.

Crovella, M. and A. Bestavros (1997, December). Self-Similarity in World Wide Web Traffic: Evidence and Possible Causes. *IEEE/ACM Transactions on Networking 5*(6), 835–846.

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

Dupuis, D. J. and M. P. Victoria-Feser (2006). A Robust Prediction Error Criterion for Pareto Modeling of Upper Tails. *Canadian Journal of Statistics 34*, 639–658.

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

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

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

Feller, W. (1971). *An Introduction to Probability Theory and its Applications* (Second ed.), Volume Vol.II. New York: Wiley Press.

Fieller, E. C. (1932). The distribution of the index in a normal bivariate population. *Biometrika 24*(3/4), 428–440.

Fiorio, C., V. Hajivassiliou, and P. Phillips (2008). Bimodal t-ratios: The impact of thick tails on inference. Downloadable from:
`http://econ.lse.ac.uk/staff/vassilis/papers/`.

Forchini, G. (2006). On the bimodality of the exact distribution of the tsls estimator. *Econometric Theory 22, No.5*, 932–946.

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

Hajivassiliou, V. (2008). Correlation versus statistical dependence and thick tail distributions: Some surprising results. LSE Department of Economics Working Paper.

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

Hillier, G. (2006). Yet more on the exact properties of iv estimators. *Econometric Theory 22, No.5*, 913 – 931.

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

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

Ibragimov, R. (2004). On the robustness of economic models to heavy-tailedness assumptions. Department of Economics, Yale University.

Jones, M. (1993). Simple boundary correction for kernel density estimation. *Statistics and Computing 3*, 135–146.

Kanter, M. and W. Steiger (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*, pp. 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*. Englewood Cliffs: Prentice-Hall.

Logan, B., C. Mallows, S. Rice, and L. Shepp (1972). Limit distributions of self-normalized sums. *Annals of Probability 1*, 788–809.

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

Maddala, G. and J. Jeong (1992). On the exact small sample disttribution of the instrumental variable estimator. *Econometrica 60*, 181–183.

Marron, J. and D. Ruppert (1994). Transformations to reduce boundary bias in kernel density estimation. *Journal of the Royal Statistical Society, Ser. B 56*, 653–671.

Marsaglia, G. (1965). Ratios of normal variables and ratios of sums of uniform variables. *Journal of the American Statistical Association 60*, 193–204.

Nelson, C. and R. Startz (1990). Some further results on the exact small sample properties of the instrumental variables estimator. *Econometrica 58*, 967–976.

Peña, V. H. d. l., R. Ibragimov, and S. Sharakhmetov (2006). Characterizations of joint distributions, copulas, information, dependence and decoupling, with applications to time series. *Lecture Notes-Monograph Series 49*, 183–209.

Phillips, P. (2006). A remark on bimodality and weak instrumentation in structural equation estimation. *Econometric Theory 22, No.5*, 947–960.

Phillips, P. and M. Wickens (1978). *Exercises in Econometrics*. Oxford: Philip Allan.

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

Phillips, P. C. B. and V. A. Hajivassiliou (1987). Bimodal t-ratios. Cowles Foundation Discussion Paper No.842.

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

Silverman, B. (1986). *Density Estimation for Statistics and Data Analysis*. London: Chapman and Hall.

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

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

Woglom, G. (2001). More results on the exact small sample properties of the instrumental variable estimator. *Econometrica 69*, 1381–1389.

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.

Zhang, S. and R. Karunamuni (1998). On kernel density estimation near endpoints. *Journal of Statistical Planning and Inference 70*, 301–316.

Zhang, S., R. Karunamuni, and M. Jones (1999). An improved estimator of the density function at the boundary. *Journal of the American Statistical Association 94*, 1231–1241.