Estimation of Stable Distributions by Indirect Inference*

René Garcia[†], Eric Renault[‡] and David Veredas[§]

First Draft February 2003, this version May 2009

Abstract

This article deals with the estimation of the parameters of an α-stable distribution with indirect inference, using the skewed-t distribution as an auxiliary model. The latter distribution appears as a good candidate since it has the same number of parameters as the α-stable distribution, with each parameter playing a similar role. To improve the properties of the estimator in finite sample, we use constrained indirect inference. In a Monte Carlo study we show that this method delivers estimators with good properties in finite sample. We provide an empirical application to the distribution of jumps in the S&P 500 index returns.

Keywords: Stable distribution, Indirect Inference, Constrained Indirect Inference, Skewed-t distribution.

JEL classification: C13, C15, G11.

*We are grateful to Marine Carrasco, Geert Dhaene, François Lamontagne, Marco Lombardi, Marc Paolella, Mark Steel and Casper de Vries for useful discussions and remarks as well as the audiences at the CIRANO-CIREQ Conference in Financial Econometrics 2004, ESEM 2004, Canadian Econometric Study Group 2004, York University, Conference on Heavy tails and stable Paretian distributions 2005 organized by the Deutsche Bundesbank and the Econometrics Seminar at the University of Copenhagen and Tilburg University. Marine Carrasco kindly provided us with the computer code for estimating α-stable distributions with continuous GMM. The first two authors gratefully acknowledge financial support from the Institut de Finance Mathématique de Montréal (IFM2), the Fonds québécois de la recherche sur la société et la culture (FQRSC), the Social Sciences and Humanities Research Council of Canada (SSHRC) and the MITACS Network of Centres of Excellence. The first author is grateful to Hydro-Québec and the Bank of Canada for financial support. An important part of this work was carried out while the third author was a postdoctoral fellow at CIRANO and later a Marie Curie postdoctoral fellow at Tilburg University. The third author thanks CIRANO, CIREQ, IFM2 and the European Community's Human Potential Programme under contract HPRN-CT-2002-00232, [MICFINMA] for financial support.

[†]Author for correspondence. Edhec Business School and CIRANO; email: rene.garcia@edhec.edu.

[‡]University of North Carolina at Chapel Hill, CIRANO and CIREQ; email: renault@email.unc.edu.

[§]ECARES, Université Libre de Bruxelles; email: dveredas@ulb.ac.be.

1 Introduction

The α-stable distribution has been widely used for fitting data in which extreme values are frequent. As shown in early work by Mandelbrot (1963) and Fama (1965a), it accommodates heavy-tailed financial series, and therefore produces more reliable measures of tail risk. The α-stable distribution is also able to capture skewness in a distribution, which is another characteristic feature of financial series. The distribution is also preserved under convolution. This property is appealing when considering portfolios of assets, especially when the skewness and fat tails of returns are taken into account to determine the optimal portfolio.[1] Stable processes have recently been used in the high-frequency microstructure literature by Ait-Sahalia and Jacod (2007, 2008) who proposed volatility estimators for some processes built from the sum of a stable process and another Levy process.

To estimate the parameters of an α-stable distribution we propose to use indirect inference (see Smith, 1993 and Gouriéroux, Monfort and Renault, 1993, GMR hereafter), a method particularly suited to situations where the model of interest is difficult to estimate but relatively easy to simulate. Indeed, the α-stable density function does not have a closed-form expression and is only characterized as an integral difficult to compute numerically, making ML estimation not very appealing in practice.[2] However, several methods are available to simulate α-stable random variables, such as the one described in Chambers, Mallows and Stuck (1976).

Indirect inference involves the use of an auxiliary model. Auxiliary parameters are recovered through maximization of the pseudo-likelihood of a model based on the fictitious *i.i.d.* sampling in a skewed-t distribution of Fernandez and Steel (1998).[3] It is a Student-t with an inverse scale factor in the positive and negative orthants, allowing for asymmetries. The distribution has four parameters which have a one-to-one correspondence with the parameters of the α-stable distribution. There is a clear and interpretable matching between the two sets, parameter by parameter.[4]

Our application of indirect inference is innovative in two respects. First, following McCulloch (1986) in the context of matching quantiles, we actually perform a con-

[1]Basic references on the α-stable distribution are Feller (1971), Zolotarev (1986) and Samorodnitsky and Taqqu (1994). Its properties motivate its use in the modelling of financial series in particular by Carr et al. (2002) and Mittnik, Paolella and Rachev (2000). For value-at-risk applications, see in particular Bassi et al. (1998) and Mittnik, Rachev and Paolella (1998). For portfolio allocation with stable distributions, see Fama (1965b), Bawa, Elton and Gruber (1979), and Ortobelli, Huber and Schwartz (2002).

[2]Nevertheless, DuMouchel (1973) has shown that the maximum likelihood (ML hereafter) estimator is consistent, asymptotically normal and reaches the Cramer-Rao efficiency bound.

[3]Hansen (1994) also proposes a skewed version of the Student-t. The way skewness in introduced differs from that of Fernandez and Steel (1998).

[4]During the course of this project, we were made aware by M. J. Lombardi that Lombardi and Calzolari (2008) use the same auxiliary model to estimate a stable distribution. The two projects were conducted independently and differ in several respects.

strained version of indirect inference, introducing an a priori constraint on one auxiliary parameter to match, namely the number of degrees of freedom of the Student-t. The theory for such constrained indirect inference (CII hereafter) has been developed in a general context by Calzolari, Fiorentini and Sentana (2004) (CFS hereafter). Second, we stress in our application that the α-stable simulator need not to take into account the actual dynamic features of the data.[5] We show that, for a reasonable level of asymmetry, the pseudo-ML estimators of the four parameters of the skewed-t distribution are asymptotically normal even when the observations are generated by an α-stable distribution.[6] Consequently, the associated indirect inference estimators of the parameters of the α-stable distribution are asymptotically normal too.

We compare our method to two moment-based estimation methods. McCulloch (1986) proposed a quantile-based estimator, by building sample counterparts of the cumulative distribution function. Another approach is to match moments produced by the characteristic function (CF hereafter). Carrasco and Florens (2000, 2002) devise an optimal generalized method of moments based on a continuum of moment conditions corresponding to the CF computed at all points. Called continuous GMM (CGMM), the method produces an efficient estimator and overcomes the necessity of choosing an arbitrary set of frequencies, which was a fundamental drawback of CF-based methods.[7]

In a Monte Carlo study, we compare our estimator with CGMM and report that it is often more efficient in finite sample. Since DuMouchel (1973) provides a way to compute the efficiency bound in the *i.i.d.* case, we are able to measure the performance of our indirect inference estimator with the ML benchmark. At least in this *i.i.d.* setting, the efficiency loss appears mainly negligible given the finite sample improvement brought about by indirect inference. We also compare our method to the simple but inefficient quantile-based estimator of McCulloch (1986). Our estimates are close to those obtained with the quantile-based method. However, our estimators appear to have a much smaller variance, both asymptotically and in finite sample.

Many of the properties of stable models are shared by GARCH models. In particular, both models share the facts that the unconditional distribution has fat tails and that the tail shape is invariant under aggregation (see Ghose and Kroner 1995 and de Vries 1991).[8] We illustrate this observational equivalence by generating different GARCH(1,1) and IGARCH(1,1) with Gaussian and Student-t innovations and aggregating the generated

[5]The use of a wrongly-specified simulator in indirect inference has not received much attention, except in Dridi, Guay and Renault (2007).

[6]According to our Monte Carlo experiments, the allowed level of asymmetry is actually consistent with the one produced by an α-stable distribution with support on the whole real line.

[7]Some authors, like Fielitz and Rozelle (1981), recommend to match only a few frequencies on the basis of Monte Carlo results, while others, like Feuerverger and McDunnough (1981), recommend on the contrary to use as many frequencies as possible.

[8]It is well-known that, except for the limiting case of the normal distribution, all the α-stable distributions have infinite variance. However, it should be remembered that a highly persistent GARCH with, by definition, finite conditional variances, may produce infinite moments at orders not much higher than two.

processes to lower frequencies. We show that the unconditional density captures very well the variance and kurtosis through aggregation and memory. The tail index α remains relatively constant under aggregation while the estimated dispersion increases.

We complete our analysis by applying our estimation procedure to a series of realized jumps filtered from the S&P 500 return series using the methodology of Tauchen and Zhou (2009). We find that the stable distribution that best characterizes these jumps is symmetric with an estimated tail index of 1.7.

The rest of the paper is organized as follows. Section 2 briefly describes the properties of α-stable distributions and their estimation by CGMM and empirical quantiles. In section 3 we detail the application of the indirect inference methodology to the α-stable distribution, using the skewed-t distribution as an auxiliary model. We discuss the primitive conditions that warrant identification of structural parameters and asymptotic normality of their indirect inference estimators. Section 4 reports the results of a Monte Carlo study where indirect inference is compared to CGMM and empirical quantiles. The superior performance of CII is documented through both asymptotic and Monte Carlo MSE. We also compare and illustrate through simulations the relationship between the fat-tailed unconditional distributions produced by highly persistent GARCH models and an α-stable model. Section 5 is devoted to an empirical application to jumps in equity returns. Section 6 concludes. Proofs to several propositions are provided in the Appendix.

2 The α-stable distributions, CGMM and empirical quantiles

The α-stable family of distributions is characterized by four parameters α, β, σ and μ, where α is the stability parameter, β the skewness parameter, σ the scale parameter, and μ the location parameter. These parameters define the natural logarithm of the characteristic function as

$$\ln \psi_\theta(t) = \ln E\left(\exp(it\,Y)\right) = i\mu t - \sigma^\alpha \left|t\right|^\alpha \left[1 - i\beta\; sign(t) \tan(\pi\alpha/2)\right]$$

where $\theta = (\alpha, \beta, \sigma, \mu) \in \Theta =]1, 2] \times [-1, 1] \times \mathbb{R}_+^* \times \mathbb{R}$, Y is the random variable following the α-stable distribution $S(\theta)$ with characteristic function $\psi_\theta(\cdot)$ and $sign(t) = t/\left|t\right|$ for $t \neq 0$ (and 0 for $t = 0$). Note that the α-stable distribution can also be defined for α smaller than 1 but we preclude this case to guarantee the existence of a finite expectation. This requirement is rather realistic for the application to financial returns we have in mind. More generally, $E(|Y|^p) < \infty$ for all $p < \alpha$ and in particular $E(Y) = \mu$.

Even though the likelihood function is not known in closed form in general, the score function for an *i.i.d.* sample of size n remains asymptotically root-n normal. Therefore, DuMouchel (1973) was able to show that the standard tools of maximum likelihood theory (mainly root-n asymptotic normality and Cramer-Rao bounds) may be applied to estimation of θ insofar as its domain is limited to $|\beta| < \min(\alpha, 2-\alpha)$. This result implies that efficient estimation of the parameters of α-stable distributions remains a sensible goal and that asymptotic normality of M-estimators like MLE or QMLE can be derived by the application of standard central limit theory to well-chosen (pseudo)-score functions rather than to moments of Y, which do not exist. This idea is the main motivation of the indirect inference strategy proposed in this paper.

Other estimation methods are available. Since the theoretical characteristic function has a closed form, estimation can be performed by fitting the sample characteristic function $n^{-1}\sum_{j=1}^{n}\exp(it_kY_j)$ to the theoretical one $\psi_\theta(t_k)$, defined on a grid of frequencies $t_k, k = 1, \ldots, K$. The problem is that it takes an infinite number of moment conditions, indexed by $t_k \in \mathbb{R}$, to summarize the informational content of the characteristic function. Consider the moment conditions:

$$E(h(t_k, Y, \theta)) = 0, \qquad \forall k = 1, \ldots, K, \tag{1}$$

where $h(t_k, Y, \theta) = \exp(itY) - \psi_\theta(t_k)$. They amount to a set of $2K$ moment restrictions $E(g_k(\theta, Y)) = 0$ that include the real and imaginary parts of $h(t_k, Y, \theta)$. Standard GMM estimates are solutions of $\min \left\|\Omega_n^{-1/2} h_n(., Y, \theta)\right\|$ where $h_n(., Y, \theta)$ is the sample counterpart of (1) and Ω_n is an estimate of the covariance operator. As first noticed by Carrasco and Florens (2002), when the grid becomes sufficiently thin (K large), it is not possible to estimate efficiently θ since the $2K$ estimating functions $g_k(\theta, Y)$ tend to become collinear when K goes to infinity. As a result, the inverse of Ω_n is not continuous and it needs to be stabilize by the introduction of a regularization parameter δ_n. This motivated Carrasco and Florens (2002) to introduce CGMM, which is based on the whole continuum of moment conditions, to estimate θ as

$$\hat{\theta}_n = \arg\min_{\theta\in\Theta} \left\| \left(\Omega_n^{\delta_n}\right)^{-1/2} h_n(., Y, \theta) \right\|,$$

where $\left(\Omega_n^{\delta_n}\right)^{-1/2} = \left(\delta_n K + \Omega_n^2\right)^{-1/2} \Omega_n^{1/2}$. Carrasco, Chernov, Florens and Ghysels (2007) show that for a sequence δ_n such that $\delta_n \longrightarrow 0$ but $n\delta_n^{5/2} \longrightarrow \infty$ when $n \longrightarrow \infty$, $\hat{\theta}_n$ is not only optimal among GMM estimators but reaches the Cramer-Rao efficiency bound.

An alternative moment-based estimation method is proposed by McCulloch (1986), extending an idea of Fama and Roll (1971). In fact, this estimator can be seen as a particular case of indirect inference. Let us denote by x_p the p-th population quantile of $S(\theta)$, i.e. $P(Y < x_p) = p$. For any different values $p, q, p', q' \in]0, 1[$,

$$\frac{x_p - x_q}{x'_p - x'_q}$$

is independent of both μ and σ. The idea of McCulloch (1986) is to define two functions of such ratios, $\Phi_1 = \Phi_1(\alpha, \beta)$ and $\Phi_2 = \Phi_2(\alpha, \beta)$, that can be interpreted as two auxiliary parameters, which will allow to back out α and β. To get accurate estimators, it is intuitive to focus the first auxiliary parameter on α (for each β) and the second on β (for each α). McCulloch (1986) proposes to define Φ_1 as a measure of the relative size of the tails with respect to the middle of the distribution:

$$\Phi_1(\alpha, \beta) = \max\left(\frac{x_{0.95} - x_{0.05}}{x_{0.75} - x_{0.25}}, 2.439\right), \tag{2}$$

where 2.439 is the smallest possible value of $\Psi_1(\alpha, \beta)$, when α increases to 2, and as it is irrespective of the value of β, $\hat{\beta}_n$ is not identified. The second function

$$\Phi_2(\alpha, \beta) = \frac{(x_{0.95} - x_{0.5}) - (x_{0.5} - x_{0.05})}{x_{0.95} - x_{0.05}} \tag{3}$$

is defined as a measure of the spread between the right part and the left part of the distribution. Replacing population quantiles x_p by their sample counterparts $\hat{x}_{p,n}$, the estimators $\hat{\alpha}_n$ and $\hat{\beta}_n$ are the solutions of $\Phi_1\left(\hat{\alpha}_n, \hat{\beta}_n\right) = \hat{\Phi}_{1,n}$ and $\Phi_2\left(\hat{\alpha}_n, \hat{\beta}_n\right) = \hat{\Phi}_{2,n}$.

On the other hand, for any $p, q \in]0,1[$, $x_p - x_q$ is independent of μ and proportional to σ, so it is natural to define an estimator of the scale parameter σ through the auxiliary parameter defined by

$$\Phi_3(\alpha, \beta) = \frac{x_{0.75} - x_{0.25}}{\sigma},$$

a standardized quantity that does not depend on μ and σ. The estimator $\hat{\Phi}_{3,n}$ of the auxiliary parameter Φ_3 is deduced from the previous estimation of (α, β): $\hat{\Phi}_{3,n} = \Phi_3\left(\hat{\alpha}_n, \hat{\beta}_n\right)$. And the estimator $\hat{\sigma}_n$ can be recovered from:

$$\hat{\sigma}_n = \frac{x_{0.75} - x_{0.25}}{\hat{\Phi}_{3,n}}. \tag{4}$$

Finally, to back out the location parameter μ, it is natural to locate it with respect to the median $x_{0.5}$ of the distribution through

$$\Phi_4(\alpha, \beta) = \frac{\mu - x_{0.5}}{\sigma} + \beta \tan\left(\Pi\frac{\alpha}{2}\right).$$

From the estimator $\hat{\Phi}_{4,n} = \Phi_4\left(\hat{\alpha}_n, \hat{\beta}_n\right)$ we can then deduce the estimator of μ from the previously defined estimators of (α, β, σ):

$$\hat{\mu}_n = \hat{x}_{0.5,n} + \hat{\sigma}_n\left(\hat{\Phi}_{4,n} - \hat{\beta}_n \tan\left(\Pi\frac{\hat{\alpha}_n}{2}\right)\right). \tag{5}$$

Therefore, to back out the indirect estimator $\hat{\theta}_n$ from the estimator $\hat{\Phi}_n$ of auxiliary parameters, one has just to invert the binding function Φ defined as (2)-(5):

$$\Phi(\theta) = \{\Phi_1(\alpha, \beta),\ \Phi_2(\alpha, \beta),\ \sigma,\ \mu\}.$$

The resulting indirect estimator $\hat{\theta}_n$ is nothing but an indirect inference estimator. It turns out that, by contrast with the most standard way to perform indirect inference, no new simulations are needed to recover the indirect inference estimator in this case. The reason for that is that the only two components of the binding function which are not known in closed form ($\Phi_1(\alpha,\beta)$ and $\Phi_2(\alpha,\beta)$) have been tabulated by McCullogh (1986). In other words, the required simulation work is already done. However, the computation of indirect inference standard errors would require to resort to simulations as usual for numerical computation of the partial derivatives of the binding function. This is required indeed only for a grid of values of α and β since the effect of the location-scale parameters μ and σ inside the binding function is known in closed form.

Last, it is worth insisting on the constraint on Φ_1. Although immaterial asymptotically when the unknown true value of α lies in the open interval $]0,2[$, this constraint may play a role in finite sample. When $\hat{\Phi}_{1,n}$ is stuck on the value 2.439, the sample is characterized by a three-dimensional parameter set $\left(\hat{\Phi}_{2,n}, \hat{\Phi}_{3,n}, \hat{\Phi}_{4,n}\right)$, which does not allow to identify the four unknown structural parameters. A sensible solution for this problem is to use the CII theory of CFS. The idea is to replace the lacking fourth auxiliary parameter by the value of the Kuhn-Tucker multiplier associated with the constraint.

For the sake of efficiency, we rather choose in section 3 below to develop the CII approach with instrumental parameters provided by a well suited pseudo-likelihood function rather than by arbitrary quantiles.

3 Constrained Indirect Inference Estimation

3.1 Indirect estimation based on a skewed-t score generator

Indirect inference involves the use of an auxiliary model through its pseudo-likelihood function. For indirect estimation almost as efficient as maximum likelihood, it matters to resort to a pseudo-likelihood whose parameters match well all the information content of the structural parameters of interest. This remark motivates our choice of a pseudo-likelihood associated to the skewed-t distribution as introduced by Fernandez and Steel (1998). This pseudo-likelihood function entails four auxiliary parameters denoted by $\Psi = (\nu,\gamma,\lambda,\omega) \in \Xi = \mathbb{R}_+^* \times \mathbb{R}_+^* \times \mathbb{R}_+^* \times \mathbb{R}$ and its log-likelihood for n observations Y_i, $i = 1,\ldots,n$, is given by

$$L_n = n\log\left(\frac{2h(\nu)}{\sqrt{\pi\nu}} \cdot \frac{1}{\lambda\left(\gamma+\frac{1}{\gamma}\right)}\right) - \frac{\nu+1}{2}\sum_{i=1}^{n}\log\left(1+\frac{1}{\nu}\left(\frac{Y_i-\omega}{\lambda}\right)^2 g_\omega(Y_i,\gamma)\right), \quad (6)$$

where

$$h(\nu) = \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)} \text{ and } g_\omega(y,\gamma) = \begin{cases} \frac{1}{\gamma^2} \text{ if } y \geq \omega \\ \gamma^2 \text{ if } y < \omega. \end{cases}$$

The degree-of-freedom parameter ν (possibly non-integer) of a Student-t distribution captures the thickness of the tails as α does for stable distributions, while the scale parameter λ and the location parameter ω can easily be introduced to match the two parameters σ and μ. Finally, the skewed-t allows one to accommodate skewness through an additional parameter γ that should be informative about β. More precisely, when observations Y_i, $i = 1, \ldots, n$, are identically distributed as Y following the stable distribution $S(\theta)$, the pseudo-log-likelihood function is integrable (i.e. $\log(1 + x^2) < |x|$ for large x) and from its expectation $L(\Psi|\theta) = E(L_1(\Psi)|\theta)$ we can define a pseudo-true value $\Psi(\theta)$ by

$$\Psi(\theta) = \arg\max_{\Psi} L(\Psi|\theta).$$

Note that $\Psi(\theta) = (\nu(\theta), \gamma(\theta), \lambda(\theta), \omega(\theta))$ defines the pseudo true value of the skewed-t parameters $(\nu, \gamma, \lambda, \omega)$ when the true marginal distribution is $S(\theta)$, irrespective of the possible dynamic structure of the data generating process. We expect the auxiliary parameters $(\nu(\theta), \gamma(\theta), \lambda(\theta), \omega(\theta))$ to be very informative about θ. To see this, note that the binding function $\theta \to \Psi(\theta)$ is not only one-to-one but each of its coefficients is well focused on the relevant corresponding structural parameter according to the natural association described in Table 1. The matching displayed in this table is justified by the following result.

Proposition 3.1: *For all* $(\alpha, \beta, \sigma, \mu) \in \Theta$

$$\begin{aligned}
\omega(\alpha, \beta, \sigma, \mu) &= \mu + \omega(\alpha, \beta, \sigma, 0) \\
\lambda(\alpha, \beta, \sigma, \mu) &= \sigma\lambda(\alpha, \beta, 1, \mu/\sigma) \\
\gamma(\alpha, \beta, \sigma, \mu) &= \gamma(\alpha, \beta, 1, 0) = \frac{1}{\gamma(\alpha, -\beta, 1, 0)} \\
\nu(\alpha, \beta, \sigma, \mu) &= \nu(\alpha, \beta, 1, 0) = \nu(\alpha, -\beta, 1, 0).
\end{aligned}$$

Such a nice correspondence between auxiliary and structural parameters leads us to hope that an indirect inference estimator $\hat{\theta}_n$ of structural parameters, obtained as the only value of θ for which $\Psi(\theta)$ equals the pseudo-maximum likelihood estimator $\hat{\Psi}_n$, should be very accurate and almost as efficient as maximum likelihood. To a large extent, this will be confirmed by our Monte Carlo experiments with an exception however concerning the estimation of the tail parameter α when its true unknown value is close to 2. The reason for that is the following: even though α and ν are both the index of regular variation characteristic of the tail of the stable and the skewed-t distribution respectively, the binding function $\nu(\alpha, \beta, \sigma, \mu) = \nu(\alpha, \beta, 1, 0)$ not only does not coincide with α but also becomes less and less informative about α when its true value gets closer and closer to 2. In fact, $\nu(\alpha, \beta, 1, 0)$ is an even function of β that blows up to infinity when α becomes arbitrarily close to 2. For this reason, the inverse problem to recover the structural parameter α from the auxiliary one ν becomes asymptotically ill-posed when α goes to 2. To see this in a simple case, let us make more explicit the symmetric $(\beta = 0)$ case. The relationship between structural and auxiliary parameters is actually even more transparent in the symmetric case. First, we can prove:

Proposition 3.2: *For all* $(\alpha, 0, \sigma, \mu) \in \Theta$

$$\begin{aligned} \omega(\alpha, 0, \sigma, \mu) &= \mu \\ \gamma(\alpha, 0, \sigma, \mu) &= 1. \end{aligned}$$

In other words, in the symmetric case, the structural parameters are exactly equal to their pseudo-true value analogs for both the location and the skewness parameters. As far as the scale and the tail parameter are concerned, we know by proposition 3.1 that in the symmetric case, they are fully characterized from the knowledge of $\lambda(\alpha, 0, 1, 0)$ and $\nu(\alpha, 0, 1, 0)$ that we denote respectively $\lambda(\alpha)$ and $\nu(\alpha)$. We can then prove:

Proposition 3.3: *For all* $\alpha \in]1, 2[$, $\lambda(\alpha)$ *and* $\nu(\alpha)$ *are determined as solution of the two equations:*

$$\begin{aligned} \frac{1}{\nu(\alpha)+1} &= E\left(\frac{Z(\alpha)}{1+Z(\alpha)}\right) \\ \frac{2h'(\nu(\alpha))}{h(\nu(\alpha))} &= E(\log(1+Z(\alpha))), \end{aligned}$$

where $Z(\alpha) = Y^2[\nu(\alpha)\lambda^2(\alpha)]^{-1}$, Y *follows* $S(\theta)$ *with* $\theta = (\alpha, 0, 1, 0)$ *and* $h(\nu) = \Gamma[(\nu+1)/2]/\Gamma[\nu/2]$.

The function $g(\nu) = E(\log(1 + (Y^2/\nu\lambda^2)))$ shows why the problem becomes almost ill-posed when α (and the corresponding ν) becomes large. The absolutely value of the derivative $g'(\nu)$ decreases fast towards its limit zero when ν increases towards infinity. As a results, small increments on α result in huge increments on $\nu(\alpha)$. According to our Monte Carlo work, an increase of α in the interval $[1.4, 1.7]$ results in an increase of ν from 2 to 10. And the variations are even steeper when α becomes close to 2. Even though sensitivity of auxiliary parameters to structural ones should be a good thing for indirect inference, the explosive behavior of ν comes with an explosive behavior of the variance of $\hat{\nu}_n$ which makes indirect inference quite imprecise about α when it is too close to two. Intuitively, when the variance of Y almost exists, the indirect inference matching becomes closer and closer to behave as if the stable distribution were the normal one (corresponding to $\alpha = 2$) and pushes accordingly ν to infinity, with huge variance of $\hat{\nu}_n$ since normality is far from being granted.

This is the reason why we have to constrain the auxiliary parameter Ψ_1, first component of Ψ, by imposing on it an upper bound, exactly as McCulloch (1986) did for his auxiliary parameter that provided information about α. We choose to impose $\nu \leq 2$, that is to redefine Ψ_1 as $\Psi_1 = min(\nu, 2)$. It is worth noting that, while $\alpha < 2$ by definition, the upper bound on ν does not have to be 2. Intuitively, choosing for ν an upper bound larger than 2 is detrimental for the estimation of α and beneficial for the estimation of β. This intuition, confirmed by our Monte Carlo experiments, can be explained as follows.

First, as explained above through the explosive behavior of ν, if we do not constrain ν to be smaller than 2, we loose the tight connection between ν and α as both measure the index of asymptotic regular variation of the tails. Second, in case of an asymmetric stable distribution, we know that the tail behavior has some informative content about β (see Samorodnitsky and Taqqu (1994), property 1.2.15). Even though the binding function $\nu(\alpha, \beta, \sigma, \mu)$ is an even function of β, it is consequently able to convey some information about $|\beta|$ that we basically waste by replacing the auxiliary parameter ν by $min(\nu, 2)$, which is the uninformative constant 2 when the true value of α is large, larger than 1.4 approximately.

Of course, as already mentioned for the quantile-based example, such a constraint on the auxiliary parameter may destroy identification when it remains at its limit value 2. CII provides the right methodology to deal with this problem.

3.2 Constrained Indirect Inference

CII is based on the Lagrangian objective function

$$Q_n(\Upsilon) = \frac{1}{n} L_n(\Psi) + \rho(2 - \nu),$$

where $\Upsilon = (\Psi', \rho)'$. The parameter $\rho \geq 0$ is the Kuhn-Tucker multiplier associated with the constraint $\nu \leq 2$. The estimator $\hat{\Upsilon}_n$ is then defined by the first-order conditions jointly with the complementary slackness restriction $\hat{\rho}_n(2 - \hat{\nu}_n) = 0$ and the inequality restriction. Note that $L_n(\Psi)$ is not differentiable with respect to ω at values that coincide with one of the n observations $Y_i, i = 1, ..., n$. These circumstances can be excluded almost surely. Likewise, let us denote $Y_i^h(\theta), h = 1, \cdots, H$, the components of H simulated paths of an α-stable process for a given value θ. The simulated path $\left(Y_i^h(\theta)\right)_{1 \leq i \leq n}$ defines a simulated criterion function:

$$Q_n^h(\Upsilon | \theta) = \frac{1}{n} L_n^h(\Psi | \theta) + \rho(2 - \nu),$$

and the corresponding simulated estimators $\hat{\Upsilon}_n^h(\theta)$ are defined by the first-order conditions jointly with the slackness condition $\hat{\rho}_n^h(\theta) \cdot \left(2 - \hat{\nu}_n^h(\theta)\right) = 0$ and the inequality restriction (with the same remark as above regarding the innocuous non-differentiability issue). Let us then consider the average estimator over the H simulated paths $\hat{\Upsilon}_{n,H}(\theta) = \frac{1}{H} \sum_{h=1}^{H} \hat{\Upsilon}_n^h(\theta)$.

The main idea of CII is to choose the estimator $\hat{\theta}_n^r$ as a value of θ that matches $\hat{\Upsilon}_{n,H}(\theta)$ against $\hat{\Upsilon}_n$. The superscript r reminds that the estimator is constrained, or restricted, by the Kuhn-Tucker multiplier. Unfortunately, we cannot rely directly on the results in GMR because the constrained estimator $\hat{\Upsilon}_n$ may not be asymptotically normal in large samples in the presence of inequality constraints on its components ν

and ρ. This is the reason why we closely follow below the relevant asymptotic theory of CII, as developed by CFS, while slightly extending it to a case of finite H. Moreover, we can simplify the exposition since we only focus on a just-identified case ($\dim \Psi = \dim \theta$). We first maintain Assumption 1 of CFS:

Assumption 1: $\frac{1}{n}L_n(\Psi)$ *converges almost surely to* $L(\Psi|\theta) = E(L_1(\Psi)|\theta)$ *uniformly in* (θ, Ψ) *as* n *goes to infinity, where* $L(\Psi|\theta)$ *is twice continuously differentiable with respect to its both arguments.*

Note that the differentiability of $L(\Psi|\theta)$ is granted with standard Lebesgue dominated convergence arguments, even with respect to the variable ω, since the non-smoothness of $L_1(\Psi)$, event of probability zero, has no impact on the expectation. For each value of θ we can define the binding function for the constrained auxiliary parameters $\Upsilon = (\Psi', \rho')'$ as $\Upsilon(\theta) = (\Psi^r(\theta)', \rho(\theta)')'$ from the value $\Psi^r(\theta) = (\nu^r(\theta), \gamma^r(\theta), \lambda^r(\theta), \omega^r(\theta))'$, which fulfills the first order conditions:

$$\frac{\partial Q}{\partial \theta}(\Upsilon|\theta)_{|\Upsilon=\Upsilon(\theta)} = 0,$$

with $Q(\Upsilon|\theta) = L(\Psi|\theta) + \rho(2 - \nu)$, and the slackness condition $\rho(\theta)(2 - \nu^r(\theta)) = 0$. In addition, we assume that $\Upsilon^r(\theta)$ is unique, in the sense that $L(\Psi^r(\theta)|\theta) > L(\Psi|\theta)$ for any $\Psi = (\nu, \gamma, \lambda, \omega)'$ in a neighborhood of $\Psi^r(\theta)$ and fulfilling $\nu \leq 2$. As a consequence, assumption 1 ensures the strong consistency when $n \to \infty$ of $\hat{\Upsilon}_n$ (resp. $\hat{\Upsilon}_{n,H}(\theta)$) for $\Upsilon(\theta^0)$ (resp $\Upsilon(\theta)$) where θ^0 denotes the true value of the parameters of the stable distribution. To ensure local identification of θ^0, we maintain a second assumption as in CFS.

Assumption 2: *Rank*$\frac{\partial \Upsilon'(\theta)}{\partial \theta} = 4$ *for any* θ *in a neighborhood of* θ^0.

To interpret assumption 2, it is first worth noting that trivially $\omega^r(\theta) = \omega(\theta)$ while hopefully $\gamma^r(\theta)$ is little different from $\gamma(\theta)$.

A constraint on the tail parameter has no impact on the location parameter and little impact on the skewness parameter, for instance always equal to one in the symmetric case. This remark is actually confirmed by our simulations. Therefore, the fact that the four rows of the matrix $\frac{\partial \Upsilon'(\theta)}{\partial \theta} = [\frac{\partial \Psi^{r\prime}(\theta)}{\partial \theta}, \frac{\partial \rho}{\partial \theta}]$ are linearly independent is expected as a likely consequence of the following arguments: First, by the natural association between structural and auxiliary parameters put forward in the former section and in Table 1, we expect the matrix $\frac{\partial \Psi^{r\prime}(\theta)}{\partial \theta}$ to be almost diagonal and at least, to have its four rows linearly independent. Therefore, the four rows of the matrix $\frac{\partial \Upsilon'(\theta)}{\partial \theta}$ must be linearly independent, at least when the constraint is not binding. Second, when the constraint is binding, we still have a natural correspondence between the two sets of parameters (β, σ, μ) and $(\gamma, \lambda, \omega)$ while the structural tail parameter α will not be captured by ν anymore ($\nu^r(\theta) = 2$) but by the Kuhn-Tucker multiplier $\rho(\theta)$. In other words, we now expect an almost diagonal matrix when considering $\frac{\partial(\gamma^r, \lambda^r, \omega^r, \rho)}{\partial \theta'}$.

For standard (unconstrained) indirect inference, we are here in a just identified setting, so that $\hat{\theta}_n^u$ is simply defined as the solution of the four-equation system $\hat{\Psi}_n^u = \hat{\Psi}_H^u\left(\hat{\theta}_n^u\right)$. The superscripts, u for unconstrained, are just a reminder that the corresponding estimators have been computed by choosing a zero Kuhn-Tucker multiplier: $\hat{\rho}_n$ and $\hat{\rho}_n^h(\theta)$ are fixed to zero, for $h = 1, \cdots, H$. Note that, from Gourieroux, Monfort and Renault (1993), we know that in this just-identified setting, the indirect inference estimator $\hat{\theta}_n^u$ numerically coincides with the score-matching estimator as put forward by Gallant and Tauchen (1996).

By contrast, in order to perform constrained indirect inference, we are faced with a seemingly overidentified problem since both $\hat{\Upsilon}_{n,H}(\theta)$ and $\hat{\Upsilon}_n$ entail five free parameters while the unknown θ is of dimension four. However, we know that this overidentification feature is just a finite sample problem since (see CSF, Proposition 1) the asymptotic distributions of $\hat{\Upsilon}_n$ and $\hat{\Upsilon}_{n,H}(\theta)$ are singular. Therefore, the overidentified finite sample matching problem can be solved by minimizing an arbitrary distance:

$$\hat{\theta}_{n,H}^r = \arg\min_\theta \left(\hat{\Upsilon}_{n,H}(\theta) - \hat{\Upsilon}_n\right)' W_n \left(\hat{\Upsilon}_{n,H}(\theta) - \hat{\Upsilon}_n\right).$$

In terms of asymptotic probability distribution of $\hat{\theta}_n^r$, the choice of the positive definite weighting matrix W_n is immaterial. In fact, Proposition 6 of CFS shows that if $dim(\Psi) = dim(\theta)$, indirect methods provide estimators that are independent of the choice of W_n for large enough n. Note however that when $\hat{\nu}_n$ is stuck at its limit value, the information content of $\hat{\Upsilon}_n$ about the structural parameters θ will go through the Kuhn-Tucker multiplier $\hat{\rho}_n = \frac{1}{n}\frac{\partial L_n}{\partial \nu}(\hat{\Psi}_n^r)$. Therefore, constrained indirect inference will not suffer from the weak identification problem about β that is currently encountered with competing estimation methods when the true unknown value of α is close to 2.

The needed regularity condition for asymptotic theory of constrained constraint indirect inference is maintained as in CFS:

Assumption 3: *There exists some nonstochastic positive definite matrices I_0^r and J_0^r such that*

$$\frac{1}{n}\frac{\partial^2 L_n^2}{\partial \Psi \partial \Psi'}(\Psi_n) \to^p J_0^r$$

for any sequence Ψ_n which is a weakly consistent estimator of $\Psi^r(\theta^0)$, and

$$\sqrt{n}\frac{\partial Q_n}{\partial \Psi}(\Upsilon(\theta^0)) \to^d \mathcal{N}(0, I_0^r).$$

Note that in order to get asymptotically normal indirect estimators, the key assumption is asymptotic normality of the pseudo-score $\sqrt{n}\frac{\partial Q_n}{\partial \Psi}$ computed at the pseudo-true value. It is worth stressing that while observations have infinite variance, the pseudo-score is expected to be well-behaved in the same way we know from DuMouchel (1973) that the true score is well-behaved. Then, standard theory of indirect inference (see GMR) can be applied insofar as the information content in Ψ is sufficient for local identification of θ (assumption 2) and as the pseudo-score is root-n asymptotically normal (assumption 3). Our Monte Carlo evidence (see also Garcia, Renault and Veredas (2006) for extended evidence) shows that we do get probability distributions close to normal for both scores and parameters for large n.[9] This makes a key difference between our approach and a conventional method of moments where asymptotic normality could not be warranted due to the infinite variance of observations.[10] In our setting, the asymptotic variance formula for $\hat{\theta}_n^u$ follows from GMR (p S93) where the cross-derivatives of L_n with respect to Υ and θ are computed by simulations. As a robustness check, which we follow in the application, we also compute the variance-covariance matrix of $\hat{\theta}_n^u$ by bootstrap methods.

As far as constrained estimation is concerned, CFS show that a similar result of asymptotic normality still holds for the parameters of interest in spite of the constraint parameters. While they derive the asymptotic distribution of the constrained indirect inference estimator for an infinite number H of simulated paths, we do extend it here with finite H. As a general principle for simulated methods of moments, we know that the only asymptotic consequence of this is to multiply the asymptotic variance matrix of $\hat{\theta}_n^r$ by a factor $\left(1+\frac{1}{H}\right)$. Thus CFS Proposition 4 (p. 951) allows us to state:

Proposition 3.4: *Under assumptions 1, 2 and 3*

$$\sqrt{n}(\hat{\theta}_{n,H}^r - \theta_0) \to^d \mathcal{N}\left(0, \left(1+\frac{1}{H}\right)(C_0^r)^{-1}\right),$$

where

$$C_0^r = \frac{\partial \Upsilon'}{\partial \theta}(\theta^0, \Upsilon(\theta^0)) J_0^r (I_0^r)^{-1} J_0^r \frac{\partial \Upsilon}{\partial \theta'}(\theta^0, \Upsilon(\theta^0)).$$

Given that the asymptotic distribution of the constrained indirect inference estimator is completely analogous to the standard one, the required asymptotic variance matrix can be consistently estimated by using procedures suggested by GMR (1993). The limit Hessian matrix J_0^r can be trivially estimated by its sample counterpart. The estimation of the asymptotic variance matrix I_0^r may be more involved since, as stressed

[9] For illustration purposes, asymptotic normality of the unconstrained QML estimators is discussed in the appendix. The key argument is that QML is based on self-weighted sample averages with finite variance by contrast with naive sample means. This argument is germane to the asymptotic normality of MLE for GARCH processes even when squared returns have infinite variance (see Francq and Zakoian (2004)).

[10] Note that fat tails may also invalidate the efficiency argument of Efficient Method of Moments (Gallant and Tauchen, 1996) since there is no more reason to hope that a seminonparametric (SNP) score generator based on Hermite expansions will be able to span the true score function. The class of densities to fit with SNP considered by Coppejans and Gallant (2002) are indeed weighted with the exponential function $\exp\left(-x^2/2\right)$ which ensures finite moments at any order.

in the introduction, we never maintain a serial independence assumption. The auxiliary score has been written as if data were *i.i.d.* but we only assume that the sequence of observations is stationary ergodic. Therefore, under standard regularity conditions:

$$I_0^r = \sum_{\tau=-\infty}^{+\infty} S_\tau(\theta_0, \Upsilon(\theta_0)),$$

where

$$\begin{aligned} S_\tau(\theta, \Upsilon) &= E\left(m_i(\Upsilon) m_{i-\tau}(\Upsilon) | \theta\right) \text{ and} \\ m_i(\Upsilon) &= \frac{\partial L_1}{\partial \Psi}(Y_i) - \rho \frac{\partial \nu}{\partial \Psi}. \end{aligned}$$

Therefore we could obtain a consistent estimator of the matrix I_0^r by any standard heteroscedasticity and autocorrelation consistent (HAC) covariance estimation procedure. In the application, we use Newey-West estimators (asymptotic variances in Table 7) and also check their validity by a bootstrap procedure.

4 A Monte Carlo study

In the first subsection, we compare our constrained indirect estimator to the CGMM and empirical quantile estimation methods for various sets of values for θ, the vector of parameters of the stable distribution, assuming that the observations are identically and independently distributed. In the second subsection we show by simulation that some GARCH processes may be observationally equivalent to a stable process from the viewpoint of the unconditional distribution.

4.1 Independent processes

We carry out a Monte Carlo experiment to determine if the good asymptotic properties of the indirect inference estimators with a skewed-t auxiliary model are maintained in a finite sample context. As we have seen in the previous section, the asymptotic distribution of $\hat{\theta}_n$ is determined by the asymptotic distribution of $\hat{\Psi}_n$. Therefore, it is worthwhile to examine the sample distribution of the parameter estimates for the auxiliary model in an experimental setting where we simulate data from a α-stable distribution with different values of the parameters.

We generate 500 samples of 1000 observations for 3 different values of α, namely 0.3, 1.5 and 1.9.[11] We keep the other parameters μ, σ and β fixed and set them equal to

[11] We simulate from the α-stable distribution following Chambers, Mallows and Stuck (1976).

0, 0.5 and 0 respectively.[12] The simulation experiment is divided in two parts. First we estimate the unconstrained skewed-t distribution. Since the true values of Ψ are unknown, we can only check by Monte Carlo the behavior of the first four moments of the estimators. The results are reported in columns 3-6 of Table 2. Concerning ω, λ and γ, skewness and kurtosis of the estimators are fairly close to zero and three respectively. In contrast, $\hat{\nu}_n$ is ill-behaved in finite sample, especially when α approaches to 2. It appears that $\hat{\nu}_n$ has a large upward bias. It is as if, in finite sample, the estimate wants to capture the spurious appearance of normality corresponding to the limit case $\alpha = 2$ and $\nu = +\infty$. Three of the four plots in Figure 1 represent the kernel density of the Monte Carlo distribution of $\hat{\nu}_n$. When $\alpha \geq 1.5$ the estimator $\hat{\nu}_n$ exhibits serious departures from normality. While these kernel densities are for a sample size of 1000, the role of sample size in the dependence of $\hat{\nu}_n$ in α is investigated in the top part of Table 3. For $\alpha = 1.9$, no less than 10000 observations are needed for fairly approaching normality. The implications of these shortcomings of the auxiliary model on the estimation of θ appear in columns 7-10 of Table 2. For $\alpha = 1.9$, skewness and kurtosis depart substantially from the respective values of 0 and 3 for a normal distribution. For $\alpha = 1.5$, the distribution of the estimates is closer to a normal. We conclude from this simulation exercise that the auxiliary model works for most cases but fails when α is close to 2.

Since the estimate $\hat{\nu}_n$ is ill-behaved in finite sample when $\alpha \to 2$, we impose an upper bound on ν, as mentioned in the previous section. To check that the estimated multiplier is well-behaved in finite sample, we plot in the bottom right corner of Figure 1 the density of $\hat{\rho}_n$ when $\alpha = 1.9$. It can be seen that, contrary to $\hat{\nu}_n$, the distribution is much closer to a normal. This confirms why $\hat{\rho}_n$ is a more relevant auxiliary parameter than $\hat{\nu}_n$ in such a case. The bottom part of Table 3 shows that in fact the finite sample behavior of the multiplier is very good for any sample size.

The Monte Carlo study is conducted for $\alpha = 1.5$ and 1.9 and $\beta = 0$ and 0.75.[13] We compare the CII method to CGMM method of Carrasco and Florens (2002) based on the characteristic function with a regularization parameter δ_n equal to 10^{-6}.[14] The second one is the empirical quantile method of McCulloch (1986).[15] The results are reported in Table 4.

As a general assessment, one can say that the constrained indirect inference method delivers consistent estimators which are close to being normally distributed. The skewness for all parameters is close to 0 and the kurtosis close to 3. Thanks to the constraint imposed on ν in the auxiliary model, the estimator behaves well even when α approaches 2. CII compares well with the two other methods. First, with respect to CGMM, it

[12] A more extensive Monte Carlo study is available in Garcia, Renault and Veredas (2006).

[13] The number of draws H in set to five. Results do not change qualitatively for $H = 1$ and 2. We choose the identity matrix for the weighting matrix W_n as explained in previous section.

[14] 10^{-6} is the same value that Carrasco and Florens (2002) chose for their Monte Carlo study with the α-stable distribution.

[15] We use a GAUSS procedure written by J. Huston McCulloch and available in his web page http://www.econ.ohio-state.edu/jhm/jhm.html

appears that it estimates much better the parameter σ, as CGMM underestimates it. The bias for $\alpha = 1.9$ is quite severe, since the mean of the 500 replications is 0.26 for a true value of 0.5. Estimates for other parameters, for example β, suffer when α gets close to 2, which is never the case for indirect inference. The CII method is also more efficient than CGMM.[16] The empirical quantile method does not seem to suffer from any systematic bias except for β at $\alpha = 1.9$. Its main weakness appears to be its lack of efficiency. Standard deviations are larger than in the case of CII and CGMM.

To judge the efficiency of the indirect inference procedure, we also compare in Table 5 the empirical standard deviations of α and β to the asymptotic Cramer-Rao bounds reported in DuMouchel (1975) for a set of parameter values. It can be seen that CII produces standard deviations that are close to the asymptotic lower bounds. Interestingly, when α is getting closer to 2, the empirical standard deviations produced by the indirect inference procedure are smaller than the asymptotic bounds. This is a finite sample phenomenon. As we increase the number of observations from 1000 to 5000 the indirect inference standard deviation becomes higher than the lower bound while remaining close to it.

4.2 Dependent Processes

Several papers have investigated the relationship between stable processes and processes with conditional heteroskedasticity such as GARCH and IGARCH. De Vries (1991) has shown that under certain conditions on the parameters of a GARCH-like process, the stable and GARCH processes are observationally equivalent from the viewpoint of the unconditional distribution. Ghose and Kroner (1995) establish that many of the properties of stable models are shared by GARCH models. However, they identify distinctive properties, namely the clustering in volatility and the distributions of extreme values, captured by their tail indices. Groenendijk, Lucas and de Vries (1995) show that it is not always the case that the tail shapes can be used to discriminate between the competing models. More recently, Deo (2000, 2002) has devised an estimation procedure for the tail index α as well as a goodness-of-fit test that is valid in the presence of m-dependence in the series.

To illustrate the results put forward in this literature, we generate GARCH(1,1) series for a subset of parameter values chosen by Ghose and Kroner (1995). As before, we simulate 500 samples of 1000 observations from a GARCH(1,1) model with either Gaussian or Student-t_5 that we aggregate every 5 and 20 periods. All processes exhibit empirically excess kurtosis, which increases with the memory of the model and does not vanish under aggregation. Table 6 shows the means and standard deviations for $\hat{\alpha}_n$ and

[16] As noted above, CGMM has been performed with a fixed ad hoc regularization coefficient. An endogenous choice of δ_n could improve results, namely suppress the bias for σ. However, in the Monte Carlo study performed by Carrasco and Florens (2002) for the α-stable distribution, the estimated σ does not change significantly when δ_n is selected in an ad hoc way or endogenously.

$\hat{\sigma}_n$. The stable density captures very well the increases in variance and kurtosis through aggregation and memory. The tail index remains relatively constant under aggregation while the estimated dispersion increases. In all cases standard errors increase with aggregation as the sample size decreases. As expected, the tail index and the dispersion are higher when the process is generated from a Student-t density that when it comes from a Gaussian probability distribution. Last, the higher the memory of the model, in particular for the last three cases, the lower the tail index and the higher the dispersion.

We have illustrated that stable distributions could serve as a good statistical tool to capture the unconditional distribution of asset returns at low frequency, even if the true DGP was a conditionally heteroscedastic process like GARCH. Of course, one might argue that what matters for financial applications is the conditional distribution and that the estimation method developed here will not be useful to capture the conditional dependence in mean and variance. Although we do not intend to show in detail how our method can be extended to characterize conditional distributions, we argue that it should be possible to introduce dependence in our indirect inference procedure: First, it should be stressed that GARCH-like processes, that is processes exhibiting conditional heteroskedasticity, can be built from stable random variables as shown by de Vries (1991). Second, Deo (2002) has used conditionally heteroscedastic processes with marginal stable distribution, which has infinite dependence, to estimate tail indices with a procedure that accommodates dependence of order m. He concludes that his procedure provides quite good estimates of the tail index.

5 An Empirical Illustration

We propose to apply our estimation methodology for the stable distribution to the characterization of the jump distribution in the S&P 500 index returns time series. A series of papers by Barndorff-Nielsen and Shephard (2004a, 2004b, 2006) and Andersen et al. (2003, 2007, 2009), as well as others, have shown that financial time series are best characterized by jump diffusion processes. Of direct interest for our purpose is the fact that the jumps are nearly *i.i.d.*. Jumps display less persistence than volatility and appear to occur at unpredictable times. Moreover, the α-stable density appears as a natural candidate for the distribution of a jump conditional on a jump occurring since Levy processes provide the theoretical foundations to jump-diffusion models.

The literature on realized variance and bi-power variation has provided financial econometricians with measures of both the diffusion and the jump parts in high frequency financial time series. Bandorff-Nielsen and Shephard (2004b) develop a method for detecting the presence of jumps. The basic idea is to compare two measures of variance, realized variance, that includes the contribution of jumps to the total variance, and bi-power variation, that is robust to the contribution of jumps. Realized variance and bi-power variation are defined as follows:

$$RV_i \equiv \sum_{j=1}^{m} r_{i,j}^2$$

$$BV_i \equiv \frac{\pi}{2}\frac{m}{m-1}\sum_{j=2}^{m} |r_{i,j}||r_{i,j-1}|,$$

where $r_{i,j}$ refers to the $j = 1, \ldots, m$ within-day return on day $i = 1, \ldots, n$. Asymptotically, the difference between the realized variance and bi-power variation is zero when there is no jump and strictly positive when there is a jump. This property has given rise to several jump detection techniques in Barndorff-Nielsen and Shephard (2004b), Andersen, Bollerslev and Diebold (2007), and Huang and Tauchen (2005).

Recently, Tauchen and Zhou (2009) extended this literature to identify large jumps in a financial time series. First, they propose to use a ratio statistics identified in this literature $RJ_i \equiv \frac{RV_i - BV_i}{RV_i}$ to construct a test statistics to detect jumps:

$$ZJ_i = \frac{RJ_i}{\sqrt{[(\pi/2)^2 + \pi - 5]m^{-1}\max(1, TP_i/BV_i^2)}},$$

where TP_i is the tri-power quarticity robust to jumps defined in Bandorff-Nielsen and Shephard (2004b). This test statistics converges in distribution to a standard normal distribution.

Then, they assume that there is at most one jump per day and that the jump size dominates the return when a jump occurs. Given a confidence level ι, it is possible to filter out the daily realized jumps as

$$\hat{J}_{n,i} = sign(r_i)\sqrt{(RV_i - BV_i)\mathbf{I}_{ZJ_i \geq \Phi_\iota^{-1}}},$$

where r_i is the daily log return.

We apply their methodology to filter the jumps out of the S&P 500 return series. We collected the 5-minute returns data from 2 January 1996 to 31 August 2006 from Tickdata.com and compute the realized variance and bi-power variation statistics leading to the test statistic separating the days in the sample between jump days and non-jump days. We then estimate a stable distribution for the $\hat{J}_{n,i}$ with CII, as well as with empirical quantiles for comparison purposes.[17]

Table 7 shows the estimation results of the α-stable and the skewed-t densities on $\hat{J}_{n,i}$. For the constrained indirect inference procedure we used $H = 5$ as in the reported Monte Carlo results. Standard errors are computed using both the asymptotic method

[17]We experienced some difficulties with the CGMM method using the fixed beta value reported in the Monte Carlo study. We decided not to pursue further since it would have involved a more elaborate data-dependent method for deriving the optimal regularization parameter.

described in GMR and a nonparametric bootstrap method. For the asymptotic method, we compute the cross-derivatives of the log-likelihood function with respect to the structural and auxiliary parameters by simulation. For the bootstrap, we generate a number of replications by drawing at random with replacement in the sample of data seen as the population.[18] We estimate the model for each bootstrap sample and obtain a measure of the standard error in the generated distribution for each parameter. Given that for each bootstrap sample our estimation procedure involves simulations, we limited our number of replications to 100.

To filter the jumps we used two different significance levels, 0.99 and 0.999, as in Tauchen and Zhou (2009). With the value of 0.99 we are detecting more jumps (828 out of 2686 days) but of smaller size, while for 0.999 jumps are less (538) but larger in size. The corresponding percentages of days with jumps are 30% and 20% respectively. It means that the distribution of daily returns for 0.999 has thicker tails, which is reflected in a smaller α (1.684 for 0.999 versus 1.722 for 0.99). The jump contribution to total variance is in the order of 10%, a bit higher than in Tauchen and Zhou (2009) who have a longer sample.[19] In Tauchen and Zhou (2009), it is shown that for a small jump contribution it is better to select the more generous test level of 0.99.

The estimates for the parameter α are smaller with empirical quantiles, implying thicker tails. The larger standard error for the empirical quantile method confirms its relative lack of efficiency. The estimate for the symmetry parameter β is statistically indistinguishable from zero for both the CII and the empirical quantile estimators. Therefore one can conclude safely that the distribution of jumps does not show any asymmetry. Finally, the estimated mean is close to 0.1 and the standard deviation is around 0.3. This compares to empirical estimates of 0.05 and 0.53 respectively in Tauchen and Zhou (2009), who impose a simple model of Poisson-mixing-Normal jump specification. The standard errors of the estimates are roughly of the same magnitude than the ones in Tauchen and Zhou (2009).

As a diagnostic check we compare by QQ plots in Figure 2 the quantiles of the estimated stable distribution with the Gaussian quantiles for the two confidence levels of the jump detection test. The left plot is for the confidence level 0.999 and the right plot for 0.99. The stable quantiles are evaluated at the estimated parameters while the Gaussian quantiles are evaluated on the sample mean and variance. Previous estimation results indicate that days with jumps are far from being Gaussian and can be better represented by a stable distribution. Indeed, the plots show clear differences between the two distributions, especially in the tails. Stable quantiles away from the median are larger than the corresponding Gaussian quantiles, indicating the presence of thick tails. Moreover, the differences in the tail behaviour are more substantial for the confidence

[18] An alternative, to account for potential serial dependence, is to do block nonparametric bootstrap with replacement. We also computed standard errors with this method for a block of size 10, but the results were very similar and are not reported for space considerations.

[19] Their sample covers the period 1986-2005 for a total of 4752 days.

level 0.999, which is natural since the jumps for this level are larger than in the 0.99 level case.

6 Conclusion

The stable distribution is particularly useful to model processes with heavy-tailed and skewed distributions which are often encountered in financial series. However, its estimation raises several challenges that we addressed in this paper. Since the density function of a stable distribution does not have a closed form but a stable series is relatively easy to simulate, we proposed an indirect inference estimation method, which is ideally suited to such characteristics. In a Monte Carlo study, we showed that the method performed well and better than competing methods in terms of efficiency. To improve the properties of the estimator in finite samples when the value of the stability parameter approaches two, we used a variant of the indirect inference method called constrained indirect inference. We also showed that this new method for estimating stable distributions proved very useful for capturing the skewness and kurtosis present in daily returns with jumps.

Although we do not consider testing in this paper, indirect inference provides as a by-product specification tests about the matched characteristics, in our case the unconditional distribution. One can envision a battery of diagnostic tools. For example, the fact that the binding function can be interpreted parameter by parameter allows independent assessments of the ability of the stable model to capture the four relevant features of the data. One can also perform an omnibus test, by matching jointly McCulloch quantile-based functions and our skewed-t auxiliary parameters to obtain an automatic overidentification test. These tests could complement the work by Deo (2000) who proposed, in the context of m-dependent sequences, a goodness-of-fit test for stable distributions. This is left for future research.

Appendix

Proof of Proposition 3.1:

Pseudo-true values are defined as maximizing the expectation of the log quasi-likelihood function, that is:

$$L\left[(\nu,\gamma,\lambda,\omega)\,|\,(\alpha,\beta,\sigma,\mu)\right] = E\left[L_1\left(Y;\nu,\gamma,\lambda,\omega\right)\right]$$

where $Y \rightsquigarrow S(\alpha,\beta,\sigma,\mu)$. First note that when $Y \rightsquigarrow S(\alpha,\beta,\sigma,\mu)$ we have for all $a \in \mathbb{R}$:

$$Y + a \rightsquigarrow S(\alpha,\beta,\sigma,\mu + a)$$

and

$$L_1(y+a;\nu,\gamma,\lambda,\omega+a) = L_1(y;\nu,\gamma,\lambda,\omega),$$

while for all $a \in \mathbb{R}_+^*$:

$$aY \rightsquigarrow S\left(\alpha,\beta,a\sigma,a\mu\right)$$

and

$$L_1(ay;\nu,\gamma,a\lambda,a\omega) = L_1(y;\nu,\gamma,\lambda,\omega).$$

Thus, for all $a \in \mathbb{R}$:

$$L\left[(\nu,\gamma,\lambda,\omega+a) \,|\, (\alpha,\beta,\sigma,\mu+a)\right] = L\left[(\nu,\gamma,\lambda,\omega) \,|\, (\alpha,\beta,\sigma,\mu)\right]$$

and for all $a \in \mathbb{R}_+*$:

$$L\left[(\nu,\gamma,a\lambda,a\omega) \,|\, (\alpha,\beta,a\sigma,a\mu)\right] = L\left[(\nu,\gamma,\lambda,\omega) \,|\, (\alpha,\beta,\sigma,\mu)\right].$$

Then, considering first $a = -\mu$ and, second, $a = 1/\sigma$, we get:

$$\begin{aligned} L\left[(\nu,\gamma,\lambda,\omega) \,|\, (\alpha,\beta,\sigma,\mu)\right] &= L\left[(\nu,\gamma,\lambda,\omega-\mu) \,|\, (\alpha,\beta,\sigma,0)\right] \\ &= L\left[\left(\nu,\gamma,\frac{\lambda}{\sigma},\frac{\omega}{\sigma}\right) \Big| \left(\alpha,\beta,1,\frac{\mu}{\sigma}\right)\right] \\ &= L\left[\left(\nu,\gamma,\frac{\lambda}{\sigma},\frac{\omega-\mu}{\sigma}\right) |(\alpha,\beta,1,0)\right]. \end{aligned}$$

We deduce straightforwardly from the above expressions that:

$$\begin{aligned} \omega(\alpha,\beta,\sigma,\mu) &= \mu + \omega(\alpha,\beta,\sigma,0) \\ \lambda(\alpha,\beta,\sigma,\mu) &= \sigma\lambda\left(\alpha,\beta,1,\frac{\mu}{\sigma}\right) \\ \gamma(\alpha,\beta,\sigma,\mu) &= \gamma\left(\alpha,\beta,1,0\right) \\ \nu(\alpha,\beta,\sigma,\mu) &= \nu\left(\alpha,\beta,1,0\right). \end{aligned}$$

Finally, note that when $Y \rightsquigarrow S\left(\alpha,\beta,1,0\right)$, $(-Y) \rightsquigarrow S\left(\alpha,-\beta,1,0\right)$ and for all y:

$$L_1(-y;\nu,\frac{1}{\gamma},\lambda,-\omega) = L_1(y;\nu,\gamma,\lambda,\omega).$$

Therefore

$$\begin{aligned} \gamma(\alpha,-\beta,1,0) &= \frac{1}{\gamma(\alpha,\beta,1,0)} \text{ and} \\ \nu(\alpha,-\beta,1,0) &= \nu\left(\alpha,\beta,1,0\right). \end{aligned}$$

This completes the proof of Proposition 3.1.

Proof of Proposition 3.2:

By proposition 3.1

$$\gamma(\alpha,\beta,\sigma,\mu)=\frac{1}{\gamma(\alpha,0,\sigma,\mu)}.$$

And thus $\gamma(\alpha,0,\sigma,\mu)=1$. Moreover, when $\gamma=1$, the asymptotic first order condition for ω is:

$$0=E\left(\frac{Y-\omega}{1+\frac{(Y-\omega)^2}{\nu\lambda^2}}\right).$$

$\omega=\mu$ is clearly a solution of the equation since, when $\beta=0$, $Y-\mu$ and $(Y-\mu)\left(1+\frac{(Y-\mu)^2}{\nu\lambda^2}\right)^{-1}$ are symmetrically distributed around 0.

Proof of Proposition 3.3:

The partial derivatives of the expected quasi loglikelihood function with respect to λ and ν are

$$\frac{\partial L}{\partial\lambda}=-\frac{1}{\lambda}-\frac{\nu+1}{2\nu}\left(-\frac{2}{\lambda^3}\right)E\left[\frac{(Y-\omega)^2 g_\omega(Y,\gamma)}{1+\frac{1}{\nu}\left(\frac{Y-\omega}{\lambda}\right)^2 g_w(Y,\gamma)}\right]$$

and

$$\begin{aligned}\frac{\partial L}{\partial\nu} &= \frac{h'(\nu)}{h(\nu)}-\frac{1}{2\nu}-\frac{1}{2}E\left[\log\left[1+\frac{1}{\nu}\left(\frac{Y-\omega}{\lambda}\right)^2 g_w(Y,\gamma)\right]\right]\\ &\quad -\frac{\nu+1}{2\lambda^2}\left(-\frac{1}{\nu^2}\right)E\left[\frac{(Y-\omega)^2 g_\omega(Y,\gamma)}{1+\frac{1}{\nu}\left(\frac{Y-\omega}{\lambda}\right)^2 g_w(Y,\gamma)}\right].\end{aligned}$$

When $Y\rightsquigarrow S(\alpha,0,1,0)$ the pseudo-true value of γ is 1 so that $g_\omega(Y,\gamma)=1$ and the above first order derivatives can be computed at the pseudo-true value $\Psi(\alpha,0,1,0)$ as:

$$\begin{aligned}\frac{\partial L}{\partial\lambda} &= -\frac{1}{\lambda}+\frac{\nu+1}{\lambda}E\left[\frac{\frac{Y^2}{\lambda^2\nu}}{1+\frac{Y^2}{\lambda^2\nu}}\right]\\ \frac{\partial L}{\partial\nu} &= \frac{h'(\nu)}{h(\nu)}-\frac{1}{2\nu}-\frac{1}{2}E\left[\log\left(1+\frac{Y^2}{\lambda^2\nu}\right)\right]+\frac{\nu+1}{2\nu}E\left[\frac{\frac{Y^2}{\lambda^2\nu}}{1+\frac{Y^2}{\lambda^2\nu}}\right].\end{aligned}$$

The first equation $\frac{\partial L}{\partial\lambda}=0$ gives zero since the first result of proposition 3.3 while when plugging in into the second equation $\frac{\partial L}{\partial\nu}=0$ we get the second result of proposition 3.3.

Asymptotic normality of the unconstrained QMLE

For sake of brevity we simplify the proof by considering that the observations Y_i are *i.i.d.* and symmetric. Serial dependence could clearly be accommodated by refereing instead to central limit theorems for mixing sequences. The symmetry assumption allows us, by virtue of proposition 3.2, to simplify the formulas by considering only three free auxiliary parameters (ω, λ, ν) to identify three structural parameters (μ, σ, α) while γ is fixed at its pseudo true value $\gamma^0 = 1$. In the general case, the functions of γ, $g_\omega(Y, \gamma)$ and $(\gamma + \frac{1}{\gamma})$ could easily be incorporated in the QML first order conditions below at the cost of cumbersome algebra.

Thanks to the simplifying constraint $\gamma = 1$, the first order conditions for the QML $(\hat{\omega}_n, \hat{\lambda}_n, \hat{\nu}_n)$ can be written:

$$\frac{\partial L_n}{\partial \omega}(\hat{\omega}_n, \hat{\lambda}_n, \hat{\nu}_n,) = 0 \quad \Leftrightarrow \quad \hat{\omega} = \frac{\sum_{i=1}^n \theta_{i,n} Y_i}{\sum_{i=1}^n \theta_{i,n}} \tag{6.7}$$

$$\text{with} \quad \theta_{i,n} = \left[1 + \frac{1}{\hat{\nu}_n}\left(\frac{Y_i - \hat{\omega}_n}{\hat{\lambda}_n}\right)^2\right]^{-1},$$

$$\frac{\partial L_n}{\partial \lambda}(\hat{\omega}_n, \hat{\lambda}_n, \hat{\nu}_n,) = 0 \Leftrightarrow \hat{\lambda}^2 = \frac{\hat{\nu}_n + 1}{2\hat{\nu}_n}\frac{1}{n}\sum_{\tau=1}^n \theta_{i,n}(Y_i - \hat{\omega}_n)^2, \tag{6.8}$$

and the first order condition

$$\frac{\partial L_n}{\partial \nu}(\hat{\omega}_n, \hat{\lambda}_n, \hat{\nu}_n,) = 0$$

is less tractable since it involves the highly nonlinear function $h(\nu)$. It could however be handled exactly in a similar way to the other first order conditions thanks to local linearization.

By (6.7):

$$\hat{\omega}_n - \omega^0 = \frac{\sum_{i=1}^n \theta_{i,n}(Y_i - \omega^0)}{\sum_{i=1}^n \theta_{i,n}} = \frac{\lambda^0\sqrt{\nu^0}\sum_{i=1}^n \theta_{i,n} Z_i}{\sum_{i=1}^n \theta_{i,n}},$$

where $Z_i = \frac{Y_i - \omega^0}{\lambda^0\sqrt{\nu^0}}$. By the uniform low of large number for the bounded double array $\theta_{i,n}$ we have

$$\frac{1}{n}\sum_{i=1}^n \theta_{i,n} \to^p E\left((1 + Z^2)^{-1}\right).$$

Thus

$$\sqrt{n}(\hat{\omega}_n - \omega^0) = \frac{\lambda^0\sqrt{\nu^0}}{E\left((1 + Z^2)^{-1}\right)}\frac{1}{\sqrt{n}}\sum_{i=1}^n \theta_{i,n} Z_i + o_p(1).$$

Since $\frac{Z_i}{1+Z_i^2}$ is bounded and symmetric around zero (see $\omega^0 = \mu^0$ by proposition 3.2), we deduce that $\sqrt{n}(\hat{\omega}_n - \omega^0)$ is asymptotically normal with asymptotic variance

$$\frac{{\lambda^0}^2\nu^0}{\left((E(1 + Z^2))^{-1}\right)^2} E\left(\frac{Z^2}{(1 + Z^2)^2}\right).$$

It is worth noting that the central limit argument above works because we manipulate sample means of variables $\frac{Z_i}{1+Z_i^2}$ which have finite variances (they are bounded), by contrast with the initial standardized $\alpha-$stable variable Z_i.

Similarly, we deduce from (6.8) that

$$\sqrt{n}(\hat{\lambda}_n^2 - {\lambda^0}^2) = \frac{\nu^0+1}{2\nu^0}\frac{1}{\sqrt{n}}\sum_{i=1}^n \big(\theta_{i,n}(Y_i-\hat{\omega})^2 - c\big) + o_p(1)$$

with

$$c = \frac{2{\lambda^0}^2\nu^0}{\nu^0+1}.$$

By standard uniformity argument and thanks to boundedness we can write, similarly to above,

$$\sqrt{n}(\hat{\lambda}_n^2 - {\lambda^0}^2) = \frac{(\nu^0+1){\lambda^0}^2}{2}\frac{1}{\sqrt{n}}\sum_{i=1}^n \left(\frac{Z_i^2}{1+Z_i^2} - \frac{2}{\nu^0+1}\right) + o_p(1).$$

Since the variables $\frac{Z_i^2}{1+Z_i^2}$ are bounded, the central limit theorem still applies.

References

[1] Ait-Sahalia, Y. and Jacod, J. (2008), "Fisher's Information for Discretely Sampled Lévy Processes", *Econometrica*, 76, 727-761.

[2] Ait-Sahalia, Y. and Jacod, J. (2007), "Volatility Estimators for Discretely Sampled Lévy Processes", *Annals of Statistics*, 35, 355-392.

[3] Andersen, T.G., Bollerslev, T. Diebold, F.X. and Labys, P. (2003), "Modelling and Forecasting Realized Volatility", *Econometrica*, 71, 579-625.

[4] Andersen, T.G., Bollerslev, T. Diebold, F.X (2007), "Roughing it up: including jump components in the measurement, modeling and forecasting or return distribution", *Review of Economics and Statistics*, 89, 701-720.

[5] Andersen, T.G., Bollerslev, T. and Diebold, F.X. (2009), *Handbook of Financial Econometrics.* L.P. Hansen and Y. Ait-Sahalia (eds.) "Parametric and Nonparametric Volatility Measures". North-Holland, Amsterdam. Forthcoming.

[6] Barndorff-Nielsen, O. and Shephard, N. (2004a), "Econometric Analysis of Realised Covariation: High Frequency Based Covariance, Regression and Correlation". *Econometrica*, 72, 885-925.

[7] Barndorff-Nielsen, O. and Shephard, N. (2004b), "Power and Bipower Variation with Stochastic Volatility Jumps". *Journal of Financial Econometrics*, 2, 1-48.

[8] Barndorff-Nielsen, O. and Shephard, N. (2006), "Econometrics of Testing for Jumps in Financial Economics Using Bipower Variation". *Journal of Financial Econometrics*, 4, 1-30.

[9] Bassi, F., Embrechts, P. and Kafetzaki, M. (1998), "Risk Management and Quantile Estimation", in *A Practical Guide to Heavy Tails*, edited by R. Adler *et al*, Birkhuser, Boston, 111-130.

[10] Bawa, V.S., Elton, E.J. and Gruber, M.J. (1979), "Simple Rules for Optimal Portfolio Selection in Stable Paretian Markets", *Journal of Finance*, 34, 1041-1047.

[11] Calzolari, G., Fiorentini, G. and Sentana, E. (2004), "Constrained Indirect Estimation", *Review of Economic Studies* 71(4), 945-973.

[12] Carr, P., Geman, H., Madan, D. B. and Yor, M. (2002), "The Fine Structure of Asset Returns: An Empirical Investigation", *Journal of Business*, 75, 305-332.

[13] Carrasco, M. and Florens, J.P. (2000), "Generalization of GMM to a continuum of moment conditions", *Econometric Theory*, 16, 797-834.

[14] Carrasco, M., Chernov, M, Florens, J.P., and Ghysels, E. (2007), "Efficient Estimation of General Dynamic Models with a continuous of Moment Conditions", *Journal of Econometrics*, 140, 529-573.

[15] Carrasco, M. and Florens, J.P. (2002), "Efficient GMM Estimation Using the Empirical Characteristic Function", GREMAQ-University of Toulouse Working Paper.

[16] Chambers, J.M., Mallows, C.L. and Stuck, B.W. (1976), "A Method for Simulating Stable Random Variables", *Journal of the American Statistical Association*, 71, 340-344. Corrections 82 (1987): 704, 83 (1988): 581.

[17] Coppejans, M. and Gallant, A.R. (2002), "Cross-Validated SNP Density Estimates", *Journal of Econometrics*, 110, 27-65.

[18] Deo, R.S. (2000), "On Estimation and Testing Goodness of Fit for m-dependent Stable sequences", *Journal of Econometrics*, 99, 349-372.

[19] Deo, R.S. (2002), "On testing the adequacy of stable processes under conditional heteroscedasticity", *Journal of Empirical Finance*, 9, 257-270.

[20] de Vries, C.G. (1991), "On the relation between GARCH and stable processes", *Journal of Econometrics*, 48, 313-324.

[21] Dridi, R. Guay, A. and R. Renault E. (2007) "Indirect Inference and Calibration of Dynamic Stochastic General Equilibrium Models", *Journal of Econometrics*, 136, 397-430.

[22] DuMouchel, W.H. (1973), "On the Asymptotic Normality of the Maximum-Likelihood Estimate when Sampling from a Stable Distribution", *Annals of Statistics*, 1, 948-957.

[23] DuMouchel, W.H. (1975), "Stable Distributions in Statistical Inference: 2. Information from Stably Distributed Samples", *Journal of the American Statistical Association,* 70, n^o 350, 386-393.

[24] Fama, E. (1965a), "Behavior of Stock Market Prices", *Journal of Business,* 38, 34-105.

[25] Fama, E. (1965b), "Portfolio Analysis in a Stable Paretian Market", *Management Science,* 11, 404-419.

[26] Fama, E. and Roll, R. (1971), "Parameter estimates for symmetric stable distributions", *Journal of American. Statistical Association,* 66, 331-338.

[27] Feller, W. (1971), *An Introduction to Probability Theory and its Applications*, Vol 2, 2nd edn. New York: John Wiley and Sons.

[28] Fernández, C. and Steel, M. (1998), "On Bayesian Modelling of Fat Tails and Skewness", *Journal of the American Statistical Association*, 93, 359-371.

[29] Feuerverger, A. and McDunnough (1981), "On the efficiency of Empirical Characteristic Function Procedures", *Journal of the Royal Statistical Society, Series B Statistical Methodology*, 43, 20-27.

[30] Fielitz, B.D. and Rozelle, J. (1981), "Method-of-Moments Estimators of Stable Distribution Parameters", *Applied Mathematics and Computation*, 8, 303-320.

[31] Francq, C. and Zakoian, J.M. (2004), "Maximum Likelihood Estimation of Pure GARCH and ARMA-GARCH processes", *Bernoulli*, 10, 605-637.

[32] Gallant, A.R. and Tauchen, G. (1996), "Which moments to match?", *Econometric Theory*, 12, 657-681.

[33] Garcia, R., Renault, E. and Veredas, D. (2006), "Estimation of Stable Distributions by Indirect Inference", CORE DP 2006/112.

[34] Ghose, D. and Kroner, K.F. (1995), "The relationship between GARCH and symmetric stable processes: Finding the source of fat tails in financial data", *Journal of Empirical Finance*, 2, 225-251.

[35] Gouriéroux, C. Monfort, A. and Renault, E. (1993). "Indirect Inference", *Journal of Applied Econometrics*, vol 8, 85-118.

[36] Groenendijk, P.A., Lucas, A. and de Vries, C.G. (1995), "A note on the relationship between GARCH and symmetric stable processes", *Journal of Empirical Finance*, 2, 253-264.

[37] Hansen, B.E. (1994), "Autoregressive Conditional Density Estimation", *International Economic Review*, vol 35, 3, 705-730.

[38] Huang, X. and Tauchen G. (2005), "The Relative Contribution of Jumps to Total Price Variance", *Journal of Financial Econometrics*, vol 3, 4, 456-499.

[39] Lombardi, M. J. and Calzolari, G. (2008), "Estimation of α-Stable Distributions and Processes", *Econometrics Journal*, 11.1, 193-208.

[40] Mandelbrot, B. (1963), "The Variation of Speculative Prices", *Journal of Business*, 36, 394-419.

[41] McCulloch, J.H. (1986), "Simple Consistent Estimators of Stable Distribution Parameters", *Communications in Statistics - Simulation and Computation*, 15, 1109-1136.

[42] Mittnik, S., Paolella, M.S. and Rachev, S.T. (2000), "Diagnosis and treating the fat tails in financial returns data", *Journal of Empirical Finance*, 7, 389-416.

[43] Mittnik, S., Rachev, S.T., Paolella, M.S. (1998), "Stable Paretian Modeling in Finance: Some Empirical and Theoretical Aspects", in *A Practical Guide to Heavy Tails*, edited by R. Adler, R. Feldman and M.S. Taqqu, Birkhauser, Boston, 79-110.

[44] Ortobelli, S., Huber, I., Schwartz, E. (2002), "Portfolio Selection with Stable Distributed Returns", *Mathematical Methods of Operations Research*, 55, 265-300.

[45] Samorodnitsky, G. and Taqqu, M.S. (1994), *Stable Non-Gaussian Random Processes*. New York: Chapman and Hall.

[46] Smith, A. A. (1993). "Estimating Nonlinear Time Series Models Using Simulated Vector Autoregressions", *Journal of Applied Econometrics* 8, S63-S84.

[47] Tauchen, G. and Zhou, H. (2009). "Realized Jumps of Financial Markets", forthcoming in the *Journal of Econometrics*.

[48] Zolotarev, V.M. (1986). *One dimensional Stable Distributions*, Amer. Math. Soc. Transl. of Math. Monographs, vol 65, Amer. Math. Soc., Providence, RI. (Transl of the original 1983 Russian).

Table 1: Relation structural and auxiliary parameters

Characteristic	Structural	Auxiliary
Tail thickness	α	ν
Skewness	β	γ
Scale	σ	λ
Location	μ	ω

Table 2: Simulation results for for the unconstrained model

α		$\hat{\nu}_n$	$\hat{\gamma}_n$	$\hat{\lambda}_n$	$\hat{\omega}_n$	$\hat{\alpha}_n$	$\hat{\beta}_n$	$\hat{\sigma}_n$	$\hat{\mu}_n$
0.3	Mean	0.21	1.01	0.08	0.00	0.28	0.05	0.52	-0.02
	Sd	0.01	0.03	0.01	0.00	0.04	0.09	0.00	0.01
	Skw	0.04	-0.06	0.06	0.14	0.09	0.12	-0.04	-0.10
	Kur	2.87	3.04	3.23	3.04	2.88	3.12	2.97	3.01
1.5	Mean	2.18	1.00	0.58	-0.01	1.47	-0.11	0.50	-0.04
	Sd	0.20	0.04	0.02	0.04	0.04	0.09	0.01	0.04
	Skw	0.58	-0.08	0.15	-0.01	0.03	0.28	0.13	0.12
	Kur	3.81	2.79	2.93	2.89	2.94	3.14	3.45	3.29
1.9	Mean	7.55	1.00	0.64	-0.01	1.82	0.16	0.50	0.11
	Sd	3.32	0.05	0.02	0.06	0.79	0.21	0.01	0.09
	Skw	3.42	0.18	-0.11	-0.02	-3.12	0.14	-0.13	1.32
	Kur	22.0	2.87	3.01	2.87	22.9	2.66	3.48	7.39

First four moments of the estimated parameters of the skewed-t (columns 3-6) and the α-stable distribution (columns 7-10) using unconstrained indirect inference and when the data generating process is $S(\alpha, 0, 0.5, 0)$ for values of α given in the first column (500 samples of 1000 observations). Sd, Skw and Kur stand for standard deviation, skewness and kurtosis.

Table 3: Sensitivity of $\hat{\nu}$ and $\hat{\rho}$ to the sample size when $\alpha = 1.9$

	N	500	1000	5000	10000
$\hat{\nu}_n$	Sd	38.82	15.46	0.83	0.58
	Skw	7.18	3.08	0.92	0.78
	Kur	56.53	16.53	4.47	4.15
$\hat{\rho}_n$	Sd	0.00	0.01	0.01	0.01
	Skw	-0.29	-0.14	-0.23	-0.17
	Kur	2.96	2.86	3.36	3.03

Standard deviations, skewness and kurtosis (denoted by Sd, Skw and Kur respectively) of the estimated ν and ρ using unconstrained and constrained indirect inference respectively. The date generating process is $S(1.9, 0, 0.5, 0)$. We generate 500 samples of $N = 500, 1000, 5000, 10000$ observations.

Table 4: Simulation results for the constrained model

			CII				CGMM				Empirical Quantiles			
α	β		$\hat{\alpha}_n$	$\hat{\beta}_n$	$\hat{\sigma}_n$	$\hat{\mu}_n$	$\hat{\alpha}_n$	$\hat{\beta}_n$	$\hat{\sigma}_n$	$\hat{\mu}_n$	$\hat{\alpha}_n$	$\hat{\beta}_n$	$\hat{\sigma}_n$	$\hat{\mu}_n$
1.5	0	Mean	1.61	-0.00	0.49	0.08	1.50	0.03	0.35	0.00	1.49	0.00	0.49	0.00
		Sd	0.15	0.09	0.02	0.03	0.07	0.15	0.02	0.06	0.08	0.15	0.02	0.07
		Skw	-0.78	0.05	-0.07	0.02	0.08	-0.10	0.14	-0.01	0.26	0.03	0.09	-0.29
		Kur	3.55	2.92	3.19	3.06	2.90	2.94	2.95	3.18	2.95	3.47	2.91	3.59
	0.75	Mean	1.53	0.75	0.50	0.01	1.50	0.75	0.35	0.01	1.51	0.78	0.49	-0.00
		Sd	0.04	0.07	0.01	0.04	0.07	0.12	0.02	0.08	0.11	0.16	0.02	0.10
		Skw	0.23	-0.13	0.04	0.11	0.02	-0.19	0.39	0.65	0.38	-0.23	0.10	0.27
		Kur	3.03	2.94	2.90	4.01	2.71	2.83	3.55	3.84	3.01	3.98	3.01	3.98
1.9	0	Mean	1.91	0.04	0.49	0.01	1.89	0.00	0.26	-0.00	1.88	0.01	0.49	0.00
		Sd	0.02	0.38	0.01	0.03	0.05	0.61	0.02	0.03	0.10	0.39	0.03	0.04
		Skw	-0.03	-0.03	0.01	0.08	-0.46	0.01	-0.01	0.05	-0.58	0.03	0.01	-0.02
		Kur	2.73	3.14	2.88	3.31	3.17	2.08	2.87	3.04	2.32	4.15	2.92	2.99
	0.75	Mean	1.90	0.75	0.49	0.01	1.88	0.63	0.26	-0.00	1.88	0.35	0.49	-0.01
		Sd	0.04	0.12	0.01	0.03	0.05	0.44	0.01	0.03	0.11	0.40	0.03	0.04
		Skw	0.20	-0.54	0.18	0.30	-0.34	-1.52	0.28	0.02	-0.59	0.38	0.16	-0.09
		Kur	3.04	3.17	3.38	3.02	2.89	5.37	3.08	2.99	2.30	2.01	3.09	3.04

First four moments of the estimated parameters of the α-stable distribution when the data generating process is $S(\alpha, \beta, 0.5, 0)$ for values of α and β given in the first and second columns respectively (500 samples of 1000 observations). Columns 4-7 are estimates using constrained indirect inference. Columns 8-11 are estimates using CGMM and columns 12-15 show the estimates using empirical quantiles. Sd, Skw and Kur stand for standard deviation, skewness and kurtosis.

Table 5: Finite Sample vs Asymptotic Standard Deviations

α	β	Sd. Ind. Inf.		Sd. Asympt.	
		$\hat{\alpha}_n$	$\hat{\beta}_n$	$\hat{\alpha}_n$	$\hat{\beta}_n$
1.5	0	0.048	0.096	0.049	0.096
	0.5	0.044	0.097	0.047	0.085
1.9	0	0.026	0.369	0.036	2.287
	0.5	0.015	0.364	0.035	0.268

Comparison of finite sample (500 samples of 1000 observations) indirect inference standard deviations (Sd. Ind. Inf.) with asymptotic deviations corresponding to the Cramer-Rao bounds (Sd. Asympt.) for α and β.

Table 6: Simulation Results when the DGP is a GARCH model.

GARCH model		Distribution		5	20
$\delta_1 = 0.1$, $\delta_2 = 0.8$	$\hat{\alpha}_n$	Normal	1.9132 [0.0348]	1.9256 [0.0424]	1.9072 [0.1095]
	$\hat{\alpha}_n$	t_5	1.7277 [0.0418]	1.8423 [0.0718]	1.7929 [0.0807]
	$\hat{\sigma}_n$	Normal	0.6800 [0.0311]	1.4731 [0.1003]	2.8884 [0.4403]
	$\hat{\sigma}_n$	t_5	0.5618 [0.0269]	1.3437 [0.1136]	2.9488 [0.4071]
$\delta_1 = 0.05$, $\delta_2 = 0.9$	$\hat{\alpha}_n$	Normal	1.9361 [0.0248]	1.9136 [0.0538]	1.7992 [0.1112]
	$\hat{\alpha}_n$	t_5	1.7459 [0.0382]	1.7785 [0.0652]	1.8071 [0.1333]
	$\hat{\sigma}_n$	Normal	0.9621 [0.0427]	2.1098 [0.1439]	4.0430 [0.5512]
	$\hat{\sigma}_n$	t_5	0.8079 [0.0381]	1.9341 [0.1744]	3.9414 [0.6127]
$\delta_1 = 0.05$, $\delta_2 = 0.95$	$\hat{\alpha}_n$	Normal	1.7526 [0.0992]	1.6941 [0.1102]	1.5987 [0.1983]
	$\hat{\alpha}_n$	t_5	1.5882 [0.1260]	1.6085 [0.1721]	1.5942 [0.1630]
	$\hat{\sigma}_n$	Normal	3.5478 [1.3464]	7.7671 [2.9350]	14.790 [3.9942]
	$\hat{\sigma}_n$	t_5	2.6441 [0.6771]	6.0090 [1.7830]	13.028 [3.6017]

Mean and standard deviation (in square brackets) of 500 estimated $\hat{\alpha}_n$'s and $\hat{\sigma}_n$'s (each draw of 1000 observations) when the DGP is a GARCH(1,1):

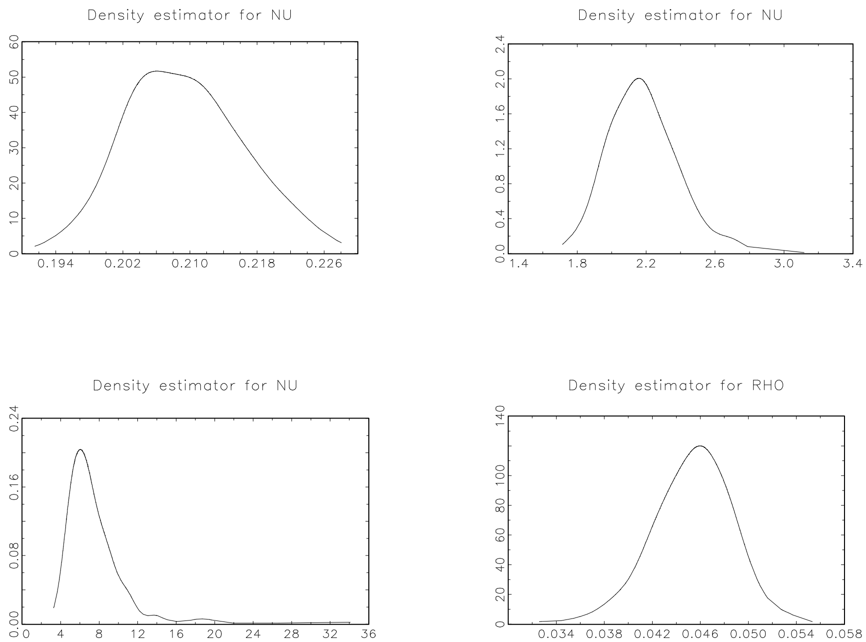
$$\begin{aligned} y_t &= \varepsilon_t, \ \varepsilon_t \sim D(0, h_t) \\ h_t &= \delta_0 + \delta_1 h_{t-1} + \delta_2 \varepsilon_{t-1}^2, \end{aligned}$$

and where D is either Gaussian or Student-t_5. Last two columns show the estimated parameters when the DGP is aggregated over 5 and 20 periods.

Table 7: Estimation results

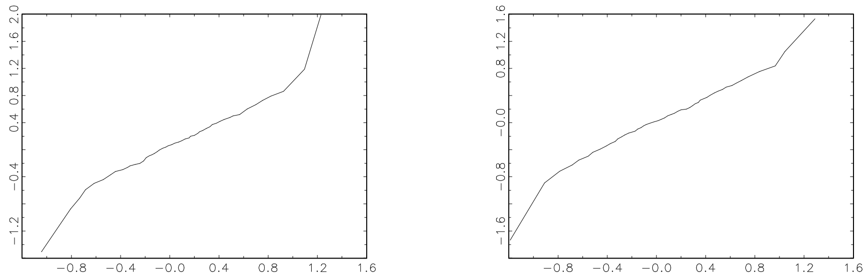
Confidence level	0.999	0.99
Days	2686	2686
Day with jumps	538	828
Jump contribution	9.44%	12%
Jump mean	0.0040	0.0038
Jump Standard Deviation	0.0036	0.0034
Equal variance test	1.12 ($F_{538,828} \simeq 1$)	
Equal mean test	0.599 ($t_{1364} = 1.96$)	
	CII	
$\hat{\alpha}_n$	1.6841 [0.0490,0.0516]	1.7224 [0.0407,0.0452]
$\hat{\beta}_n$	0.0298 [0.1822,0.2772]	-0.170 [0.1681,0.2164]
$\hat{\sigma}_n$	0.2992 [0.0176,0.0228]	0.2966 [0.0137,0.0160]
$\hat{\mu}_n$	0.1067 [0.0355,0.0319]	0.0752 [0.0314,0.0251]
	Empirical Quantiles	
$\hat{\alpha}_n$	1.5631 [0.0885]	1.6188 [0.0796]
$\hat{\beta}_n$	0.0064 [0.1600]	0.0029 [0.1307]
$\hat{\sigma}_n$	0.2909 [0.0202]	0.2906 [0.0129]
$\hat{\mu}_n$	0.0853 [0.0300]	0.0829 [0.0232]
	Skewed-t	
$\hat{\nu}_n$	4.5414 [0.8541]	5.1589 [0.8677]
$\hat{\gamma}_n$	1.0354 [0.0569]	0.9910 [0.0455]
$\hat{\lambda}_n$	0.4075 [0.0277]	0.4032 [0.0200]
$\hat{\omega}_n$	0.0625 [0.0366]	0.0705 [0.0319]

Descriptive statistics (top panel), estimation results of the α-stable distribution with CII and empirical quantiles (middle panels), and estimation results of the skewed-t distributions (bottom panel). Estimation is based on days when a jump has been detected. To detect them a test has been used with two confidence levels 0.999 and 0.99. Standard deviations for CII have been computed with nonparametric bootstrap (left numbers within the squared brackets) and the asymptotically (right numbers). Standard deviations for empirical quantiles have been computed with nonparametric bootstrap.



From top to bottom and from left to right, kernel densities of $\hat{\nu}_n$ when the true α are 0.3, 1.5 and 1.9. The bottom right plot shows the kernel density of $\hat{\rho}_n$ when the true α is 1.9.

Figure 1: Kernel Densities for $\hat{\nu}_n$ and $\hat{\rho}_n$



Stable quantiles (y-axis) against the Gaussian quantiles (x-axis). The left plot is for confidence level 0.999 in the jump detection test and the right plot for confidence level 0.99. The stable quantiles are evaluated at the estimated parameters. The Gaussian quantiles are evaluated at the estimated sample mean and variance.

Figure 2: QQ plots Stable vs. Gaussian