Dipartimento di Statistica
"Giuseppe Parenti"

Indirect estimation of alpha-stable distributions and processes

Marco J. Lombardi,
Giorgio Calzolari

Università degli Studi di Firenze

WORKING PAPER 2004/07

Dipartimento di Statistica "G. Parenti" – Viale Morgagni 59 – 50134 Firenze - www.ds.unifi.it

Statistics, Econometrics

Indirect estimation of α-stable distributions and processes

Marco J. Lombardi* Giorgio Calzolari

Abstract

The α-stable family of distributions constitutes a generalization of the Gaussian distribution, allowing for asymmetry and thicker tails. Its practical usefulness is coupled with a marked theoretical appeal, as it stems from a generalized version of the central limit theorem in which the assumption of the finiteness of the variance is replaced by a less restrictive assumption concerning a somehow regular behavior of the tails. Estimation difficulties have however hindered its diffusion among practitioners.

Since simulated values from α-stable distributions can be straightforwardly obtained, the indirect inference approach could prove useful to overcome these estimation difficulties. In this paper we provide a description of how to implement such a method by using a skew-t distribution as an auxiliary model. The indirect inference approach will be introduced in the setting of the estimation of the distribution parameters and then extended to linear time series models with α-stable disturbances. The performance of this estimation method is then assessed on simulated data. An application on time-series models for the inflation rate concludes the paper.

1 Introduction

The central limit theorem is one of the cornerstones of statistical inference. In the formulation provided by Lindeberg and Lévy, it basically states that, given a sequence of n independent and identically distributed random variables with finite variance, their sum converges, as n grows, to a normal distribution regardless of the individual shape. This is of crucial importance in statistical inference for two basic reasons:

– most of the sample statistics are built by adding up random variables related to the individuals in the sample.

– several phenomena of statistical interest may be thought as aggregations of contributions of smaller factors.

*Corresponding author; email mjl@ds.unifi.it. A preliminary version of this paper was presented at the conference S.Co. 2003 in Treviso. We thank Adelchi Azzalini, Silvano Bordignon and Mauro Grigoletto for their insightful comments.

The consequence of this result is that the normal distribution is quite widespread both in statistical inference and in statistical modelling. As an example, if we hypothesize that the noise term in regression and time series models is the result of a large number of small effects with finite variances, its distribution should be normal. Since it turns out that the estimation residuals are often roughly normal-like, the theoretical property of the normal distribution as a limit law matches with the empirical evidence: these two aspects support and encourage the widespread use of the normal distribution in statistical applications.

However, there are situations in which empirical findings clash with what one would expect provided the theoretical assumptions made. In the specific case, one may observe that in some cases the estimation residuals turn out to have much thicker tails than those expected according to the normal law. This means that one of the two assumptions we made, i.e. that the noise is given by the contribution of a high number of factors and that those factors have finite variance, must be wrong.

1.1 α-Stable Distributions

When the central limit theorem fails because of the non-finiteness of the variance, one should not expect anymore the Gaussian distribution as a limit law. Instead, provided that the following condition concerning the tail behavior

$$\lim_{x\to\infty} \frac{x^2\left[1-F(x)+F(-x)\right]}{u(x)} = \frac{2-\alpha}{\alpha} < \infty, \tag{1}$$

where $u(x)$ is a slowly varying function, holds, one should observe a α-stable limiting distribution. This generalized version of the central limit theorem and the related family of distributions were introduced by Gnedenko & Kolmogorov (1954); the Gaussian distribution is thus a particular case of α-stable distribution. This family of distributions has a very interesting pattern of shapes, allowing for asymmetry and thick tails, that makes them suitable for the modelling of several phenomena; moreover, it is closed under linear combinations.

The family is identified by means of the characteristic function

$$\phi_1(t) = \begin{cases} \exp\left\{\mathrm{i}\delta_1 t - \gamma^\alpha |t|^\alpha \left[1 - \mathrm{i}\beta \operatorname{sgn}(t)\tan\frac{\pi\alpha}{2}\right]\right\} & \text{if } \alpha \neq 1 \\ \exp\left\{\mathrm{i}\delta_1 t - \gamma|t|\left[1 + \mathrm{i}\beta\frac{2}{\pi}\operatorname{sgn}(t)\ln|t|\right]\right\} & \text{if } \alpha = 1 \end{cases} \tag{2}$$

which depends on four parameters: $\alpha \in (0, 2]$, measuring the tail thickness (thicker tails for smaller values of the parameter), $\beta \in [-1, 1]$ determining the degree and sign of asymmetry, $\gamma > 0$ (scale) and $\delta_1 \in \mathbb{R}$ (location).

While the characteristic function (2) has a quite manageable expression and can straightforwardly produce several interesting analytic results, it unfortunately has a major drawback for what concerns estimation and inferential purposes: it is not continuous with respect to the parameters, having a pole at $\alpha = 1$.

An alternative way to write the characteristic function that overcomes this problem, due to Zolotarev (1986), is the following:

$$\phi_0(t) = \begin{cases} \exp\left\{ \mathrm{i}\delta_0 t - \gamma^\alpha |t|^\alpha \left[1 + \mathrm{i}\beta \tan\frac{\pi\alpha}{2} \mathrm{sgn}(t) \left(|\gamma t|^{1-\alpha} - 1\right)\right]\right\} & \text{if } \alpha \neq 1 \\ \exp\left\{ \mathrm{i}\delta_0 t - \gamma |t| \left[1 + \mathrm{i}\beta \frac{2}{\pi} \mathrm{sgn}(t) \ln(\gamma |t|)\right]\right\} & \text{if } \alpha = 1 \end{cases} \tag{3}$$

In this case, the distribution will be denoted as $\mathcal{S}(\alpha, \beta, \gamma, \delta_0)$. The formulation of the characteristic function is, in this case, quite more cumbersome, and the analytic properties have less intuitive meaning; but it is much more useful for what concerns statistical purposes and, unless otherwise stated, we will refer to it in the following. The only parameter that takes needs to be "translated" according to the following relationship is δ:

$$\delta_0 = \begin{cases} \delta_1 + \beta\gamma \tan\frac{\pi\alpha}{2} & \text{if } \alpha \neq 1 \\ \delta_1 + \beta\frac{2}{\pi}\gamma \ln\gamma & \text{if } \alpha = 1 \end{cases} \tag{4}$$

On the basis of the above equations, a $\mathcal{S}_1(\alpha, \beta, 1, 0)$ distribution corresponds to a $\mathcal{S}_0(\alpha, \beta, 1, -\beta\gamma\tan\frac{\pi\alpha}{2})$, provided that $\alpha \neq 1$.

Unfortunately, (2) and (3) cannot be analytically inverted to yield a closed-form density function except for a very few cases: $\alpha = 2$, corresponding to the normal distribution[1], $\alpha = 1$ and $\beta = 0$, yielding the Cauchy distribution, and $\alpha = \frac{1}{2}, \beta = \pm 1$ for the Lévy distribution.

This difficulty, coupled with the fact that moments of order greater than α do not exist whenever $\alpha \neq 2$, has made impossible the use of standard estimation methods such as maximum likelihood and the method of moments. Researchers have thus proposed alternative estimation procedures, mainly based on quantiles (McCulloch 1986) or on the empirical characteristic function (Koutrouvelis 1980), the performance of which is judged unsatisfactory in a number of respects.

With the availability of powerful computing machines, it has become possible to employ computationally-intensive estimation methods for the estimation of α-stable distributions; in particular, likelihood-based inference has been carried out by approximating the density with the FFT of the characteristic function (Mittnik, Doganoglu & Chenyao 1999) or with numerical quadrature (Nolan 1997). However, the accuracy of both these approximations is quite poor for small values of α because of the spikedness of the density function. The latter method, furthermore, is of very difficult implementation. The Bayesian approach has also benefited from the introduction of modern computers: simulation based MCMC methods have been proposed by Buckle (1995), Qiou & Ravishanker (1998), Lombardi (2004) and Casarin (2004).

Despite the computational burden associated with the evaluation of the probability density function, stably distributed pseudo-random numbers can be straightforwardly simulated using the algorithm proposed by Chambers, Mallows & Stuck (1976). Let W be a random variable with exponential distribution of mean 1 and let U be an uniformly distributed random variable on $\left[-\frac{\pi}{2}, \frac{\pi}{2}\right]$. Furthermore, let

[1]Note, though, that in this case β becomes unidentified.

$\zeta = \arctan\left(\beta \tan \frac{\pi\alpha}{2} / \alpha\right)$. Then

$$Z = \begin{cases} \dfrac{\sin \alpha(\zeta + U)}{\sqrt[\alpha]{\cos \alpha\zeta \cos U}} \left[\dfrac{\cos\left(\alpha\zeta + \alpha U - U\right)}{W}\right]^{\frac{1-\alpha}{\alpha}} & \text{if } \alpha \neq 1 \\ \dfrac{2}{\pi}\left[\left(\dfrac{\pi}{2} + \beta U\right)\tan U - \beta \ln \dfrac{\frac{\pi}{2} W \cos U}{\frac{\pi}{2} + \beta U}\right] & \text{if } \alpha = 1 \end{cases} \tag{5}$$

has $\mathcal{S}_0(\alpha, \beta, 1, 0)$ distribution. Random numbers for the general case containing also the position and scale parameters δ and γ may be straightforwardly obtained exploiting the fact that, if $X \sim \mathcal{S}(\alpha, \beta, \gamma, \delta)$, then $Z = \frac{X-\delta}{\gamma} \sim \mathcal{S}(\alpha, \beta, 0, 1)$. Similarly, random numbers from an α-stable distribution expressed in parameterization (2) can be readily obtained using (4).

1.2 α-Stable ARMA Processes

One of the most promising fields of applications of α-stable distributions is that of time series models. As one can in fact note, several empirical phenomena that are observed over time exhibit asymmetry and leptokurtosis (e.g. intensity and duration of rainfalls analyzed in environmetrics, activity time of CPUs and networks or noise in degraded audio samples in engineering, asset returns in finance).

Formally, a process is said to be ARMA (p, q) with α-stable innovations if it takes the form

$$Y_t = \sum_{i=1}^{p} \varphi_i Y_{t-i} + \sum_{j=1}^{q} \psi_j \epsilon_{t-j} + \epsilon_t, \quad \epsilon_t \sim \mathcal{S}_k(\alpha, \beta, \gamma, 0)\ \forall t,\ k = 0, 1, 2. \tag{6}$$

By defining a lag operator L such that $L^q y_t = y_{t-q}$, it is possible to rewrite (6) as

$$\Phi(L) Y_t = \Psi(L) \epsilon_t. \tag{7}$$

Provided that $\Phi(z)$ and $\Psi(z)$ do not have common roots and that the roots of the former are outside the unit circle, the process can be expressed as an infinite moving average:

$$Y_t = \sum_{j=0}^{\infty} c_j \epsilon_{t-j}, \tag{8}$$

where the c_js are the coefficients of the series expansion of $\frac{\Psi(z)}{\Phi(z)}$. From (8), it is straightforward to note that Y_t, being a linear combination of α-stable random variables, is α-stable too with the same characteristic index (Samorodnitsky & Taqqu 1994). It is also immediate to observe that the sequence (8) is strictly stationary; however it is important to remark that, being the variance infinite, the concept of covariance stationarity is meaningless. It can be also demonstrated (Kokoszka & Taqqu 1994) that the c_js decrease at an exponential rate, so that there exists a $M > 1$ such that $|c_j| < M^{-j}$ and the resulting process is short memory.

Since the variance does not exists, however, one cannot use the autocovariance function in order to describe the dependence structure of the process and get insights about an appropriate specification as in the Gaussian case. Methods to identify the appropriate order of AR and MA lags to be used are discussed in Nardelli (1997).

2 Indirect inference for α-stable distributions

The indirect inference (Gouriéroux, Monfort & Renault 1993) is an inferential approach which is suitable for every situation in which the estimation of the statistical model of interest is too difficult to be performed directly while it is straightforward to produce simulated values from the same model. It was first motivated by econometric models with latent variables, but it can be applied in virtually every situation in which the direct maximization of the likelihood function turns out to be difficult.

The underlying principle is very simple: suppose we have a sample of T observations $\mathbf{y}$ and a model whose likelihood function $\mathcal{L}^\star(\mathbf{y};\theta)$ is difficult to handle and maximize; the model could also depend on a matrix of explanatory variables $\mathbf{X}$. The maximum likelihood estimate of $\theta \in \Theta$, given by

$$\hat{\theta} = \max_{\theta\in\Theta} \sum_{t=1}^{T} \ln \mathcal{L}^\star(\theta; y_t),$$

is thus unavailable. Let us now take an alternative model, depending on a parameter vector $\zeta \in Z$, which will be indicated as *auxiliary model*, easier to handle, and suppose we decide to use it in the place of the original one. Since the model is misspecified, the quasi-ML estimator

$$\hat{\zeta} = \max_{\zeta\in Z} \sum_{t=1}^{T} \ln \tilde{\mathcal{L}}(\zeta; y_t),$$

is not necessarily consistent: the idea is to exploit simulations performed under the original model to correct for inconsistency.

The first step consists of computing the quasi maximum likelihood estimate of ζ, which will be denoted as $\hat{\zeta}$. Next, one simulates a set of S vectors of size T from the original model on the basis of an arbitrary parameter vector $\hat{\theta}^{(0)}$. Let us denote each one of those vectors as $\mathbf{y}^s(\hat{\theta}^{(0)})$. The simulated values are then estimated using the auxiliary model, yielding

$$\tilde{\zeta}(\hat{\theta}^{(0)}) = \max_{\zeta\in Z} \sum_{s=1}^{S} \sum_{t=1}^{T} \ln \tilde{\mathcal{L}} \left[\zeta; y_t^s(\hat{\theta}^{(0)})\right]. \tag{9}$$

The idea is to numerically update the initial guess $\hat{\theta}^{(0)}$ in order to minimize the distance

$$\left[\hat{\zeta} - \tilde{\zeta}(\theta)\right]' \Omega \left[\hat{\zeta} - \tilde{\zeta}(\theta)\right], \tag{10}$$

where Ω is a symmetric nonnegative matrix defining the metric.

An alternative but similar approach, introduced by Gallant & Tauchen (1996), considers directly the score function of the auxiliary model:

$$\sum_{t=1}^{T} \frac{\partial \ln \tilde{\mathcal{L}}(\zeta; y_t)}{\partial \zeta}, \tag{11}$$

which is clearly zero for the quasi-maximum likelihood estimator of β. The idea is to make as close as possible to zero the score computed on the simulated observations, namely

$$\min_{\theta} \left\{ \sum_{s=1}^{S} \sum_{t=1}^{T} \frac{\partial \ln \tilde{\mathcal{L}}\left[\zeta; y_t^s(\theta)\right]}{\partial \zeta} \right\}' \Sigma \left\{ \sum_{s=1}^{S} \sum_{t=1}^{T} \frac{\partial \ln \tilde{\mathcal{L}}\left[\zeta; y_t^s(\theta)\right]}{\partial \zeta} \right\}, \tag{12}$$

where Σ is a symmetric nonnegative definite matrix. This approach is especially useful when an analytic expression for the gradient of the auxiliary model is available, since it allows us to avoid the numerical optimization routine for the computation of the $\hat{\zeta}(\theta)$s.

The indirect inference estimators are consistent and asymptotically normal under certain regularitiy conditions. The most difficult one to establish is that the *binding function*, that is the function that maps the parameter space of the auxiliary model onto the parameter space of the true model, is one-to-one. In general, the binding function cannot be expressed analytically and the above condition needs to be verified numerically.

Once one manages to specify an adequate auxiliary model, indirect inference estimators for the parameters of α-stable distributions can be readily implemented and exploited by relying on the simulation algorithm of Chambers et al. (1976).

2.1 The auxiliary model

The auxiliary model we have decided to use is the skew-t distribution recently introduced by Azzalini & Capitanio (2003). The idea follows from an extension of the skew-normal distribution (Azzalini 1985), in which the symmetry of the density function is perturbated by means of the distribution function evaluated at a certain point. More formally, the univariate skew-normal density function is defined as:

$$f(x; \tilde{\beta}, \mu, \sigma) = 2 f_{\mathcal{N}}(z) F_{\mathcal{N}}(\tilde{\beta} z), \tag{13}$$

where $f_{\mathcal{N}}$ and $F_{\mathcal{N}}$ denote, respectively, the density and the distribution function of the standard normal distribution and $z = \frac{x-\mu}{\sigma}$. The parameter[2] $\tilde{\beta} \in \mathbb{R}$ deals with the degree of skewness of the distribution and thus determines the shape of the density function.

[2]In the original papers, $\tilde{\beta}$ is denoted by α; in this work we have adopted this different notation to avoid confusion and mark similarities with the α-stable distribution parameters.

The skewed variant of the t distribution is defined by means of the same perturbation strategy.

$$\begin{aligned} f(x;\nu,\tilde{\beta},\sigma,\mu) &= \frac{2}{\sigma} f_t(z;\nu) F_t\left(\tilde{\beta} z \sqrt{\frac{\nu+1}{z^2+\nu}};\nu+1\right) \\ &= 2\frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\sigma\Gamma\left(\frac{\nu}{2}\right)\sqrt{\pi\nu}}\left[1+\frac{z^2}{\nu}\right]^{-\frac{\nu+1}{2}} F_t\left(\tilde{\beta} z \sqrt{\frac{\nu+1}{z^2+\nu}};\nu+1\right), \end{aligned} \tag{14}$$

where, as before, $z_i = \frac{x_i-\mu}{\sigma}$.

This distribution has four parameters: since it has similarities to a α-stable distribution, given the potential to accommodate asymmetry and heavy tails, it is a good candidate for our purposes. The preferred estimation method for skew-t-based models is maximum likelihood. The log-likelihood function for a skew-t sample of n observations is:

$$\begin{aligned} \ln \mathcal{L}(\nu,\tilde{\beta},\sigma,\mu|\mathbf{x}) &= n\left[\ln\tfrac{2}{\sigma} + \ln\Gamma\left(\tfrac{\nu+1}{2}\right) - \ln\Gamma\left(\tfrac{\nu}{2}\right) - \tfrac{1}{2}\ln(\pi\nu)\right] \\ &\quad + \sum_{i=1}^{n} \ln F_t\left(\tilde{\beta} z_i \sqrt{\frac{\nu+1}{z_i^2+\nu}};\nu+1\right) \\ &\quad - \frac{\nu+1}{2}\sum_{i=1}^{n}\ln\left(1+\frac{z_i^2}{\nu}\right). \end{aligned} \tag{15}$$

The analytic expressions of the first-order derivatives of the log-likelihood function were worked out by Azzalini & Capitanio (2003) and are of great advantage for the implementation of a indirect inference approach, allowing the use of the less computationally-intensive method of Gallant & Tauchen (1996). Since we are dealing with an auxiliary model with one or more constraints, indirect estimation is possible following the method developed by Calzolari, Fiorentini & Sentana (2004). Setting

$$\tau_i = \tilde{\beta} z_i \sqrt{\frac{\nu+1}{z_i^2+\nu}},$$

the analytic gradient is reported below in (16):

$$
\begin{aligned}
\frac{\partial \ln \mathcal{L}}{\partial \nu} &= \frac{n}{2}\left[\Psi\left(\frac{\nu+1}{2}\right) - \Psi\left(\frac{\nu}{2}\right) - \frac{1}{\nu}\right] + \qquad (16)\\
&\quad + \frac{1}{2}\sum_{i=1}^{n}\left[\frac{\nu+1}{\nu^2}\frac{z_i^2}{1+z_i^2/\nu} - \ln\left(1+\frac{z_i^2}{\nu}\right)\right];\\
\frac{\partial \ln \mathcal{L}}{\partial \tilde{\beta}} &= \sum_{i=1}^{n} z_i \frac{f_t(\tau_i;\nu+1)}{F_t(\tau_i;\nu+1)}\sqrt{\frac{\nu+1}{z_i^2+\nu}};\\
\frac{\partial \ln \mathcal{L}}{\partial \sigma} &= -\frac{n}{\sigma} + \sum_{i=1}^{n}\left[\frac{\nu+1}{\sigma\nu}\frac{z_i^2}{1+z_i^2/\nu}\right.\\
&\quad \left. + z_i^2\frac{\tilde{\beta}}{\sigma}\frac{f_t(\tau_i;\nu+1)}{F_t(\tau_i;\nu+1)}\sqrt{\frac{\nu+1}{\left(z_i^2+\nu\right)^3}} - \sqrt{\frac{\nu+1}{z_i^2+\nu}}\right];\\
\frac{\partial \ln \mathcal{L}}{\partial \mu} &= \frac{1}{\sigma}\sum_{i=1}^{n} z_i\left[\frac{\nu+1}{\nu}\left(1+\frac{z_i^2}{\nu}\right)^{-1} + \tilde{\beta} z_i \frac{f_t(\tau_i;\nu+1)}{F_t(\tau_i;\nu+1)}\sqrt{\frac{\nu+1}{\left(z_i^2+\nu\right)^3}}\right] +\\
&\quad - \frac{\tilde{\beta}}{\sigma}\sum_{i=1}^{n}\frac{f_t(\tau_i;\nu+1)}{F_t(\tau_i;\nu+1)}\sqrt{\frac{\nu+1}{z_i^2+\nu}}.
\end{aligned}
$$

2.2 The binding function

The binding function is in general very difficult to be expressed in analytic terms. In order to assess that the estimator is indeed consistent, one must thus rely on graphical information. The most striking difference between skew-t and α-stable distributions is that, for the latter, the asymmetry parameter becomes unidentified as α approaches two; in the sequel we will see that this could be a serious problem. Nevertheless, the binding function seems to generally behave remarkably well, as illustrated in figure 1.

The behavior of the binding function is however less pleasant as α approaches 2, since in such a case β is unidentified. As one can glance from the first graph in figure 1, when α is very close to 2, the binding curves for two very different values of β are nearly indistinguishable. The three-dimensional plot of the binding function displayed in figure 2 highlights this situation: as α approaches 2, the surface gets very steep with respect to $\tilde{\beta}$ and completely flat with respect to β. We will show in what follows that this can be a major source of trouble in the estimation procedure.

2.3 Simulation Results

The simulation study we have conducted to explore the properties of indirect inference estimators yields very promising results; each of the experiments we will present is based on a set of 1000 replications with $S = 10$ and was run on a

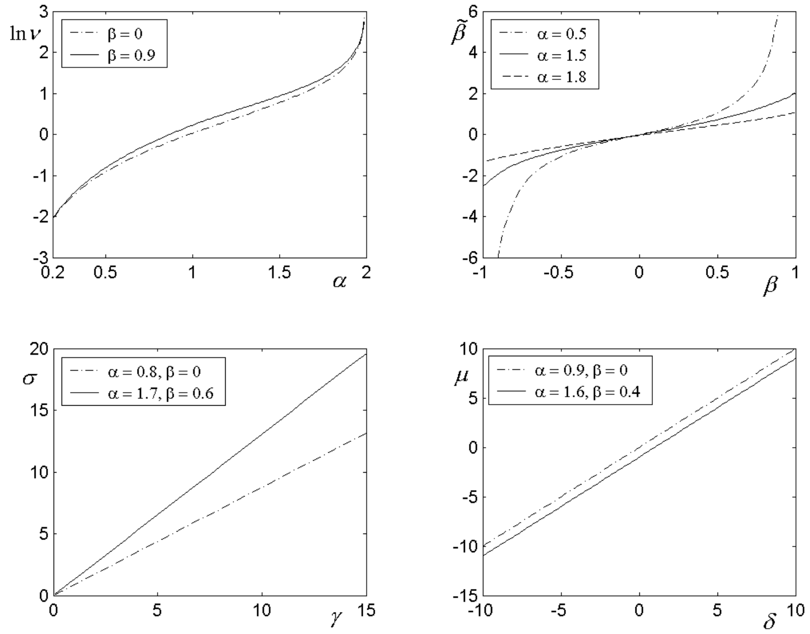

Figure 1: Profiles of the binding function for various parameter values.

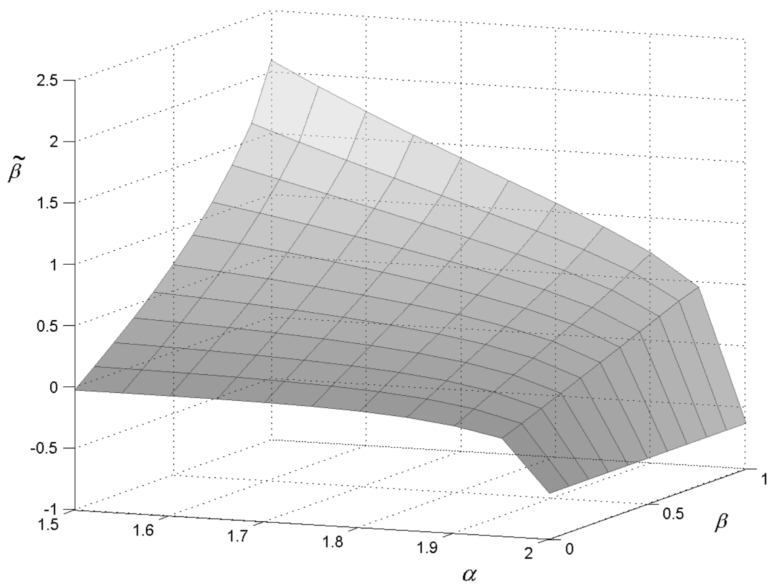

Figure 2: Surface of the binding function with respect to $\tilde{\beta}$ as α and β vary.

2.66GHz Pentium IV processor using a Fortran 77 compiler; 30Mb of RAM were sufficient to handle the largest dimensional case ($N = 3000, S = 10$). The first experiment we have conducted was aimed at assessing the general consistency properties of the indirect inference estimators. Random samples of three different sizes, namely 500, 1000 and 3000, were generated from α-stable distribution with different parameter choices. For this first validation experiment, the starting values supplied to the optimization algorithm were "not too wrong", that is not too far from the actual ones; the effect of the choice of starting values will be examined in one of the following experiments. Results are reported in table 1.

Table 1: Monte Carlo mean and standard error (in parentheses) for various parameter values and sample sizes.

	$\alpha = 1.4$	$\beta = 0$	$\gamma = 1$	$\delta = 0$
$N = 500$	1.4058	0.0025	0.9987	0.0012
	(0.0760)	(0.1268)	(0.0559)	(0.0831)
$N = 1000$	1.4049	0.0023	0.9987	– 0.0011
	(0.0527)	(0.0886)	(0.0388)	(0.0569)
$N = 3000$	1.4014	0.0007	0.9992	0.0003
	(0.0296)	(0.0517)	(0.0222)	(0.0335)
	$\alpha = 1.1$	$\beta = 0.7$	$\gamma = 2$	$\delta = 10$
$N = 500$	1.1044	0.7060	2.0021	10.0053
	(0.0579)	(0.0693)	(0.1224)	(0.1573)
$N = 1000$	1.1028	0.7028	1.9958	10.0001
	(0.0397)	(0.0505)	(0.0866)	(0.1118)
$N = 3000$	1.1010	0.7009	1.9986	10.0011
	(0.0222)	(0.0281)	(0.0491)	(0.0646)
	$\alpha = 0.7$	$\beta = -0.3$	$\gamma = 2$	$\delta = 10$
$N = 500$	0.7035	– 0.2959	1.9989	9.9971
	(0.0360)	(0.0621)	(0.1725)	(0.1166)
$N = 1000$	0.7027	– 0.2996	1.9971	9.9974
	(0.0249)	(0.0438)	(0.1196)	(0.0774)
$N = 3000$	0.7006	– 0.2997	1.9956	10.0002
	(0.0146)	(0.0255)	(0.0725)	(0.0454)

The second experiment we have performed consisted in evaluating whether different choices of the scale and position parameters affect the performance of the estimators for α and β. The results, displayed in table 2, suggest that the estimator are still asymptotically unbiased, but the presence of "low" or "high" values of the scale γ negatively affects the standard error of both γ and δ and has a very mild effect on α and β, whereas different values of δ have apparently no effect.

The estimator provides reliable and consistent results, at least for what concerns values of α and β situated far away from the boundary. Furthermore, the empirical distribution of the estimator behaves remarkably well, as the exemplifi-

Table 2: Monte Carlo mean and standard error (in parentheses) for changing scale and location, $N = 1000$.

	$\alpha = 1.5$	$\beta = 0.5$	Varying γ	$\delta = 10$
$\gamma = 0.5$	1.5037	0.5052	0.4993	10.0000
	(0.0532)	(0.0961)	(0.0180)	(0.0298)
$\gamma = 3$	1.5037	0.5052	2.9955	10.0003
	(0.0532)	(0.0961)	(0.1082)	(0.1789)
$\gamma = 30$	1.5037	0.5052	29.9550	10.0026
	(0.0532)	(0.0960)	(1.0819)	(1.7887)
	$\alpha = 1.5$	$\beta = 0.5$	$\gamma = 3$	Varying δ
$\delta = -5$	1.5036	0.5052	2.9955	– 4.9997
	(0.0532)	(0.0961)	(0.1082)	(0.1789)
$\delta = 0$	1.5037	0.5051	2.9957	0.0005
	(0.0532)	(0.0960)	(0.1081)	(0.1788)
$\delta = 5$	1.5037	0.5052	2.9955	5.0003
	(0.0532)	(0.0961)	(0.1082)	(0.1789)

cation presented in figure 3 reveals.

For what concerns the limiting cases for α and β, the situation is a little bit different and the optimization procedure tends to fail quite often. The solution for perfectly skewed (or apparently symmetric) distributions is to fix β to ± 1 (or to 0). The situation when α is close to 2 is quite different, and is often encountered in practical applications when heavy tailed distributions border normality. In this case, the indirect inference approach tends to fail because, as it can be easily glanced from (2) or (3), β loses relevance and eventually becomes unidentified. This difficulty can be overcome by leaving out β by pre-estimating it[3], possibly with a quantile-based method, or by fixing it to 0 whenever the empirical distribution looks symmetric enough. Although this approach rules out inferential considerations on the asymmetry parameter, the results it provides are quite satisfactory, as displayed in table 3.

Table 3: Monte Carlo mean and standard error for values of α close to 2, $N = 1000$. *Conv.* reports the percentage of replications for which the estimation procedure of the auxiliary model converged.

	α			$\gamma = 1$		$\delta = 0$	
	Conv.	*Mean est.*	*Std. err.*	*Mean est.*	*Std. err.*	*Mean est.*	*Std. err.*
$\alpha = 1.9$	99.8%	1.9026	0.0458	1.0002	0.0298	– 0.0005	0.0499
$\alpha = 1.95$	98.1%	1.9510	0.0357	1.0003	0.0281	– 0.0005	0.0491
$\alpha = 1.99$	68.3%	1.9836	0.0186	0.9962	0.0254	– 0.0003	0.0484

The other problem we have encountered is that the estimation of the auxiliary

[3]This obviously implies pre-testing issues that, at this stage, were not considered.

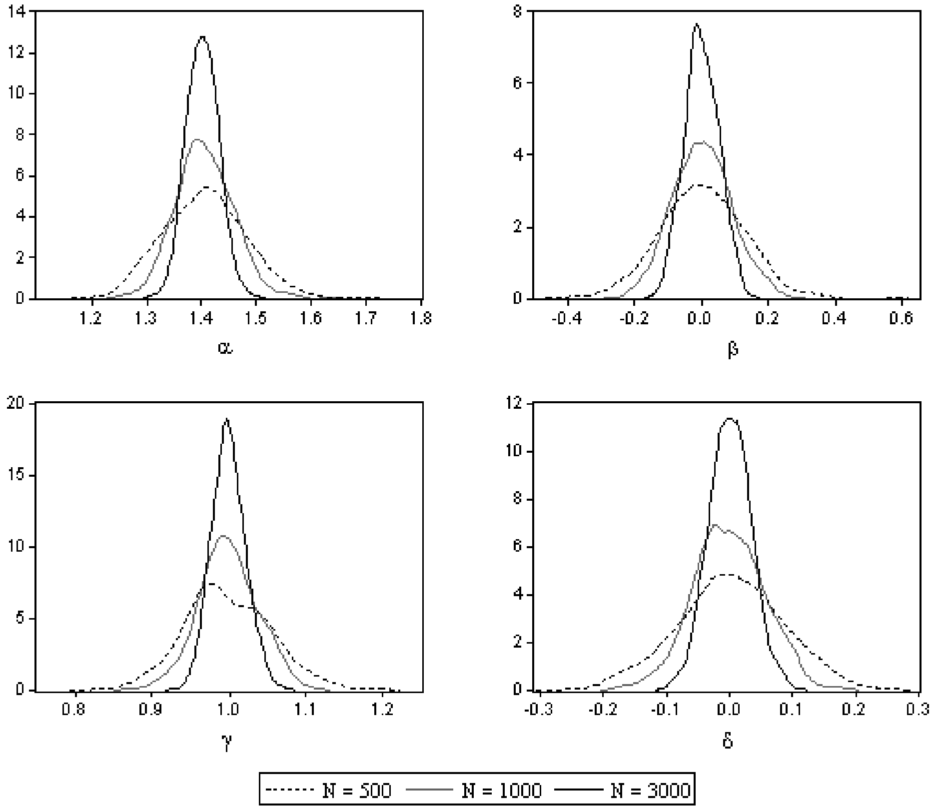

Figure 3: Kernel densities of the parameter estimators, $\alpha = 1.4$, $\beta = 0$, $\gamma = 1$, $\delta = 0$.

model tends to fail[4] as α approaches 2; those cases were thus discarded and the results were computed according to the actual number of replications used. It is worth remarking that the decrease in the standard error of $\hat{\alpha}$ as α approaches 2 is caused by the fact that the asymptotic distribution gets more and more skewed to the left, cutting off the right tail because of the parameter boundary.

In the case presented in table 3, however, β was fixed to its true value. This is obviously not the case in real estimation problems, when one can only guess a value and hope it is close enough to the actual one. Luckily enough, using a guess as greater as 20% than the true value of β and fixing $\hat{\beta}$ to this value seems to have no relevant impact on the standard errors of the other estimates, as shown in table 4.

Table 4: Monte Carlo mean and standard error of parameter estimates when β, whose true value is 0, is fixed to two different values, $N = 1000$.

	$\alpha = 1.8$		$\gamma = 1$		$\delta = 0$	
	Mean est.	*Std. err.*	*Mean est.*	*Std. err.*	*Mean est.*	*Std. err.*
$\beta = 0$	1.8030	0.0538	0.9990	0.0321	– 0.0003	0.0512
$\beta = 0.2$	1.8031	0.0538	0.9997	0.0408	0.0282	0.0512

The last experiment we have performed aims at assessing how the starting values[5] supplied to the optimization algorithm affect the estimates. The parameters of the DGP were set to $\theta = [1.5, 0.5, 1, 0]$. The "wrong" starting values were set to $\hat{\theta}^{(0)} = \hat{\zeta}^{(0)} = [0.6, -0.8, 3, 2.5]$; those "slightly wrong" were $\hat{\theta}^{(0)} = [1.3, 0.8, 1.5, 0.5]$ and $\hat{\zeta}^{(0)} = [2.0, 0.9, 1.5, -0.3]$ and finally, for what concerns the "true" values, besides the obvious choice $\hat{\theta}^{(0)} = [1.5, 0.5, 1, 0]$, we employed $\hat{\zeta}^{(0)} = [2.3, 0.8, 1.3, -0.5]$. Those values were chosen according to the binding function. A quick glance highlights that, apart from the obvious increase in computation time, different starting values do yield completely identical results.

Table 5: Monte Carlo mean and standard error (in parentheses) for different starting values, $N = 1000$. The column "*Time*" reports the average time to convergence in seconds.

	$\alpha = 1.5$	$\beta = 0.5$	$\gamma = 1$	$\delta = 0$	*Time*
True	1.5037	0.5052	0.9985	0.0001	4.3477
	(0.0532)	(0.0960)	(0.0360)	(0.0596)	
Slightly wrong	1.5037	0.5052	0.9985	0.0001	5.6508
	(0.0532)	(0.0960)	(0.0360)	(0.0596)	
Wrong	1.5037	0.5052	0.9985	0.0001	28.6396
	(0.0532)	(0.0960)	(0.0360)	(0.0596)	

[4] In this case the skew-*t* distribution converges to the skew-normal (Azzalini 1985) and thus involves the estimation difficulties associated with this distribution.

[5] Note that, in a indirect inference framework, one has two sets of starting values: those related to the estimation of the auxiliary model, namely $\hat{\zeta}^{(0)}$, and those of the true model, $\hat{\theta}^{(0)}$.

Finally, we have compared the results with those obtained by approximate maximum likelihood. As we have already remarked, the quadrature-based numerical approach of Nolan (1997) is very difficult to implement. Although the author distributes a program to perform basic estimation, its source code was not made public. We have thus confined our attention to the FFT-based approach of Mittnik, Rachev, Doganoglu & Chenyao (1999); the spacing between each point of the grid for the FFT was set to 0.01. Furthermore, for observations lying at a distance greater than 30γ away from δ, we have employed a series expansion in order to avoid having a too large number of points for the FFT. For both the estimation approaches, starting values were set equal to the actual parameter values. Results, displayed in table 6, point out that the indirect inference is only slightly slower with respect to maximum likelihood. One has to keep in mind, however, that the likelihood optimization routine ended up in weak convergence[6] 18% of the times. In table 6 we will thus report both the Monte Carlo results computed on the whole set of simulation and those obtained excluding weak convergences. The mean estimates are quite similar, except for the case of α, whereas for the standard errors a major discrepancy can be highlighted for δ.

Table 6: Monte Carlo mean and standard errors (in parentheses) of the indirect inference and approximate maximum likelihood estimators for various parameter values, $N = 1000$. The column "*Time*" reports the average time to convergence in seconds.

	$\alpha = 1.4$	$\beta = 0$	$\gamma = 1$	$\delta = 0$	*Time*
Ind. inf.	1.4049	0.0023	0.9987	– 0.0011	6.4339
	(0.0527)	(0.0886)	(0.0388)	(0.0569)	
ML, no weak	1.4012	0.0004	0.9959	– 0.0022	4.4421
	(0.0489)	(0.0882)	(0.0251)	(0.0554)	
ML, complete	1.3752	– 0.0016	0.9936	– 0.0030	4.6426
	(0.0499)	(0.0896)	(0.0253)	(0.1167)	

3 Indirect inference for α-stable ARMA processes

The main selling point of this computationally-intensive approach is that, contrary to what happens for the other estimation methods, it is very flexible and can be embedded in a variety of structures, provided one can identify a well-behaved skew-t based auxiliary model. In linear regression models, this carries out straightforwardly: if one wishes to estimate a linear regression model the error term of which has an α-stable distribution, it is sufficient to use the analog model with skew-t error distribution.

[6]For weak convergence we mean that the linear search procedure cannot find a better value along the direction indicated by the gradient.

The issue is a little bit more complex for ARMA time series models, which we will consider in what follows. The idea one could pursue is to use as auxiliary model the skew-t analog of the "true" model of interest, e.g. for an α-stable ARMA(1,1) an auxiliary skew-t ARMA(1,1) model. As far as simple AR models are concerned, this carries out straightforwardly and a just-identified approach performs well. Unfortunately, the analytic derivatives of the MA terms of the auxiliary model cannot be obtained by analytic means; the use of the analog skew-t model thus leads to computational slowness. One could thus use, as an auxiliary model, a simple AR structure, e.g. for an α-stable MA(1) an auxiliary skew-t AR(1) model, for which the analytic gradient is available. In a general MA(q) framework, as long as the roots of the polynomial $1 + \sum_{k=1}^{q} \psi_k z^k$ are outside the unit circle, the MA model is invertible and can be expressed as an AR(∞), making thus possible to establish a correspondence between the true and the auxiliary model.

3.1 Simulation results

The first simulations we have performed concern the estimation of simple AR(1) and MA(1) with α-stable noise models by means of an auxiliary skew-t AR(1) model. Results are based on a set of 1000 independent replications, each one consisting of 1000 observations, and are reported in table 7.

Table 7: Monte Carlo mean and standard error for the estimation of an α-stable AR(1) model with skew-t AR(1) and MA(1) auxiliary.

			AR(1)		
	$\alpha = 1.5$	$\beta = 0.5$	$\gamma = 2$	$\delta = 0$	$\varphi = 0.5$
Mean est.	1.5054	0.5077	1.9955	0.0016	0.4994
Std. err.	0.0553	0.0999	0.0727	0.1221	0.0121
	$\alpha = 1.7$	$\beta = -0.2$	$\gamma = 1$	$\delta = 0$	$\varphi = -0.8$
Mean est.	1.7036	– 0.2003	0.9982	– 0.0014	– 0.7994
Std. err.	0.0572	0.1473	0.0337	0.0609	0.0128
			MA(1)		
	$\alpha = 1.5$	$\beta = 0.5$	$\gamma = 2$	$\delta = 0$	$\psi = 0.4$
Mean est.	1.5036	0.5065	1.9920	0.0020	0.3988
Std. err.	0.0600	0.1039	0.0896	0.1294	0.0455
	$\alpha = 1.8$	$\beta = -0.4$	$\gamma = 1$	$\delta = 0.5$	$\psi = -0.5$
Mean est.	1.8034	– 0.4321	0.9951	0.5148	– 0.5080
Std. err.	0.0608	0.2324	0.0416	0.1017	0.0612

This approach performs satisfactorily as long as the model of interest does not combine AR and MA terms. In this latter case, unfortunately, the binding function is no longer one-to-one; if one naively tries to use the indirect inference anyway, e.g. tries to estimate an α-stable ARMA(1,1) with a just-identified skew-t AR(2) as an auxiliary model, (s)he would face a bimodal distribution for both the AR and the MA parameters.

A possible approach to overcome this difficulty is to increase the AR order of the auxiliary model. In the experiments we have performed it appears that, for

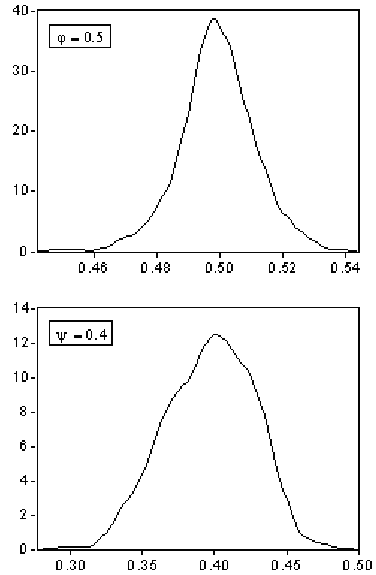

Figure 4: Kernel densities of the AR and MA parameters estimators of table 7.

what concerns an ARMA(1,1) model, an AR(4) auxiliary structure is sufficient (cf. figure 5) to get a well-behaved binding function (cf. Di Iorio & Calzolari 2004). As a rule of thumb, one could thus suggest to double the number of AR coefficients. This finding, however, deserves further attention.

The results of the simulation experiment we have performed in this framework are presented in table 8 and are based on a set of 500 replications. We can observe that the performance of the method is acceptable for all the parameters but ψ. In this case, the estimator displays a strong bias associated with a markedly high standard deviation. This could be a signal that, albeit the binding function is well-behaved, the AR order of the auxiliary model is not large enough. Using more AR lags, however, makes the procedure much slower and unsuitable for a detailed simulation study. For this reason, we do not report any result concerning this issue.

Table 8: Monte Carlo mean and standard error for the estimation of an α-stable ARMA(1,1) model with skew-t AR(4) auxiliary.

	$\alpha = 1.5$	$\beta = 0.3$	$\gamma = 1$	$\delta = 0$	$\varphi = 0.7$	$\psi = 0.1$
Mean est.	1.5527	0.3116	0.9981	0.0086	0.7126	0.0453
Std. err.	0.0691	0.1168	0.0375	0.0634	0.0422	0.1475

3.2 An empirical application

As an illustration on how our model performs on real data, we will deal with the estimation of a simple AR(1) model for the inflation rate of the consumer price

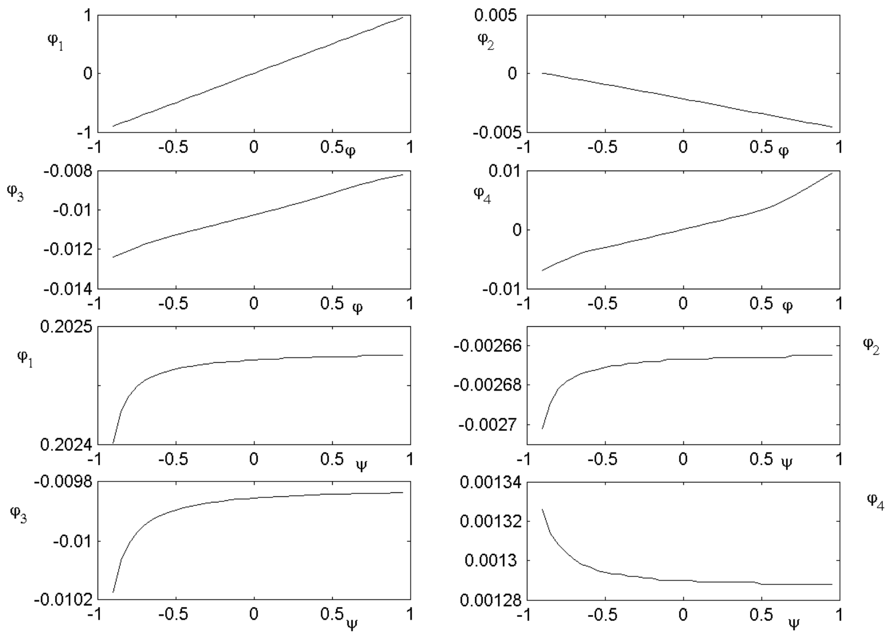

Figure 5: Various profiles of the binding function for an ARMA(1,1) model with an AR(4) auxiliary model. The parameters of the underlying α-stable noise are $\alpha = 1.5$, $\beta = 0.5$, $\gamma = 1$, $\delta = 0$. The first row reports the binding function with respect to the AR parameter φ with the MA parameter $\psi = 0.2$, the second is respect to the MA parameter with $\varphi = 0.2$.

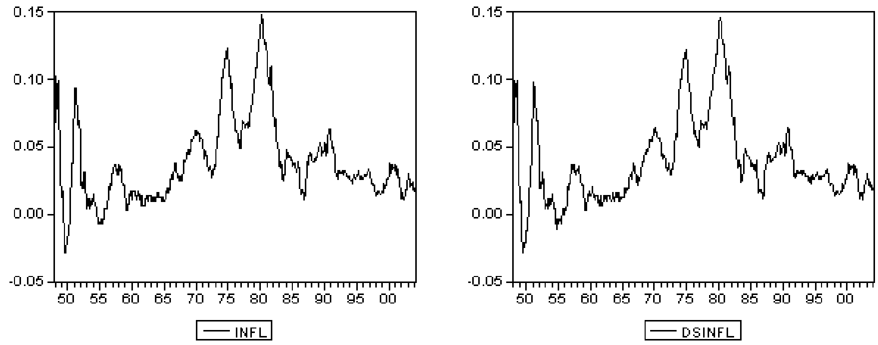

Figure 6: Monthly data on raw (left) and seasonally adjusted (right) yearly inflation rates in the US, January 1948 – March 2004.

index in the United States. The data set we will consider is composed of 675 monthly observations on yearly inflation rates and ranges from January 1948 to March 2004 (cf. figure 6). We will employ both the raw and the seasonally adjusted versions of the consumer price index, obtained from the Federal Reserve of St. Louis. Although not too far from the normal, the estimation residuals of a Gaussian AR(1) model display an excess of kurtosis.

Estimation results for both the Gaussian and the α-stable AR(1) models are displayed in table 9. As one could expect, the standard deviation of the Gaussian model is much larger than the scale parameter of the α-stable model, due to the heavy tailed features of the noise. For what concerns the AR parameter φ, the α-stable model indicates much less persistence than the Gaussian. Again, this is a direct consequence of the heavy tailed features of the noise: large deviations of the inflation rate away from its expected value are much likely to be interpreted as noise than in the Gaussian case. As a result, the fitted series is notably smoother, as displayed in figure 7.

Table 9: Estimates and standard errors (in parentheses) for the parameters of a Gaussian and an α-stable AR(1) models for the inflation rate.

	α-Stable					Gaussian		
	α	β	γ	δ	φ	σ	μ	φ
Raw	1.5713	– 0.4309	0.2097	0.0321	0.7596	0.4577	0.0304	0.9876
	(0.0507)	(0.1370)	(0.0106)	(0.0507)	(0.0248)	(0.0221)	(0.0147)	(0.0057)
S.adj.	1.5520	– 0.2586	0.2055	0.0392	0.8171	0.4411	0.0290	0.9871
	(0.0490)	(0.0989)	(0.0119)	(0.0309)	(0.0444)	(0.0215)	(0.0138)	(0.0055)

4 Conclusions

We have introduced a novel indirect inference approach to the estimation of the α-stable distributions parameters that makes use of a skew-t distribution as auxil-

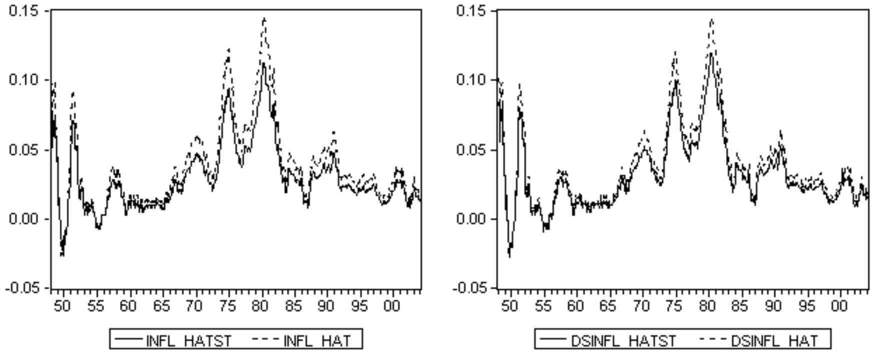

Figure 7: Fitted series for the Gaussian (dotted lines) and the α-stable (solid lines) AR(1) models for both the raw (left panel) and the seasonally adjusted (right panel) inflation series.

iary model. This approach was shown to perform satisfactorily on simulated data, although we have to note that some adjustments to the standard indirect inference scheme are required when the parameters α and β are close to their boundaries. We further remarked that its computational requirements are competitive with respect to the approximate maximum likelihood.

We have then extended this approach to basic ARMA time series models, highlighting that AR-based auxiliary models perform well whenever AR and MA terms are not combined. In the case of more complex ARMA models, one can use as auxiliary model an over-identified AR structure. As a practical example, the indirect inference approach was then used to estimate an α-stable AR(1) model for the US inflation rate.

This approach seems very promising, especially because it can be easily extended to various kinds of statistical models. In particular, an extension that could be very interesting in the setting of time series models is to consider long memory ARFIMA models (Hosking 1981), whose estimation could be carried out using an AR auxiliary model.

References

Azzalini, A. (1985), 'A class of distributions that include the normal one', *Scandinavian Journal of Statistics* **12**, 171–178.

Azzalini, A. & Capitanio, A. (2003), 'Distributions generated by perturbation of symmetry with emphasis on a multivariate skew *t* distribution', *Journal of the Royal Statistical Society B* **65**, 367–389.

Buckle, D. (1995), 'Bayesian inference for stable distributions', *Journal of the American Statistical Association* **90**, 605–613.

Calzolari, G., Fiorentini, G. & Sentana, E. (2004), 'Constrained indirect estimation', *Review of Economic Studies, forthcoming* .

Casarin, R. (2004), Bayesian inference for mixture of stable distributions. Cahier du CEREMADE 0415.

Chambers, J., Mallows, C. & Stuck, B. (1976), 'A method for simulating stable random variables', *Journal of the American Statistical Association* **71**, 340–344.

Di Iorio, F. & Calzolari, G. (2004), Indirect estimation of ARFIMA models: Computational performance on medium-large samples. Dipartimento di Statistica, Università degli studi di Napoli Federico II, mimeo.

Gallant, A. & Tauchen, G. (1996), 'Which moments to match?', *Econometric Theory* **12**, 657–681.

Gnedenko, B. & Kolmogorov, A. (1954), *Limit Distributions for Sums of Independent Random Variables*, Addison-Wesley, Reading.

Gouriéroux, C., Monfort, A. & Renault, E. (1993), 'Indirect inference', *Journal of Applied Econometrics* **8**, S85–S118.

Hosking, J. (1981), 'Fractional differencing', *Biometrika* **68**, 165–176.

Kokoszka, P. & Taqqu, M. (1994), 'Infinite variance stable ARMA processes', *Journal of Time Series Analysis* **15**, 203–220.

Koutrouvelis, I. (1980), 'Regression-type estimation of the parameters of stable laws', *Journal of the American Statistical Association* **75**, 919–928.

Lombardi, M. (2004), Bayesian inference for α-stable distributions: A random walk MCMC approach, *in* 'Atti della XLII riunione scientifica', Vol. Sessioni spontanee, Società Italiana di Statistica, CLEUP, Padova, pp. 299–302.

McCulloch, J. (1986), 'Simple consistent estimators of stable distribution parameters', *Communications in Statistics – Simulation and Computation* **15**, 1109–1136.

Mittnik, S., Doganoglu, T. & Chenyao, D. (1999), 'Computing the probability density function of the stable paretian distribution', *Mathematical and Computer Modelling* **29**, 235–240.

Mittnik, S., Rachev, S., Doganoglu, T. & Chenyao, D. (1999), 'Maximum likelihood estimation of stable paretian models', *Mathematical and Computer Modelling* **29**, 275–293.

Nardelli, S. (1997), Time series models with stable innovations, Master's thesis, European University Institute, Fiesole.

Nolan, J. (1997), 'Numerical computation of stable densities and distribution functions', *Communications in Statistics – Stochastic Models* **13**, 759–774.

Qiou, Z. & Ravishanker, N. (1998), 'Bayesian inference for time series with stable innovations', *Journal of Time Series Analysis* **19**, 235–249.

Samorodnitsky, G. & Taqqu, M. (1994), *Stable Non-Gaussian Random Processes*, Chapman & Hall, Boca Raton.

Zolotarev, V. (1986), *One-dimensional Stable Distributions*, American Mathematical Society, Providence.