AISM (2006) 58: 779–807
DOI 10.1007/s10463-006-0042-z

Anestis Antoniadis · Andrey Feuerverger
Paulo Gonçalves

Wavelet-based estimation for univariate stable laws

Received: 10 November 2003 / Revised: 6 July 2005 / Published online: 11 May 2006

Abstract Stable distributions are characterized by four parameters which can be estimated via a number of methods, and although approximate maximum likelihood estimation techniques have been proposed, they are computationally intensive and difficult to implement. This article describes a fast, wavelet-based, regression-type method for estimating the parameters of a stable distribution. Fourier domain representations, combined with a wavelet multiresolution approach, are shown to be effective and highly efficient tools for inference in stable law families. Our procedures are illustrated and compared with other estimation methods using simulated data, and an application to a real data example is explored. One novel aspect of this work is that here wavelets are being used to solve a parametric problem, rather than a nonparametric one, which is the more typical context in wavelet applications.

Keywords Wavelets · Stable laws · Empirical characteristic function · Estimation · Efficiency · Completeness · Regularization

1 Introduction

In recent years, new classes of functions called wavelets have been discovered which span the usual Hilbert space $L^2(\mathbb{R})$ and which possess local adaptivity,

A. Antoniadis
Laboratoire IMAG-LMC, University Joseph Fourier,
BP 53, 38041 Grenoble Cedex 9, France

A. Feuerverger (✉)
Department of Statistics, University of Toronto,
Toronto, ON, Canada M5S 3G3
E-mail: andrey@utstat.toronto.edu

P. Gonçalves
INRIA Rhône-Alpes, ZIRST, 655 Avenue de l'Europe,
38330 Monbonnot Saint Martin, France

approximation and computational properties which not only are remarkable, but also give rise to applications of wavelet-based methods in many areas of science and engineering. A key characteristic in many such applications is the ability of wavelet functions to capture effectively local features of the processes modelled. Hence, in statistics, for example, wavelets have been used primarily to deal with problems of a nonparametric character, such as those arising in the context of regression analysis, or when estimating functions such as densities, spectral densities, or hazard rates. See, for example, Antoniadis et al. (1994, 1999), Donoho et al. (1996), Gao (1993), Hall and Patil (1995), and Johnstone et al. (1992). This leaves open a question which is natural to ask within a statistical perspective, namely, can these new classes of functions be used to advantage to solve statistical problems which are purely of a *parametric* nature? For example, can wavelet methods be used effectively to obtain estimates of unknown parameters in a parametric statistical model, given data generated according to that model? It is to this question that the present paper is addressed, and we argue the answer is affirmative.

In particular, we will demonstrate the effectiveness of procedures based on wavelet functions to estimate parameters of the stable laws—and specifically on regression methods involving the wavelet transform. The stable laws are a natural family to consider in this context, not only because of their importance in an increasingly wide range of applications in such fields as finance, chaos, hydrology, telecommunications, and physics, but also because the Stable Laws do not have closed forms for their densities so that the classical inference procedures are generally difficult to implement. Thus alternative methods of inference for this family are of particular interest.

We shall show that wavelets provide an effective, natural, and numerically interesting class of techniques for carrying out parametric inference. These techniques are complementary to maximum likelihood and often applicable in situations (e.g. convolution families) where densities may not have a tractable form. These new wavelet-based methods are highly efficient, inherently robust, and in particular, allow one to deal effectively with distributions possessing heavy tails. These wavelet transformation methods also have the property of 'disbalancing' certain correlation structures and rearranging the (Fisher) information of data in ways that can prove helpful in the context of nearly singular statistical models. A key novelty is that wavelets are being used here to solve a *parametric* problem instead of a *nonparametric* one, the latter being more typical in applications of wavelet methods.

This paper is organized as follows. In Sect. 2 we give a brief overview of the Stable Laws as well as some recent methods that have been proposed to estimate parameters in such families. Our main wavelet-based parametric estimation procedures are developed in Sect. 3. In particular, there we explore a multiresolution approach to solve parametric estimation problems using an iterative weighted nonlinear least squares procedure. We adopt this approach specifically because many of the issues arising in the solution of such least squares problems may be handled effectively by multiresolution methods. As with many high-dimensionality regression problems, ill-conditioning of the resulting equations needs to be overcome by regularization.

Implementation of such methods involve substantive numerical considerations; these aspects are discussed in Sect. 4 which deals with the algorithmic details of our

procedures. In Sect. 5 we summarize the results of numerical experimentation and report on a small simulation study intended to assess the performance of the methods in finite sample situations. We also provide an example of the application of our procedure to a real data example arising from the financial markets. Finally some further discussion, indications of possible extensions, and concluding remarks are collected in Sect. 6.

2 The stable laws

The univariate stable laws comprise the class of limiting distributions for normalized sums of independent and identically distributed random variables. The members of this class are characterized by four parameters: an index or shape parameter α, a skewness parameter β, a scale parameter σ, and a shift (or location) parameter μ, where $0 < \alpha \leq 2$, $-1 \leq \beta \leq 1$, $\sigma > 0$, and $-\infty < \mu < +\infty$. The parameter α measures the degree of peakedness and the heaviness of the tails of the stable distribution; when $\alpha = 2$, it corresponds to a normal distribution with mean μ and variance $2\sigma^2$, while when $\alpha < 2$, the variance becomes infinite and the tails of the density then have order $\mathcal{O}(|x|^{-1-\alpha})$ as $|x| \to \infty$. This characteristic of the stables (together with their 'central limit' origins) makes them useful in modelling certain types of data that admit observations of large magnitude, while the parameter β allows incorporation of skewness. See, for example, Christof and Wolf (1992), Feller (1971), Gnedenko and Kolmogorov (1954), Ibragimov and Linnik (1971), Lukacs (1970), Samorodnitsky and Taqqu (1994), and Zolotarev (1986). A wide range of applications of the stable laws are discussed, for example, in Janicki and Weron (1994), Mandelbrot (1963, 1972), Mittnik and Rachev (1993, 2000), and Uchaikin and Zolotarev (1999).

Except for $(\alpha, \beta) = (2, \cdot)$, $(1, 0)$ or $(1/2, \pm 1)$, the densities of stable random variable are not available in closed form, which makes inference by standard techniques, such as maximum likelihood, difficult. This makes methods based on the characteristic function of stable distributions a natural tool for inference. In fact the stable distributions are given most simply by means of their characteristic function $\phi(t)$, whose logarithm is

$$\log \phi(t) = \begin{cases} \mathrm{i}\mu t - |\sigma t|^{\alpha} \left\{1 - \mathrm{i} \tan\left(\frac{\pi}{2}\alpha\right) \beta \operatorname{sgn}(t)\right\} & \text{when } \alpha \neq 1 \\ \mathrm{i}\mu t - |\sigma t| \left\{1 + \mathrm{i}\frac{2}{\pi} \beta \log |t| \operatorname{sgn}(t)\right\} & \text{when } \alpha = 1 \end{cases}. \tag{1}$$

For statistical applications, the canonical representation (Eq. 1) has the disagreeable feature that $\phi(t)$ is discontinuous in the parameters when α and $\beta \neq 0$ vary continuously across the line $\alpha = 1$. However, this discontinuity is of an inessential kind and should be removed in nonsymmetric cases, as in DuMouchel (1971), Chambers et al. (1976), or Feuerverger and McDunnough (1981b). We will denote an α-stable distribution by $S_\alpha(\sigma, \beta, \mu)$. When $\mu = 0$ and $\sigma = 1$ the distribution is called standard stable.

The probability densities of stable variables exist and are continuous and unimodal, but as already noted they are not available in closed form and this severely hampers the problem of estimating their parameters. In particular, conventional statistical methods (such as maximum likelihood) cannot be used directly. However, several methods for estimating stable parameters have been considered and

found useful. For example, Fama (1965) and Fama and Roll (1968, 1971) suggested estimation by means of sample fractiles for symmetric stable distributions when $1 \leq \alpha \leq 2$; Fama–Roll's method is simple but suffers from a small asymptotic bias in $\hat{\alpha}$ and $\hat{\sigma}$ and from restrictions on α and β. Their procedure was generalized and improved by McCulloch (1986) who provided consistent estimators of all four parameters, for $0.6 \leq \alpha \leq 2$, while retaining the computational simplicity of Fama–Roll's method. As well, sample characteristic function methods were developed by many authors. In particular, Koutrouvelis (1980, 1981) adapted Press's (1972) moment procedure into a two-step regression-type method and developed an iterative weighted regression fitting procedure. Paulson et al. (1975) used an optimization approach, selecting as estimators the values which minimize a weighted second-mean distance between the fitted characteristic function and the empirical one. Hill (1975) suggested an approach to estimation by assuming a parametric form only for the tails of the distribution function. See also Feuerverger and McDunnough (1981a,b) and Kogon and Williams (1998). For further review, see e.g., Mittnik and Rachev (1993).

Using a multinomial approximation to the likelihood, DuMouchel (1971, 1973a,b, 1983) was the first to obtain approximate ML estimates for α and σ (assuming $\mu = 0$); DuMouchel (1973a,b) also proved that the maximum likelihood estimates for the full four-parameter stable law family are asymptotically normal with covariance matrix determined by the inverse of the Fisher information matrix in the usual way provided $\alpha > \epsilon$ for some fixed $\epsilon > 0$. (The stable likelihood function is unbounded near $\alpha = 0$.)

Although DuMouchel also required that the parameter values stay away from $\alpha = 1$, his proof appears to carry over to the continuous reparameterizations referred to above. Maximum likelihood procedures for estimating stable parameters were also implemented by Feuerverger and McDunnough (1981b), McCulloch (1998), Chen (1991), and Nolan (1997, 2001).

To our knowledge, no wavelet-based method for parametric estimation of the stable laws has been explored in the literature, and this is the approach we adopt.

3 Wavelet estimation procedures

The estimation procedures we develop are based on wavelet decompositions. For expositions of mathematical aspects of wavelets we refer, for example, to Daubechies (1992), Mallat (1998), Chui (1992), or Holschneider (1995), while expositions on the use of wavelets in statistical settings are given, for example, in Abramovich et al. (2000), Antoniadis (1997), Ogden (1997), and Vidakovic (1999).

To simplify exposition, we consider only the one-dimensional case and orthonormal wavelet bases of $L^2(\mathbb{R})$ generated by dyadic dilation and translation of a compactly supported scaling function and a compactly supported mother wavelet. To fix context and notation, wavelet analysis thus requires a description of a *scaling function* $\varphi(x)$ and a *wavelet* $\psi(x)$. The function $\varphi(x)$ is a solution of a two-scale difference equation $\varphi(x) = \sqrt{2} \sum_{k \in \mathbb{Z}} h_k \varphi(2x - k)$ with normalization $\int_{\mathbb{R}} \varphi(x)\, \mathrm{d}x = 1$, and the function $\psi(x)$ is then defined by $\psi(x) = \sqrt{2} \sum_{k \in \mathbb{Z}} (-1)^k h_{1-k} \varphi(2x - k)$. Through careful choice of the *filter coefficients* h_k, wavelet functions with desirable properties can be constructed. A wavelet

system then consists of the infinite collections of translated and scaled versions of φ and ψ defined via

$$\varphi_{j,k}(x)=2^{j/2}\varphi(2^j x-k) \quad \text{and} \quad \psi_{j,k}(x)=2^{j/2}\psi(2^j x-k), \quad j,k\in\mathbb{Z}. \tag{2}$$

Additional conditions on the filter coefficients imply that $\{\varphi_{j,k},\ k \in \mathbb{Z}\}$ is an orthonormal system in $L^2(\mathbb{R})$ for each $j \in \mathbb{Z}$, and $\{\psi_{j,k},\ j,k \in \mathbb{Z}\}$ is an orthonormal basis of $L^2(\mathbb{R})$.

The wavelet representation for $g \in L^2(\mathbb{R})$ is then given by $g = \sum_{j\in\mathbb{Z}} \sum_{k\in\mathbb{Z}} d_{j,k}$ $\psi_{j,k}$ where the wavelet coefficients $d_{j,k} = \int_{\mathbb{R}} g(t)\psi_{j,k}(t)\,\mathrm{d}t = \langle g, \psi_{j,k}\rangle$, and $\langle\cdot,\cdot\rangle$ is the usual inner product. For any $f\in L^2(\mathbb{R})$ and $j,k \in \mathbb{Z}$ we will sometimes denote by $\eta_{jk} \equiv \eta_{jk}(f) = \langle f, \psi_{jk}\rangle$ the wavelet coefficients of f, with respect to this orthonormal wavelet basis. Typically, we want algorithms with linear or linear-logarithmic complexity to traverse between a function and its wavelet coefficients. Such *fast wavelet transforms* are often obtained through the multiresolution framework of Mallat (1989). Such algorithms are available in several standard implementations, for example, in the `S-Plus` packages `WaveThresh` (Nason and Silverman 1994) or `S+Wavelets` (Bruce and Gao 1994), or in the `MatLab` packages `WaveLab` (Buckheit and Donoho 1995) or `FracLab` (INRIA Project Fractales, 2002). In particular, the computations described in this paper were based primarily on `FracLab`.

We now represent the parameters of the stable laws by a column vector $\theta = (\alpha, \beta, \mu, \sigma)'$, and the parameter space by Θ, for some characteristic function representation for the stables. The corresponding density, cumulative distribution function, and characteristic function (cf) will be denoted as $f_\theta(x)$, $F_\theta(x)$, and $c_\theta(t)$. Data will be denoted as $x_1, x_2, \ldots, x_n$, these being realizations of independent, identically distributed random variables $X_1, X_2, \ldots, X_n$ sampled according to some unknown value $\theta = \theta_0$. The empirical distribution function and empirical characteristic function (ecf) corresponding to these data will be denoted as $F_n(x)$ and $c_n(t)$.

The wavelet estimation procedures we propose are based on the following general result of Feuerverger and McDunnough (1984). Consider a class of functions $\{g_t(x);\ t \in \mathcal{T}\}$ indexed by $\mathcal{T}$, and set

$$\mathcal{E}_\theta\, g_t(X) \equiv G_\theta(t) \equiv \int g_t(x)\,\mathrm{d}F_\theta(x) \tag{3}$$

and

$$\mathcal{E}_n\, g_t(X) \equiv G_n(t) \equiv \int g_t(x)\,\mathrm{d}F_n(x) = \frac{1}{n}\sum_{j=1}^{n} g_t(X_j). \tag{4}$$

Here $\mathcal{E}_\theta$ and $\mathcal{E}_n$ represent expectations with respect to the distributions F_θ and F_n. Suppose now that we estimate θ by "fitting" $G_\theta(t)$ to $G_n(t)$ at a finite collection of values $t_1, t_2, \ldots, t_k \in \mathcal{T}$ (for some k) using generalized, nonlinear least squares. (The term "generalized" is used here in its ordinary meaning of least squares with the covariance structure of the error terms taken appropriately into account.) Then such procedures can be made to attain arbitrarily high asymptotic efficiency (provided we use a sufficiently extensive grid $\{t_j\}_{j=1}^k \subset \mathcal{T}$) if and only if the closure in

the weighted space $L^2(f_{\theta_0})$ of the finite linear combinations $\sum_{j=1}^{k} a_j\, g_{t_j}(x)$, where k and the $t_1, ..., t_k$ are arbitrary, includes the components of the true vector 'score' function $\partial \log f_\theta(x)/\partial\theta\,|_{\theta=\theta_0}$. (Here $L^2(f_{\theta_0})$ is the linear space of functions $g(\cdot)$ whose norm $\left\{\int_{\mathbb{R}} g^2(x)\, f_{\theta_0}(x)\, \mathrm{d}x\right\}^{1/2}$ is finite.) Since the true parameter value θ_0 is ordinarily unknown, it is more practical to require this criterion to hold for all θ_0 in the parameter space Θ. The situation will, of course, be particularly simple when the $\sum_{j=1}^{k} a_j\, g_{t_j}(x)$ span all of $L^2(f_\theta)$ for every θ.

Now, since for any $\theta \in \Theta$, the density of a stable distribution (w.r.t. Lebesgue measure) is bounded above, it is easy to see that if the collection $\{g_t(x);\ t \in \mathcal{T}\}$ is a basis of $L^2(\mathbb{R})$, then it will also span the weighted space $L^2(f_\theta)$ of the stable law. (This follows since $L^2(\mathbb{R}) \subset L^2(f_\theta)$ and, by the boundedness of f_θ, any approximating sequence in $L^2(\mathbb{R})$ is also an approximating sequence in $L^2(f_\theta)$.) Hereafter, the families $\{g_t(x);\ t \in \mathcal{T}\}$ of interest to us will be complete wavelet families and their Fourier transforms, typically denoted as $\{\psi_{j,k}(x);\ j, k \in \mathbb{Z}\}$ and $\{\hat{\psi}_{j,k}(t);\ j, k \in \mathbb{Z}\}$. In view of Fourier isometry, both these sets of functions form complete families spanning all of $L^2(\mathbb{R})$. Furthermore, for reasonable continuous parameterizations, all score functions of the stable laws belong to their corresponding $L^2(f_\theta)$ spaces, a result that was proved by DuMouchel (1971, 1973a). Therefore procedures based on fitting population to sample moments defined by wavelets or their Fourier transforms, using generalized (nonlinear) least squares, will permit arbitrarily high asymptotic efficiency.

We are thus led to consider (in one instance) the nonlinear generalized least squares fitting, in obvious notation, of

$$\langle \mathrm{d}F_n, \hat{\psi}_{j,k}\rangle = \langle f_\theta, \hat{\psi}_{j,k}\rangle + \epsilon_{j,k}, \tag{5}$$

or equivalently (by Parseval's equality) of

$$\langle \mathrm{d}F_n, \hat{\psi}_{j,k}\rangle = \langle c_\theta, \psi_{j,k}\rangle + \epsilon_{j,k}\,, \tag{6}$$

using many values of j and k, where $\{\psi_{j,k}\}$ is a complete wavelet family, and $\{\hat{\psi}_{j,k}\}$ are its Fourier transforms. In fact, writing the left hand side of Eq. (6) as

$$Y_{j,k} \equiv \langle \mathrm{d}F_n, \hat{\psi}_{j,k}\rangle \equiv \frac{1}{n}\sum_{\ell=1}^{n} \hat{\psi}_{j,k}(X_\ell) \tag{7}$$

allows us to exhibit Eq. (6) in standard least squares format:

$$Y_{j,k} = \langle c_\theta, \psi_{j,k}\rangle + \epsilon_{j,k}\,. \tag{8}$$

Here the $\epsilon_{j,k}$ are (complex-valued) error terms whose means are 0 and whose covariance matrix—to be denoted by Σ— will be discussed further below. While the form of Eq. (8) is simple, evaluation of the empirical wavelet coefficients as in Eq. (7) can be computationally demanding, especially if the functions $\hat{\psi}_{j,k}$ do not have a closed form. However, one may apply Parseval's equality yet again to obtain an alternative expression for the "observations" $Y_{j,k}$ which is computationally more tractable, namely

$$Y_{j,k} = \langle \mathrm{d}F_n, \hat{\psi}_{j,k}\rangle = \langle c_n, \psi_{j,k}\rangle\,. \tag{9}$$

In Eq. (9), the quantities $\langle c_n, \psi_{j,k}\rangle$ can thus be computed by means of a fast wavelet transform applied to the empirical characteristic function c_n, whereas in Eq. (7), the quantities $\langle \mathrm{d}F_n, \hat{\psi}_{j,k}\rangle$ are typically much more difficult to compute.

We note in passing here that if (in the other instance) we reverse the roles of $\{\psi_{j,k}\}$ and $\{\hat{\psi}_{j,k}\}$ in the last paragraph, we then obtain estimating equations of the form

$$\begin{aligned} Y'_{j,k} \equiv \frac{1}{n}\sum_{\ell} \psi_{j,k}(X_\ell) = \langle \mathrm{d}F_n, \psi_{j,k}\rangle \;&=\; \langle f_\theta, \psi_{j,k}\rangle + \epsilon'_{j,k} \\ &= \langle c_\theta, \hat{\psi}_{j,k}\rangle + \epsilon'_{j,k}\,. \end{aligned} \tag{10}$$

These procedures are also asymptotically arbitrarily highly efficient. However, computation of the $\langle c_\theta, \hat{\psi}_{j,k}\rangle$ is problematical in that no fast algorithm is known for obtaining them, while computation of the equivalent forms $\langle f_\theta, \psi_{j,k}\rangle$ first requires obtaining the densities f_θ by Fourier inversion of the c_θ which we obviously seek to avoid here. For this reason we do not pursue this option below.

We next discuss the covariance matrix Σ of the error terms $\epsilon_{j,k}$ in Eq. (5). Our initial convention is that each row (and each column) of Σ corresponds to a particular pair (j, k), i.e., a particular wavelet function; a typical entry of Σ is therefore (in obvious notation) denoted as $\Sigma_{(j,k),(j',k')}$. However, quantities such as $\langle \mathrm{d}F_n, \hat{\psi}_{j,k}\rangle$ and $\langle c_\theta, \psi_{j,k}\rangle$ are actually complex valued, and their real and imaginary parts need to be 'separated'. Thus for each j, k an equation such as Eq. (6) in fact represents two equations: one for the real part and one for the imaginary part. Correspondingly, the number of rows and columns of the covariance matrix Σ must in fact equal twice the number of wavelet coefficients being used, with real and imaginary parts of every wavelet coefficient each corresponding to its own row and column. The entries of Σ are thus not based on covariances among the $\hat{\psi}_{j,k}(X)$, but rather among all their real and imaginary components.

In complex form (our 'initial convention') the entry of Σ corresponding to the (j, k) and (j', k') wavelets is given via

$$\begin{aligned} n\Sigma_{(j,k),(j',k')} &\equiv n\mathrm{Cov}\left(Y_{j,k},\; Y_{j',k'}\right) \\ &= \mathrm{Cov}\left(\hat{\psi}_{j,k}(X), \hat{\psi}_{j',k'}(X)\right) \\ &= \mathbb{E}\left(\hat{\psi}_{j,k}(X)\,\overline{\hat{\psi}_{j',k'}(X)}\right) - \mathbb{E}\left(\hat{\psi}_{j,k}(X)\right)\,\overline{\mathbb{E}\left(\hat{\psi}_{j',k'}(X)\right)}\,. \end{aligned} \tag{11}$$

By Parseval's identity,

$$\begin{aligned} \mathbb{E}(Y_{j,k}) = \mathbb{E}(\hat{\psi}_{j,k}(X)) &= \int \hat{\psi}_{j,k}(x)\; f_\theta(x)\,\mathrm{d}x \\ &= \int \psi_{j,k}(x)\, c_\theta(x)\,\mathrm{d}x; \end{aligned} \tag{12}$$

therefore, the quantities $\mathbb{E}\left(Y_{j,k}\right)$ in Eq. (11) are just the wavelet coefficients of c_θ which can readily be computed and which we denote by $\eta_{j,k}(c_\theta)$. Next,

$$\begin{aligned} E\left(\hat{\psi}_{j,k}(X)\,\overline{\hat{\psi}_{j',k'}(X)}\right) &= \int \hat{\psi}_{j,k}(x)\,\overline{\hat{\psi}_{j',k'}(x)}\,f_\theta(x)\,\mathrm{d}x \\ &= \int\int\int \psi_{j,k}(s)\mathrm{e}^{-\mathrm{i}sx}\,\overline{\psi_{j',k'}(t)\mathrm{e}^{-\mathrm{i}tx}}\,f_\theta(x)\,\mathrm{d}s\,\mathrm{d}t\,\mathrm{d}x \\ &= \int\int \psi_{j,k}(s)\,\overline{\psi_{j',k'}(t)}\,c_\theta(t-s)\,\mathrm{d}s\,\mathrm{d}t\,. \end{aligned} \tag{13}$$

Consequently

$$n\Sigma_{(j,k),(j',k')} = \int\int \psi_{j,k}(s)\,\overline{\psi_{j',k'}(t)}\,\left[c_\theta(t-s) - c_\theta(s)c_\theta(-t)\right]\,\mathrm{d}s\,\mathrm{d}t\,. \tag{14}$$

This approach can now be used to determine the full covariance structure among the real and imaginary components of the $Y_{j,k}$; however, the following alternative viewpoint is also useful. First note, for any cf $c(\cdot)$ and its empirical version $c_n(\cdot)$, the well known and easily derived identity

$$n\mathrm{Cov}\left(c_n(s), c_n(t)\right) = c(s-t) - c(s)c(-t). \tag{15}$$

From this, using the conjugate symmetry of characteristic functions, it follows that

$$\begin{aligned} n\mathrm{Cov}\left(\Re\, c_n(s), \Re\, c_n(t)\right) &= \frac{1}{2}\left[\,\Re\, c(s-t) + \Re\, c(s+t)\,\right] \\ &\quad -\Re\, c(s)\,\Re\, c(t)\,, \end{aligned} \tag{16}$$

$$\begin{aligned} n\mathrm{Cov}\left(\Re\, c_n(s), \Im\, c_n(t)\right) &= \frac{1}{2}\left[\,\Im\, c(s-t) + \Im\, c(s+t)\,\right] \\ &\quad -\Re\, c(s)\,\Im\, c(t), \end{aligned} \tag{17}$$

and

$$\begin{aligned} n\mathrm{Cov}\left(\Im\, c_n(s), \Im\, c_n(t)\right) &= \frac{1}{2}\left[\,\Re\, c(s-t) - \Re\, c(s+t)\,\right] \\ &\quad -\Im\, c(s)\,\Im\, c(t). \end{aligned} \tag{18}$$

(Here $\Re$ and $\Im$ denote real and imaginary parts.) Then using the linearities of covariance and inner product and the fact that the functions $\psi_{j,k}$ that we use are real, we obtain

$$\begin{aligned} \mathrm{Cov}\left(\Re Y_{j,k}, \Re Y_{j',k'}\right) &= \mathrm{Cov}\left(\langle \Re c_n, \psi_{j,k}\rangle, \langle \Re c_n, \psi_{j',k'}\rangle\right) \\ &= \int \psi_{j,k}(s)\,\mathrm{Cov}\left(\Re c_n(s), \Re c_n(t)\right)\psi_{j',k'}(t)\,\mathrm{d}s\,\mathrm{d}t. \end{aligned} \tag{19}$$

In this way we obtain the relation

$$
\begin{aligned}
&n\mathrm{Cov}\left(\Re Y_{j,k}, \Re Y_{j',k'}\right) \\
&\quad = \int\int \psi_{j,k}(s)\,\psi_{j',k'}(t)\left\{\frac{1}{2}\left[\,\Re c(s-t) + \Re c(s+t)\,\right] - \Re c(s)\,\Re c(t)\right\}\mathrm{d}s\,\mathrm{d}t \\
&\quad = \frac{1}{2}\int\int \psi_{j,k}(s)\,\psi_{j',k'}(t)\Re\left[c(s-t) + c(s+t)\right]\mathrm{d}s\mathrm{d}t \\
&\qquad -\eta_{j,k}(\Re\left[c_\theta\right])\,\eta_{j',k'}(\Re[c_\theta])\,, \qquad (20)
\end{aligned}
$$

and likewise

$$
\begin{aligned}
&n\mathrm{Cov}\left(\Re Y_{j,k}, \Im Y_{j',k'}\right) \\
&\quad = \int\int \psi_{j,k}(s)\,\psi_{j',k'}(t)\left\{\frac{1}{2}\left[\,\Im c(s-t) + \Im c(s+t)\,\right] - \Re c(s)\,\Im c(t)\right\}\mathrm{d}s\,\mathrm{d}t \\
&\quad = \frac{1}{2}\int\int \psi_{j,k}(s)\,\psi_{j',k'}(t)\Im\left[c(s-t) + c(s+t)\right]\mathrm{d}s\,\mathrm{d}t \\
&\qquad -\eta_{j,k}(\Re\left[c_\theta\right])\,\eta_{j',k'}(\Im[c_\theta])\,, \qquad (21)
\end{aligned}
$$

and

$$
\begin{aligned}
&n\mathrm{Cov}\left(\Im Y_{j,k}, \Im Y_{j',k'}\right) \\
&\quad = \int\int \psi_{j,k}(s)\,\psi_{j',k'}(t)\left\{\frac{1}{2}\left[\,\Re c(s-t) - \Re c(s+t)\,\right] - \Im c(s)\,\Im c(t)\right\}\mathrm{d}s\,\mathrm{d}t \\
&\quad = \frac{1}{2}\int\int \psi_{j,k}(s)\,\psi_{j',k'}(t)\Re\left[c(s-t) - c(s+t)\right]\mathrm{d}s\,\mathrm{d}t \\
&\qquad -\eta_{j,k}(\Im\left[c_\theta\right])\,\eta_{j',k'}(\Im[c_\theta])\,. \qquad (22)
\end{aligned}
$$

The entries of Σ corresponding to the real and imaginary parts of the wavelet coefficients are thus given by the expressions (20)–(22).

In the symmetric family $c_\theta(t) = \exp\{\mathrm{i}\mu t - |\sigma t|^\alpha\}$, if we work with centred variates such as $\tilde{X}_j = \left(X_j - \hat{\mu}\right)/\hat{\sigma}$, where $\hat{\mu}$ and $\hat{\sigma}$ are consistent estimates, the covariance between the real and imaginary parts of the ecf and hence of the wavelet coefficients $\mathrm{Cov}\left(\Re Y_{j,k}, \Im Y_{j',k'}\right)$ are null [see Eq. (21)] so that Σ becomes block diagonal. In this case our updating equations will separate into two parts: one involving only α and σ and the other involving only μ. This phenomenon is related to orthogonalities in the Fisher information matrix for symmetric scale-location families and is discussed further in Sect. 4.

The weighted least squares wavelet procedure is now based on a grid of size T, say, of scale-position pairs (j,k), and involves minimizing (with respect to θ) a quadratic form in the $2T$ quantities $\Re(Y_{j,k} - \eta_{j,k}(c_\theta))$ and $\Im(Y_{j,k} - \eta_{j,k}(c_\theta))$. Once we have estimated the $2T \times 2T$ matrix Σ, then starting from any consistent estimate of θ, our nonlinear least squares updating algorithm is given by

$$
\hat{\theta}_{\mathrm{new}} = \hat{\theta}_{\mathrm{old}} + (G'\,\Sigma^{-1}G)^{-1}G'\,\Sigma^{-1}\eta_{\theta_{\mathrm{old}}}. \qquad (23)
$$

(The primes, here and elsewhere, represent the transpose of a matrix or a vector.) Here the length $2T$ column vector η_θ has entries consisting of the real and imaginary parts $\Re\left(Y_{j,k} - \eta_{j,k}(c_\theta)\right)$ and $\Im\left(Y_{j,k} - \eta_{j,k}(c_\theta)\right)$, and G is the $2T \times 4$ matrix

of partial derivatives of η_θ with respect to the four components of θ, these terms all being evaluated at $\theta = \theta_{\text{old}}$.

We mention here that Eq. (15), as well as Eq. (16)–(18), belong to the class of operators that are made sparse under the action of the wavelet transform (see, e.g., Flandrin 1992). Therefore one may hope to exploit the sparse and near-diagonal structure of the wavelet covariance matrix Σ when constructing the inverses in the updating scheme (23). A difficulty (which occurs in many inverse problems) is that for moderate n and large T, the true and estimated Σ will be ill-conditioned. Indeed, when selecting the grid (of size T) of scale-position pairs (j, k) there are two somewhat conflicting objectives. To eliminate instabilities arising from multicolinearities we seek to delete components whose variances are very small. But, at the same time, it is undesirable to delete components having large correlations with the $c_n(t)$ since these carry substantial information for θ. Because the estimation procedure is carried out in scale-space, one strategy for choosing the grid (to overcome ill-posedness) is to employ a multiscale regularization technique based on a truncated singular value decomposition of the estimated Σ; this amounts to choosing T by deleting all eigenvectors of the SVD corresponding to eigenvalues smaller than some cut-off level. The choice of this cut-off level is somewhat arbitrary, but if the matrix Σ is first rescaled into a correlation matrix (which has an average eigenvalue equal to 1) a cut-off value between 0.01 and 0.1 appears to be useful in practice. A more sophisticated method would be to add to Σ a small diagonal matrix, with scale-dependent diagonal weights (as in ridge regression) before proceeding to the inversion of the matrices in the updating algorithm in a scale-recursive manner.

Since the performance of regularization methods is context-dependent, it is difficult to make general statements about the superiority of one regularization method over another. Whatever method is chosen, we used a QR decomposition (see, e.g., Strang 1986) to iteratively solve Eq. (23); this is the standard method to solve overdetermined systems by least squares. Specifically, at each iteration, we factorize the part $\Sigma^{-1}G = QR$ with R upper triangular and Q orthogonal in the sense that $Q\,\Sigma^{-1}Q = I$ and then solve the resulting triangular system $Q'\,\eta_{\theta_{\text{old}}} = R\,(\theta - \theta_{\text{old}})$ by substitution.

4 Numerical algorithms

In this section we provide some details of the numerical algorithms used in our experiments. In particular we indicate how the theoretical details of Sect. 3 (expressed there in continuous form) were implemented by means of discrete algorithms.

To simplify our numerical work, we experimented only with symmetric stable families, (i.e., $\beta = 0$). An important consideration is that the 4×4 Fisher information matrix then becomes block diagonal, with 2×2 blocks corresponding to parameter groupings (α, σ) and (β, μ) (DuMouchel 1975, Theorem 2). This allows us to concentrate efforts on the key parameter α, and on the scale parameter σ thereby associated with it. We saved additional efforts by using only a $\sqrt{n}$-consistent estimator for μ throughout. The Fisher information orthogonalities (when $\beta = 0$) imply that our estimators for α and σ may then (asymptotically) be

fairly compared to estimators for these two parameters obtained from algorithms which fit all four stable parameters simultaneously.

As before, denote by $\{x_j\}_{j=1}^n$ the data to be analysed; the x_j are independent variables from a stable distribution f_θ, with $\theta = (\alpha, \beta, \mu, \sigma)'$. As mentioned, we always used $\beta = 0$. Further, due to the invariance in location, we always used $\mu = 0$. It is thus convenient to change notation, and henceforth $\theta = (\alpha, \sigma)'$, a column vector of length two. We therefore speak of parameter estimates $\hat{\theta} = (\hat{\alpha}, \hat{\sigma})'$, and these are first initialized by the values $\hat{\alpha}_1$ and $\hat{\sigma}_1$ obtained using the four parameter estimation procedure of Koutrouvelis (1980), as implemented in `FracLab`. The parameter μ was estimated by the $\sqrt{n}$-consistent estimator $\hat{\mu}_0$, defined as the mean of the subset of the data obtained by removing 28% of the smallest and 28% of the largest values. (See Fama and Roll 1971)

Our estimation updating algorithm now proceeds iteratively. Let k denote the iterative step in the nonlinear least squares fitting procedure. We determined, after experimentation, that three iterations always sufficed for satisfactory convergence; consequently k here always ranged over the values 1–3. Each step, say the kth, of the estimating loop, thus proceeds as follows. Data are centred (by removing the trimmed mean $\hat{\mu}_0$) and then renormalized using the current estimate $\hat{\sigma}_k$ to $\{\tilde{x}_j\}_{j=1}^n$, where $\tilde{x}_j = (x_j - \hat{\mu}_0)/\hat{\sigma}_k$. We then compute the empirical characteristic function

$$c_n(s_\ell) = \frac{1}{n} \sum_{j=1}^{n} \exp\{i s_\ell \tilde{x}_j\} \tag{24}$$

over an interval $[-S, S)$, using sampled gridpoints $s_\ell = \pm\delta\ell$, where $\ell = 0, \ldots, M$, and sampling rate $\delta = S/M$.

We next compute the analytic c_θ at the current θ, namely $\hat{\theta}_k = (\hat{\alpha}_k, \hat{\sigma}_k)'$; here β is implicitly 0, while μ is also 0 in view of the centring at $\hat{\mu}_0$. Analytic forms of the derivatives of c_θ w.r.t. α, σ, and μ (required for the nonlinear regression fitting) are then also computed over the same grids and evaluated at the current parameter values; denote these dimension $2M$ column vectors as $\partial_\alpha c_\theta$, $\partial_\sigma c_\theta$, and $\partial_\mu c_\theta$, respectively. These quantities are also computed over $[-S, S)$, using gridpoints $s_\ell = \delta\ell$, but to obtain the covariance structures, c_θ must actually be computed over a support $[-2S, 2S]$ that is twice as large. No tapering was required.

We then computed the full tree wavelet decompositions (see, e.g., Mallat 1998) for each of the finite vectors defined above; we denote such decompositions by $W[c_\theta]$, $W[\partial_\alpha c_\theta]$, $W[\partial_\sigma c_\theta]$, and $W[\partial_\mu c_\theta]$. Assuming a signal of $N \equiv 2M = 2^J$ points, such decomposition consist of $N/2$ wavelet coefficients at the finest scale $j = -1$, $N/4$ wavelet coefficients at the next finest scale $j = -2, \ldots\ldots, 2$ wavelet coefficients at scale $j = -J + 1$, and finally, 1 wavelet coefficient at the coarsest scale $j = -J$, as well as 1 scaling coefficient also at this scale—a total of $N/2 + N/4 + \cdots + 2 + 1 + 1 = N$ coefficients in all.

In our experiments, we used the real, compactly supported Daubechies wavelets with various degrees of regularity. We also experimented using various numbers of wavelet coefficients. Thus, at each scale we would retain, say, only the N_ψ wavelet coefficients lying nearest to the singularity (i.e., origin) of the characteristic function. At coarser scales, when the number of available coefficients falls below N_ψ, all coefficients were kept.

We then form the $2M \times 2$ and $2M \times 1$ matrices $D\,\theta^R = \{W[\partial_\alpha c_\theta],\ W[\partial_\sigma c_\theta]\}$ and $D\,\theta^I = W[\partial_\mu c_\theta]$ (remark: when not all wavelet coefficients are used these matrices will, of course, have fewer than $2M$ rows) as well as the $(2M+1) \times (2M+1)$ real Toeplitz matrix T whose first row just constitutes the values of $c_\theta(s_\ell)$:

$$T = \begin{bmatrix} c_\theta(0) & c_\theta(\delta) & c_\theta(2\delta) & \cdots & c_\theta(2M\delta) \\ c_\theta(\delta) & c_\theta(0) & c_\theta(\delta) & \cdots & c_\theta((2M-1)\delta) \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ c_\theta(2M\delta) & c_\theta((2M-1)\delta) & c_\theta((2M-2)\delta) & \ldots & c_\theta(0) \end{bmatrix}.$$

We also built the $(2M+1) \times (2M+1)$ real cross-product matrix

$$R = c'_\theta\, c_\theta = \begin{bmatrix} c_\theta^2(M\delta) & c_\theta(M\delta)c_\theta((M-1)\delta) & \cdots & c_\theta^2(M\delta) \\ c_\theta((M-1)\delta)c_\theta(M\delta) & c_\theta^2((M-1)\delta) & \cdots & c_\theta((M-1)\delta)c_\theta(M\delta) \\ \vdots & \vdots & \vdots & \vdots \\ c_\theta^2(M\delta) & c_\theta(M\delta)c_\theta((M-1)\delta) & \cdots & c_\theta^2(M\delta) \end{bmatrix}.$$

The theoretical covariance matrix of the real part of the empirical cf is then given by

$$\Sigma^R = \frac{1}{2}(T + \tilde{T}) - R\,, \tag{25}$$

where $\tilde{T}$ is the 90° anticlockwise rotated version of T, while the covariance matrix for the imaginary part is just

$$\Sigma^I = \frac{1}{2}(T - \tilde{T})\,. \tag{26}$$

The quantities (25) and (26) correspond to Eqs. (16) and (18); the matrix corresponding to Eq. (17) is null for symmetric laws.

The wavelet regression procedure requires the covariance matrix of the wavelet coefficients; for coefficients corresponding to the real part of the empirical cf this is

$$\Psi^R = \mathcal{W}\,\Sigma^R\,\mathcal{W}', \tag{27}$$

where $\mathcal{W}$ is the matrix of the 1-D wavelet transform. The (nonstandard) 2D wavelet transform (27) thus corresponds to a 1D orthogonal wavelet transform applied to each column of Σ^R, together with a 1D wavelet transform applied to each row. The covariance matrix for the wavelet coefficients of the imaginary part of the empirical cf is obtained similarly as $\Psi^I = \mathcal{W}\,\Sigma^I\,\mathcal{W}'$, but was not required. (The cross-covariance matrix here is null.) When not all wavelet coefficients are retained in the fitting, Ψ^R is pruned so only the required covariances are kept.

Finally, the procedure used to update $\hat{\theta}$ was $\hat{\theta}_{k+1} = \hat{\theta}_k + \rho\,\mathrm{d}\theta$ where $\rho < 1$ is a *damping factor*, and $\mathrm{d}\theta$ is the weighted least squares solution of the nonlinear regression

$$W[\Re\{c_N\}] - W[\Re\{c_\theta\}] = \left(D\,\theta^R\right)\mathrm{d}\theta + \mathcal{E}. \tag{28}$$

Here, the error vector $\mathcal{E}$ is regarded as approximately normal with zero mean and covariance Ψ^R, while θ on the left is at the current estimate. The solution (which can be highly singular if many coefficients are used) was obtained using MatLab's `lscov.m`, which is based on QR decomposition, and does not require Ψ^R to be inverted. In particular, `lscov` solves weighted least squares minimization problems for overdetermined linear regression systems of the form $\|Ab - y\|^2$ with A having more rows than columns and y having a known covariance Σ. We found that $\rho = 0.9$ resulted in good overall performance in our algorithm. Thus ends one iteration; as noted previously, three iterations (k = 1, 2, and 3) were found to be adequate always.

Finally, we mention that if location μ were also to be fitted using our wavelet procedure then Eq. (28) would need to be augmented by the one-parameter regression $W[\Im\{c_N\}] - W[\Im\{c_\theta\}] = \left(D\,\theta^I\right)\mathrm{d}\mu + \mathcal{E}'$ for that parameter. If the full four parameter family is to be fitted, the covariance structure of the real and imaginary terms would not separate (we would not have the parameter orthogonalities), so we would need to fit a regression using the real and imaginary ecf terms combined as one vector, together with a full covariance matrix which will now include nonzero cross terms off the diagonal blocks, since the terms (17), and hence (21), are not null in nonsymmetric families. We remark that the quantities $\partial_\mu c_\theta$ and $D\,\theta^I$ are not actually required when only α and σ are being fitted; these would be required only to update the estimates for μ and (in nonsymmetric families) for β.

5 Summary of numerical experiments

In this section we describe some numerical experiments carried out using the algorithms described in Sect. 4. However, we make two general observations. The first is that wavelet transformation tends to 'disbalance' or 'alter the presentation' of data; it also tends to diagonalize, or make sparse, the covariance and correlation structures of signals to which wavelet transformation is applied. (See, e.g., Flandrin 1992) To illustrate this, Figs. 1 and 2 show the covariance and correlation functions (matrices), multiplied by n, associated with the real part of the ecf on the interval $t \in [-1, 1]$ for the standardized stable distribution with $\alpha = 0.5$, while Figs. 3 and 4 show the corresponding functions for the imaginary part of the ecf; these plots are based on the linear colour-code scales at the right in these figures. (We chose a low α in order to exhibit high levels of correlation; as α increases, the illustrated effects only become more striking.) The covariances are seen to range from 0 to just under 1 for the real part and from approximately -1 to $+1$ for the complex part, with substantial correlations evident throughout both Figs. 1 and 3. In fact, there are also substantial correlations on the horizontal and vertical axes here, but these are not evident in the covariance plots because the variance of the ecf is zero at the origin; these correlations are evident in Figs. 2 and 4. Figures 5 and 6 show the absolute values of the covariances and the correlations of the wavelet coefficients for the real part of the ecf based on a grid of 128 points per unit (corresponding to the previous figures.) Here there are $J = \log_2 256 = 8$ wavelet scales, with the finest scale occupying the leftmost half of the horizontal axis, and bottom half of the vertical axis. Many covariances after wavelet transformation are so low that the colour coding for this plot (shown on the right) is based on a logarithmic dynamic going from 0 db down to -55 db below the maximum (10 db equals a

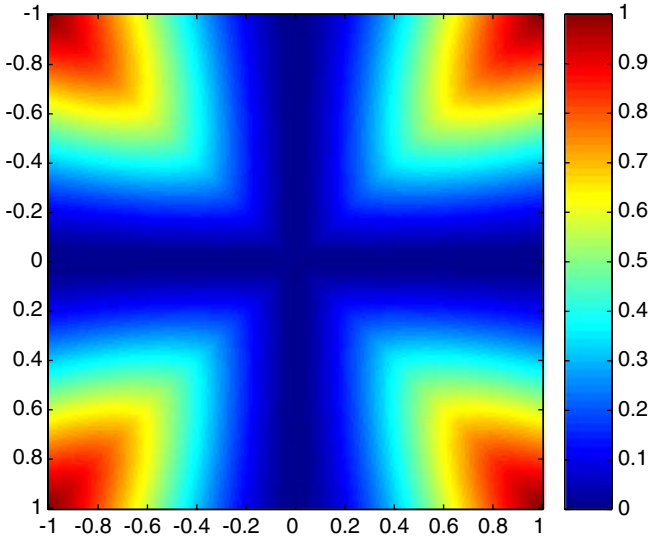

Fig. 1 Covariance matrix of $\Re\, c_n$. The covariance matrix of $\Re\, c_n(t)$ (multiplied by n) on $t \in [-1, 1]$, in the standardized case $\alpha = 0.5$, with linear colour scale

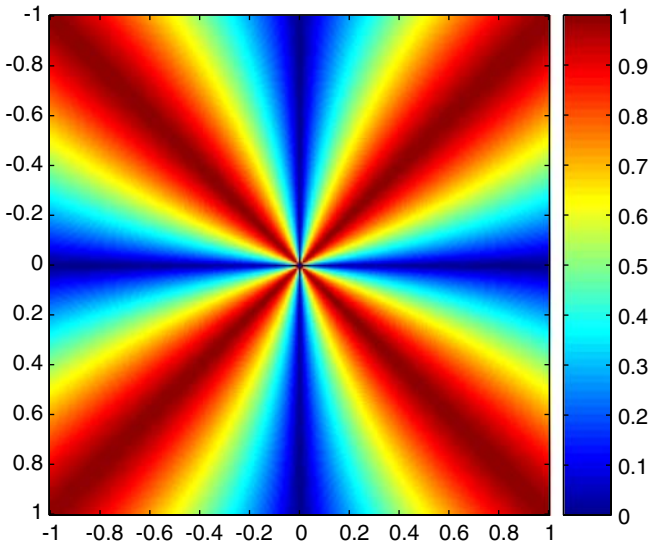

Fig. 2 Correlation matrix of $\Re\, c_n$. The correlation matrix corresponding to Fig. 1

multiplicative factor of 10). The wavelet covariance structure is seen to be highly concentrated at the coarsest scales (top and right of the image) with otherwise many near zero covariances throughout. The 'fishnet' structure in these plots traces the diagonal variance terms within the block matrices for each scale (with anti-diagonal terms mirroring the symmetry of $\Re\, c_n$), as well as covariance terms across scales corresponding to different wavelet functions which are centred at (nearly)

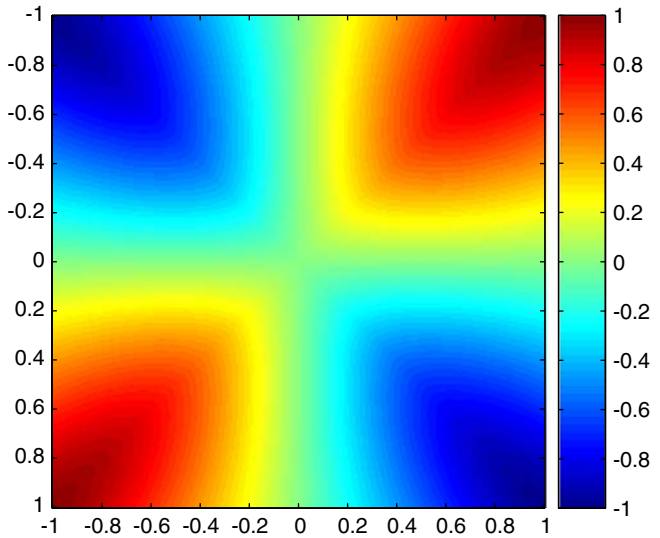

Fig. 3 Covariance matrix of $\Im c_n$. The covariance matrix of $\Im c_n(t)$ (multiplied by n) on $t \in [-1, 1]$, in the standardized case $\alpha = 0.5$, with linear colour scale

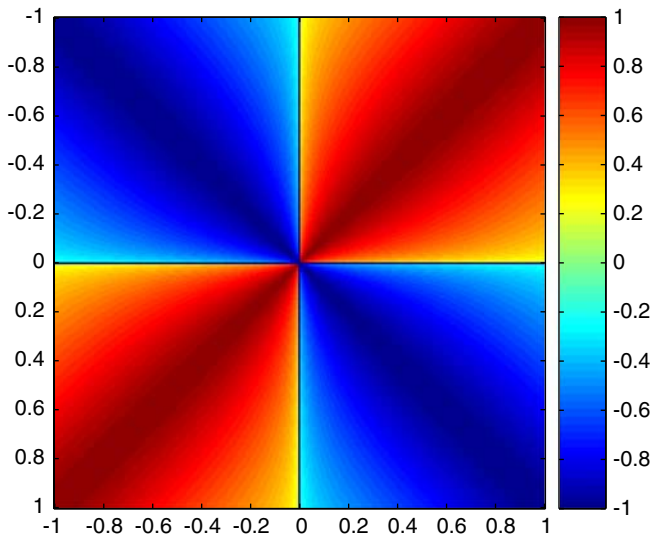

Fig. 4 Correlation matrix of $\Im c_n$. The correlation matrix corresponding to Fig. 3

identical locations; the (dark blue) remaining correlations are all essentially null. The covariance and correlation plots for wavelet coefficients of the imaginary part of the ecf are very similar to Figs. 5 and 6 and therefore omitted.

Our second observation is that wavelet transformation tends to redistribute and 'concentrate' (Fisher) information in data. This point is difficult to make rigorously because the quantities we are examining (in this case the 'view' of the data

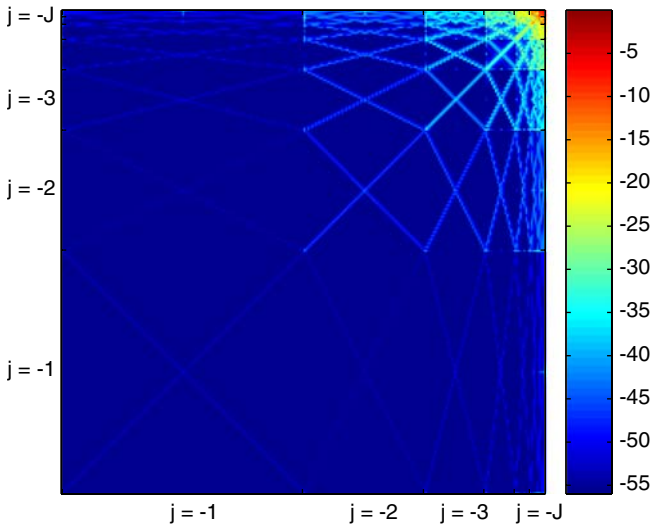

Fig. 5 Covariance matrix of $\mathcal{W}\,\Re\,c_n$. The covariance matrix of the wavelet coefficients $\mathcal{W}\,\Re\,c_n$ corresponding to Fig. 1, with logarithmic colour scale

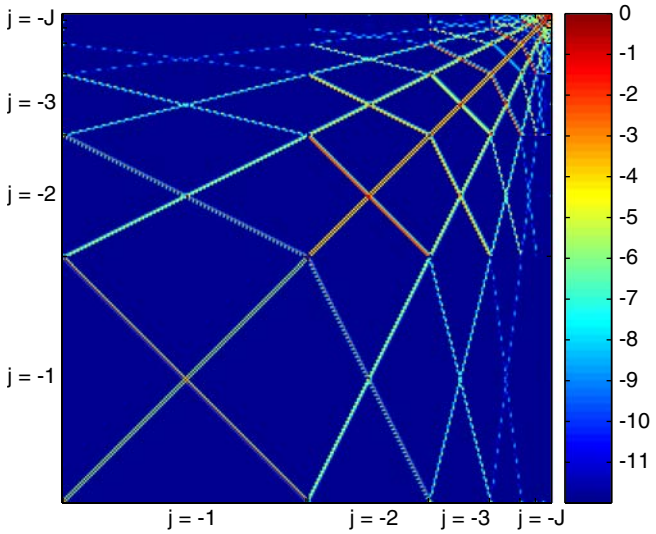

Fig. 6 Correlation matrix of $\mathcal{W}\,\Re\,c_n$. The correlation matrix corresponding to Fig. 5, with logarithmic colour scale

given by points on the ecf) are correlated, and this nonorthogonality entails (Fisher) information structure which is not simply additive. However, this phenomenon can be illustrated heuristically by looking at the information in individual terms. For example, if all parameters are known except one—say α—and if this must be estimated from only one point, say $\Re\,c_n(t)$, on the real ecf, by fitting to it the true cf $\Re\,c_\theta(t)$ there, then a simple Taylor expansion argument shows that the asymptotic

variance of this estimator is

$$\mathrm{Var}\,\hat{\alpha} \;=\; [\,\mathrm{Var}\,\Re\, c_n(t)\,]\left[\frac{\mathrm{d}\,\Re\, c_\theta(t)}{\mathrm{d}\alpha}\right]^{-2}.$$

In that sense, the Fisher information per observation in $\Re\, c_n(t)$ for the parameter α is

$$[\,n\mathrm{Var}\,\Re\, c_n(t)\,]^{-1}\left[\frac{\mathrm{d}\,\Re\, c_\theta(t)}{\mathrm{d}\alpha}\right]^{2}. \tag{29}$$

(Alternately, we could refer to this as the 'statistical sensitivity' of $\Re\, c_n(t)$ to this parameter.) Formulas analogous to Eq. (29) also hold for the other stable law parameters, as well as for $\Im\, c_n(t)$, and also for individual wavelet coefficients.

Figure 7 shows the information per observation for α, σ, and μ—in the sense of Eq. (29)—at each point of the ecf for a stable distribution with $\alpha = 1.75, \sigma = 1$, and $\mu = 0$; the curves for α and σ are computed using the real part of the cf while that for μ is computed using the imaginary part, these being the relevant information components for each parameter. In these plots, the ecf is viewed on the domain $t \in [-4, 4]$ with a gridding of 32 points per unit. Figure 8 shows the information per observation for the corresponding individual wavelet coefficients (with coarsest scales leftmost and only wavelets centred on the nonnegative axis included). Compared to Fig. 7, it is seen that only a few wavelet coefficients contain any appreciable amount of information; to show this clearly, the plots for α, σ, and μ are not superimposed here. To further illustrate these points, Figs. 9, 10, and 11 plot the information quantities for the individual ecf points (over the positive domain) as well as for their corresponding wavelet coefficients, for each of the parameters α, σ, and μ, in descending order (of information). Comparing the

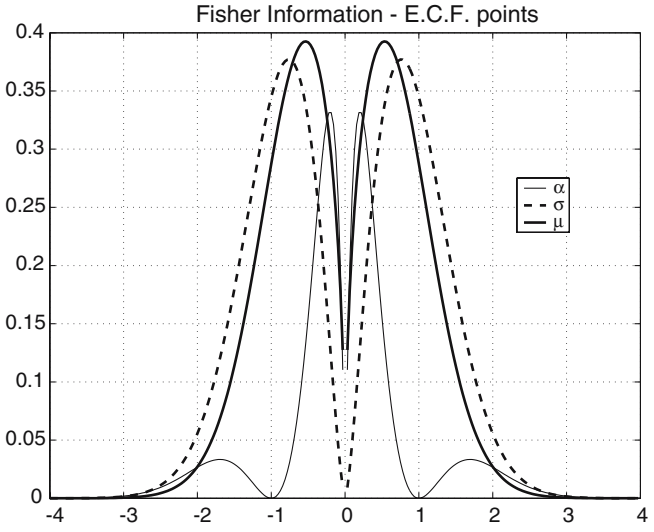

Fig. 7 Information for α, σ, and μ in $\Re\, c_n$. The pointwise 'information' (or statistical sensitivity) per observation for the estimation of α, σ, and μ, using points on the real ecf

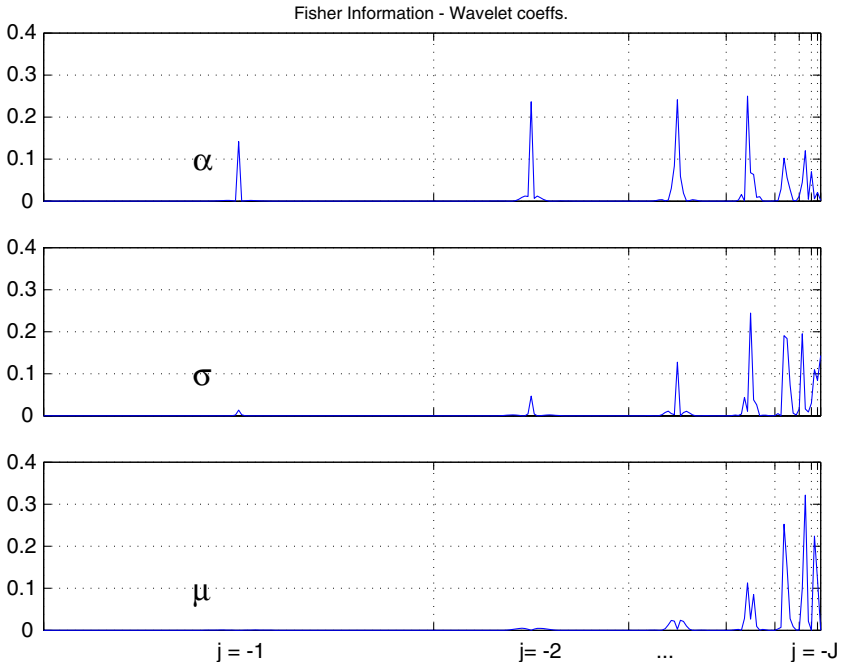

Fig. 8 Information for α, σ, and μ in $\mathcal{W}\Re\, c_n$. The 'information' (or statistical sensitivity) per observation for the estimation of α, σ, and μ using individual wavelet coefficients of the real ecf

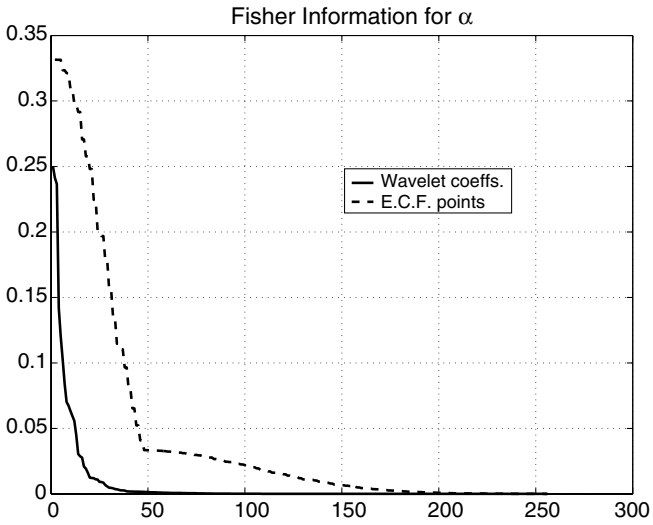

Fig. 9 The distribution of information for α. The information content for α of individual wavelet coefficients and ecf points (corresponding to the real part of the ecf), ordered from largest to smallest

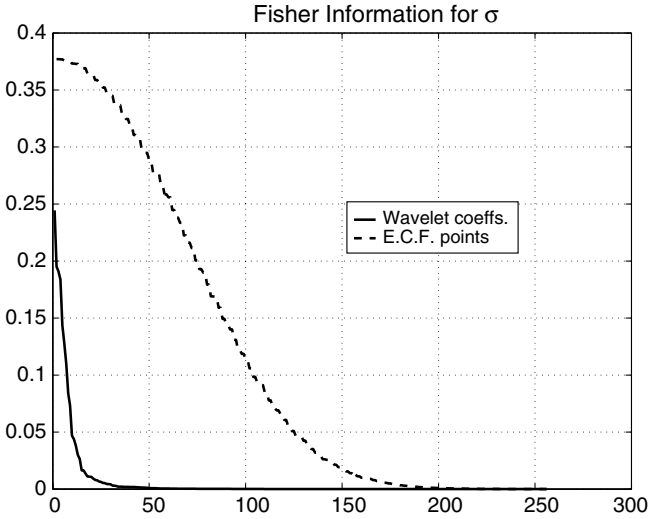

Fig. 10 The distribution of information for σ. The information content for σ of individual wavelet coefficients and ecf points (corresponding to the real part of the ecf), ordered from largest to smallest

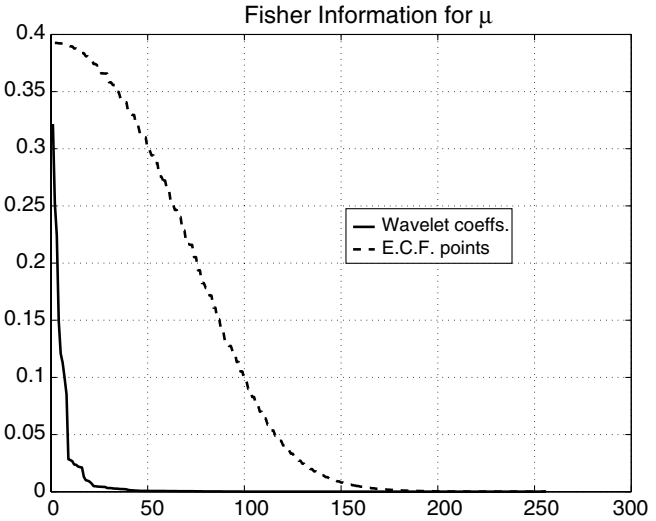

Fig. 11 The distribution of information for μ. The information content for μ of individual wavelet coefficients and ecf points (corresponding to the imaginary part of the ecf), ordered from largest to smallest

rates of descent of the two curves in each plot again emphasizes that fewer wavelet coefficients carry greater proportions of the available information. A practical consequence of this rearrangement and concentration of Fisher's information is that typically fewer wavelet coefficients are required (than, say, ecf coefficients) to obtain the same level of statistical performance for estimation and inference. The use of fewer and less correlated coefficients also carries numerical advantages.

We turn now to the results of a simulation study summarized compactly in Figs. 12 and 13. These figures each consist of 6×4 arrays, with each array consisting of a collection of 'box and whisker' plots. Each such box plot has lines at the lower quartile, median, and upper quartile, with 'whiskers' showing the extent of the remaining data, and with outliers, if any, indicated beyond the whiskers. Each box and whisker plot summarizes the results of 100 trials. (The same 100 simulated data sets were used throughout for each sample size.) For comparative purposes, box plots for estimates obtained from the four parameter maximum likelihood procedure of McCulloch (1998) and for estimates obtained from the four parameter regression procedure of Koutrouvelis (1980) are included within each array. Both the McCulloch and Koutrouvelis estimators used here are based on algorithms implemented in `FracLab`. In all of the results here, we used the compactly supported Daubechies wavelet of regularity 2 (i.e., four nonzero filter coefficients and two vanishing moments, the most symmetrical case). However, in our experimentation we found that the exact choice of wavelet basis did not result in appreciable differences.

Figure 12 summarizes the performance of the estimators for α. The six rows of Fig. 12 correspond to true values of $\alpha = 1.9$ (top), 1.75, 1.5, 1.0, 0.75, and 0.5 (bottom). Estimates of α were always truncated to lie in $[0, 2]$. The box and whisker plots within the column 1 arrays compare the McCulloch, Koutrouvelis, and wavelet estimators for α for the sample sizes $n = 25$, 200, and 1, 600 as indicated on (the horizontal, i.e.) the x-axis. (These sample sizes correspond to those used in Koutrouvelis 1980.) The wavelet estimators within the column 1 arrays are based on wavelet coefficients of the real ecf taken over the interval $[-2, 2]$ with grid spacing of 16 points per unit and with all 32 wavelet coefficients included in the regression fit. The columns 2, 3, and 4 arrays are all based on samples of size $n = 200$ and show the effects of varying individually various parameters of the wavelet regression algorithm. (Results for the other sample sizes are not included here.) The column 2 arrays show the effect on the wavelet estimator of varying the length of the cf support interval (0.1, 0.5, 1, 2, 4, 10) as shown on the x-axis. The column 3 arrays show the effect of varying the number of gridpoints (8, 16, 32, 64, 128, 256) used on the ecf, as indicated on the x-axis. And, finally, the column 4 arrays show the effect of using a reduced number of wavelet coefficients (32, 8, 4, 2) as indicated on the x-axis. Algorithm parameters not being varied in column arrays 2, 3, and 4 were otherwise held fixed at the values established for the column 1 arrays. Within each array of columns 2, 3, and 4, the first two box plots are just the McCulloch and Koutrouvelis box plots (of sample size 200) that appear within the leftmost column arrays. To facilitate comparisons, a full horizontal line is drawn at the true parameter value across each of the box-plot arrays.

Figure 13 is the same as Fig. 12 except it summarizes performance of the estimators for the scale parameter σ. The rows in these figures still correspond to the values of α equal to 1.9, 1.75, 1.5, 1.0, 0.75, and 0.5, as before. Due to the statistical invariance under scaling, only the true value $\sigma = 1$ is used throughout.

We have not included summaries here for estimates of the location parameter μ but note that estimates of μ obtained from fitting the symmetric three-parameter family cannot be compared directly to estimates of μ obtained from fitting a full four-parameter stable family due to the fact that μ is not orthogonal to the skewness parameter β (even when the true $\beta = 0$). It is worth noting that DuMouchel

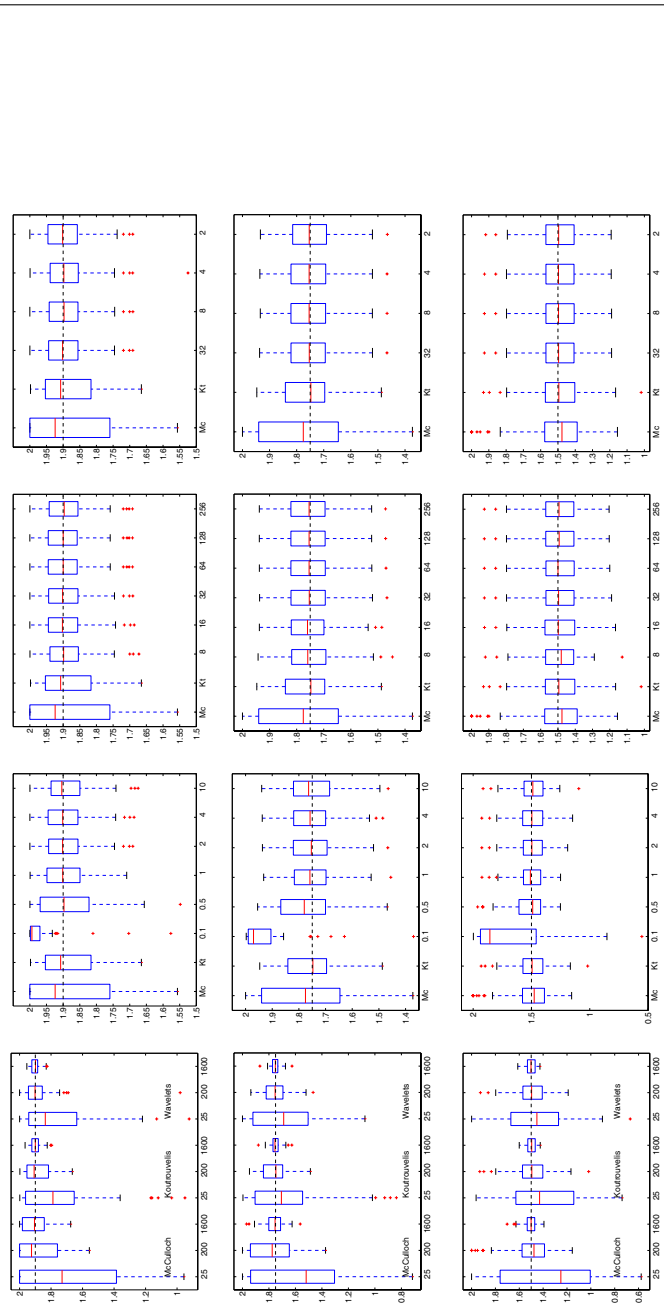

Fig. 12 Performance of the wavelet estimators for α. Array consisting of 6 rows $\times$ 4 columns of collections of 'box and whisker' plots. Rows 1 through 6 correspond to values of $\alpha = 1.9, 1.75, 1.5, 1.0, 0.75$, and 0.5. Box and whisker plots in column 1 compare the McCulloch, Koutrouvelis, and Wavelet estimators for α for sample sizes $n = 25, 200$, and $1,600$. The wavelet estimators in column 1 are based on the real ecf over the interval $[-2, 2]$, with grid spacing of 16 points per unit and with all 32 wavelet coefficients included in the fitting procedure. Box and whisker plots in columns 2, 3, and 4 are based on the sample size $n = 200$. Column 2 shows the effects on the wavelet estimator of varying the length of the cf support interval, but with the remaining algorithm parameters otherwise fixed as in column 1. Column 3 shows the effects of varying the number of gridpoints on the ecf, and column 4 shows the effects of using fewer than all wavelet coefficients (see text for further details)

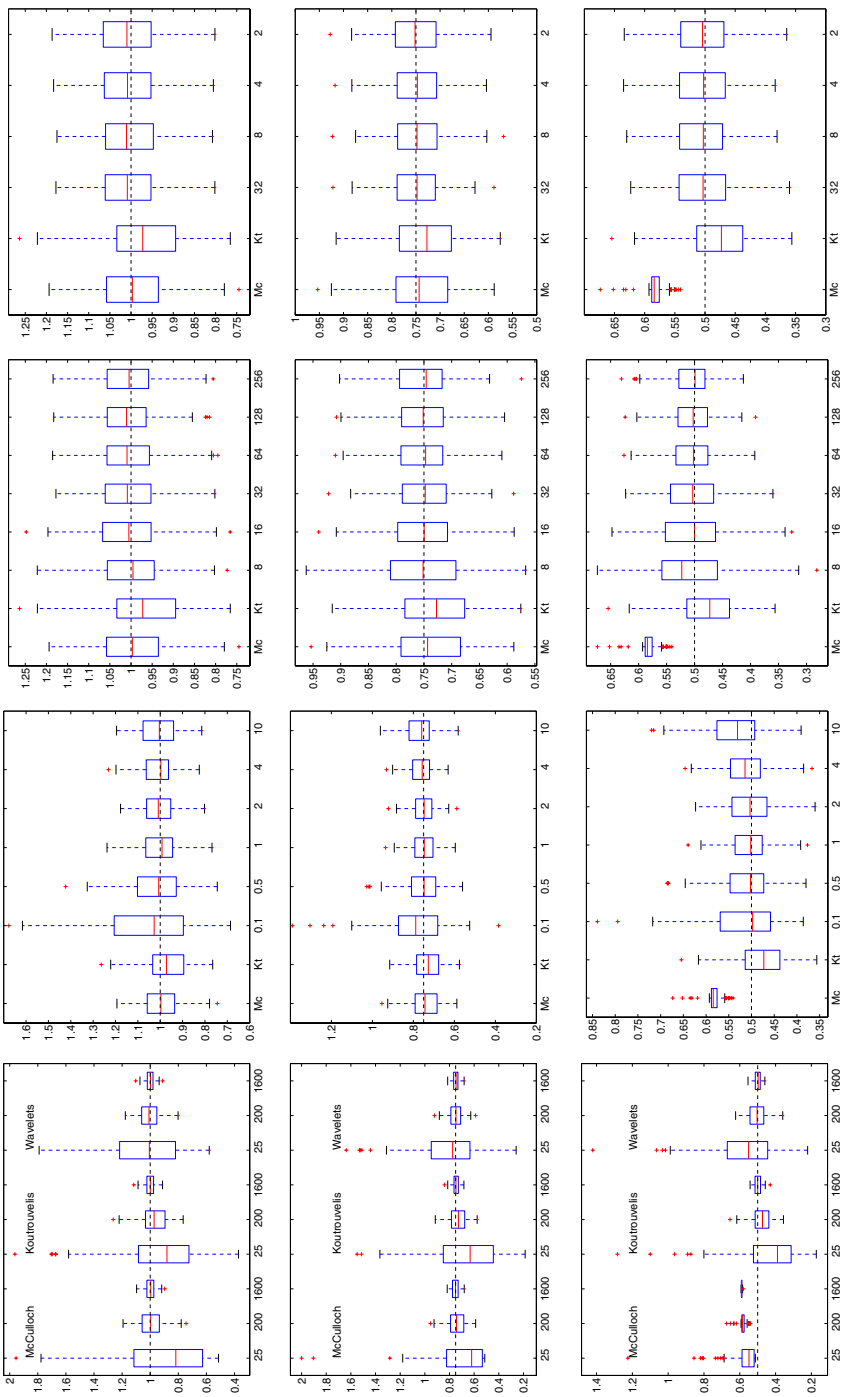
α = 0.5
α = 0.75
α = 1
McCulloch
Koutrouvelis
Wavelets
Mc
Kt
0.1
0.5
1
2
4
8
16
32
64
128
256
1600
200
25

Fig. 12 (Contd.)

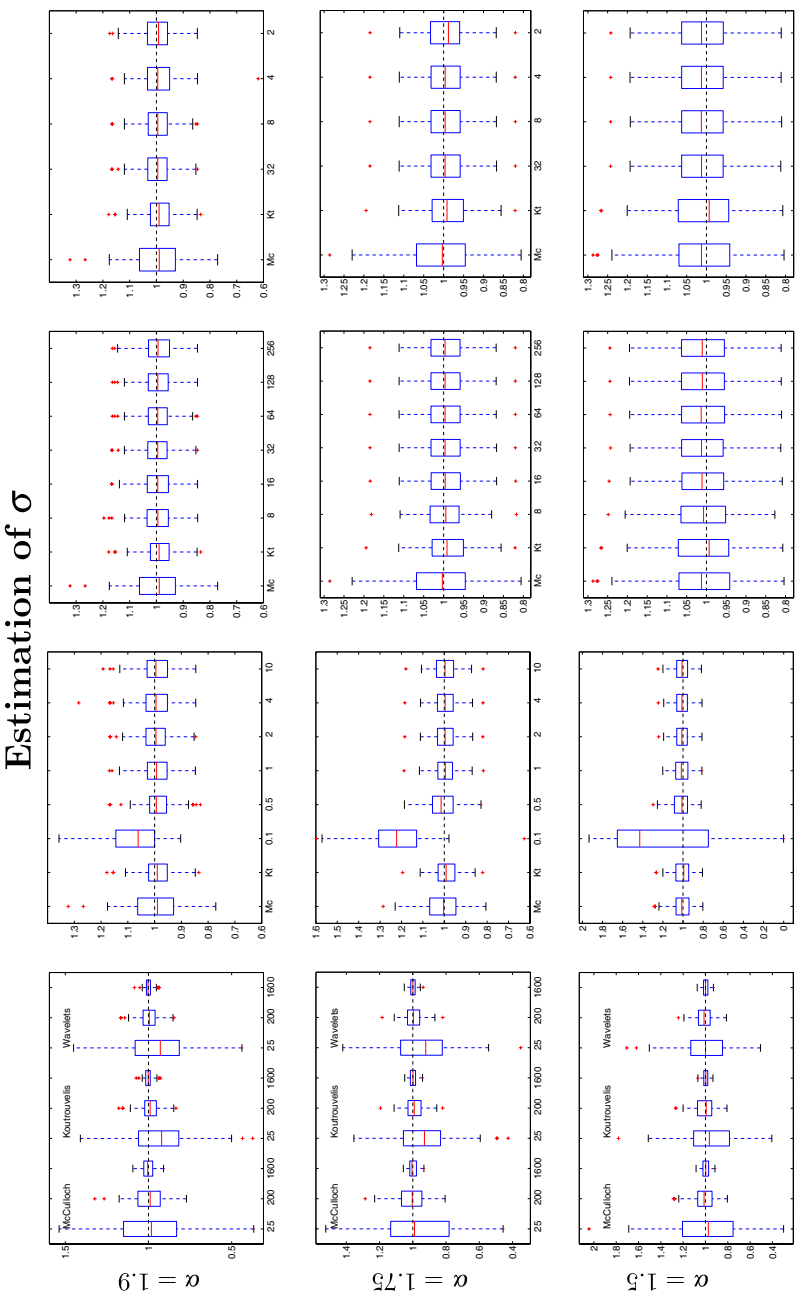

Fig. 13 Performance of the wavelet estimators for σ. Same as Fig. 12 except for estimators of the scale parameter σ. Rows correspond to values of α as in Fig. 12, with columns as before. True value of $\alpha = 1$ (see text for details)

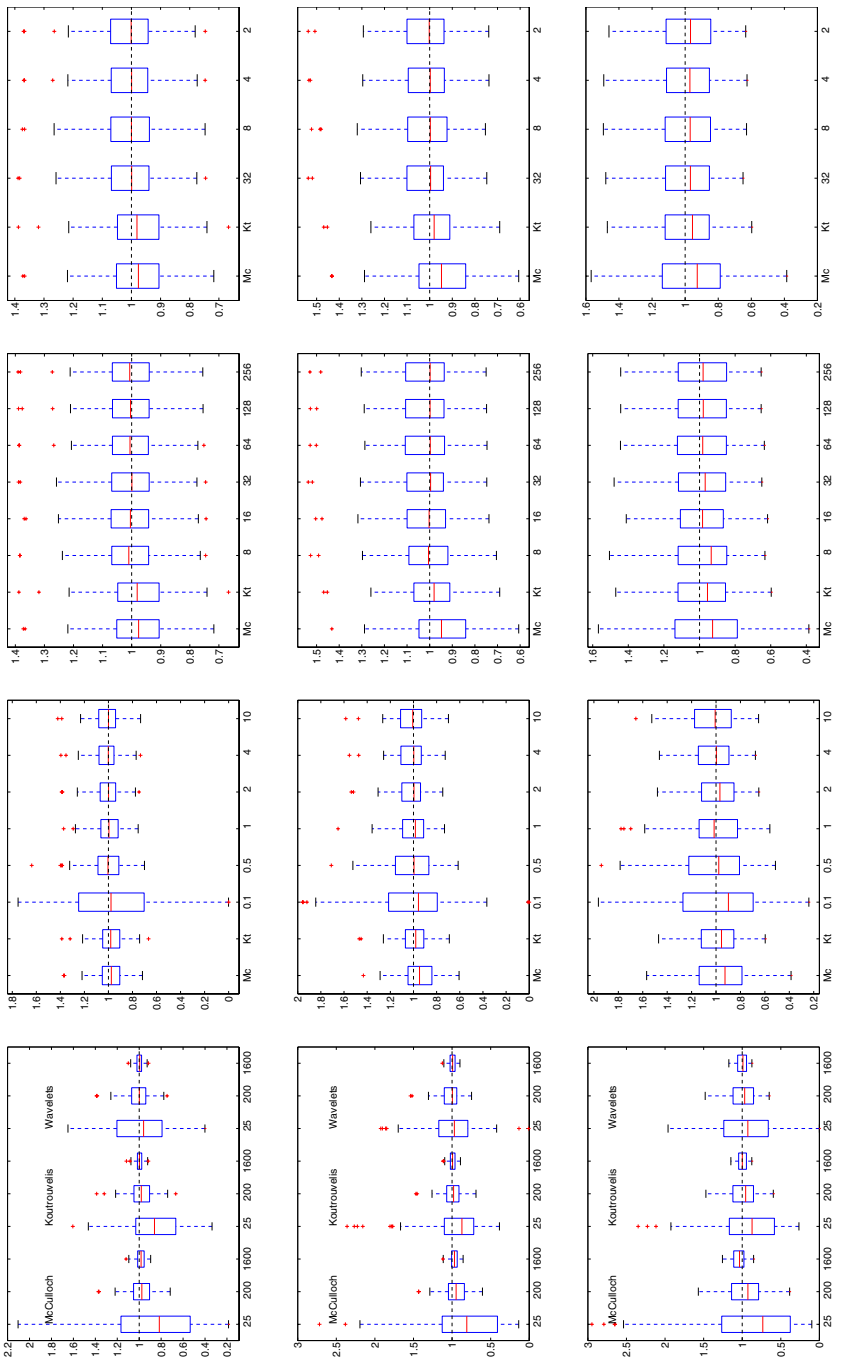

Fig. 13 (Contd.)

(1975, Table 1) does provide some asymptotic correlations $\rho_{\beta,\mu}$ for these correlated parameters for several values of α; see also Nolan (2001). Variances of estimates of μ with $\beta = 0$ assumed known versus assumed unknown may therefore be adjusted by applying a correction factor $\left(1 - \rho^2_{\beta,\mu}\right)$. As mentioned in the previous section, for α and σ we do not require similar asymptotic corrections. Note, however, that DuMouchel (1975) uses a discontinuous representation for the stables, so use of his Table 1 for values of β other than 0 may require adjustment of the value of β.

Examining Figs. 12 and 13 we see that for most combinations of algorithm parameters our wavelet-based estimation procedure somewhat outperforms (or at least matches) the performance of the standard estimators, especially for α close to 2 and small values such as 0.5. This holds true even for the smaller sample sizes, although these results are not shown in the summary figures here. We found that for almost all configurations of sample size, ecf sampling density, and number of wavelet coefficients retained, there is an optimal cf support range of around 1 or 2. This likely reflects the fact that the most useful information is concentrated around the origin of the cf; there also appears to be some trade-off concerning the density of wavelet coefficients around this origin.

Finally, we applied our algorithms to a real data set and summarize these results here. The data set, downloaded from `http://chart.yahoo.com`, consists of the Standard and Poors 500 ('S&P500') daily stock market closing prices index from January 1950 to December 1999 inclusive. Interest focussed on exploring the tails of the daily log-return series with a view of examining if and how the high tail differs from the low tail, and how both tails have evolved over the decades. Accordingly, this data set was divided into five parts, corresponding to the decades 1950–1959, 1960–1969, 1970–1979, 1980–1989, and 1990–1999. Within each decade, the positive and negative log-returns were separated (zero returns were discarded), and within each of the resulting 10 data sets the signs of the returns were randomized so as to produce essentially symmetrical distributions representative of the various upper and lower distributional tails. The resulting 'lower tail data sets' consisted, on average, of fewer observations than the corresponding 'upper tail data sets'; these sample sizes are shown in Table 1. Although we note this disparity, we do not take it further into account below.

Because interest focussed on tails of the distributions, we applied our wavelet procedures only to selected segments of the empirical characteristic functions near the origin. (Although the distributions resulting here will not be exactly stable, it is expected that useful estimates will result in view of analytical connections between the decay of density in the tails and behaviour of the cf around the origin.) The resulting estimates for α using cf supports $[-a, a]$ with $a = 0.5$, 1.0, and 2.0

Table 1 Sample sizes for the S&P500 data sets

S&P500 data	Positive values	Negative values
1950–1959	1,407	1,119
1960–1969	1,320	1,139
1970–1979	1,277	1,230
1980–1989	1,334	1,187
1990–1999	1,354	1,171

Table 2 Stability index α estimates for the S&P500 data sets

S&P500 Data	Koutrouvelis	McCulloch	Wavelets 0.5	Wavelets 1.0	Wavelets 2.0
1950–1959 pos	1.8748	1.8559	1.8882	1.8502	1.8454
1950–1959 neg	1.7025	1.5641	1.8041	1.6743	1.6368
1960–1969 pos	1.7632	1.7956	1.7493	1.7416	1.7422
1960–1969 neg	1.6765	1.5060	1.7878	1.6570	1.6096
1970–1979 pos	1.7750	1.5962	1.8307	1.7408	1.7122
1970–1979 neg	1.8788	1.6785	1.9383	1.8602	1.8538
1980–1989 pos	1.7930	1.6110	1.8776	1.7873	1.7465
1980–1989 neg	1.7337	1.5566	1.7728	1.7140	1.6713
1990–1999 pos	1.7911	1.5798	1.8609	1.7720	1.7420
1990–1999 neg	1.6152	1.4477	1.8185	1.6392	1.5652

are given in Table 2. The cf scaling was based on first standardizing the returns by dividing by their interquartile ranges; we remark that (with this scaling) the ecf functions appeared to follow the cfs for stable distributions very closely up to approximately $t = \pm 1.5$ only. The estimates shown in Table 2 are each based on 64 sample points and 64 wavelet coefficients using the Haar wavelet. As a general rule, it may be seen that the stability index estimates tend to be somewhat higher when smaller cf supports are used; the consistency of this variation in the estimates is indicative of the fact that stable distributions provide only an approximate fit to this data. (Similar effects were also noted in the estimates of scale, but to conserve space these results are not included here.) In terms of discernible patterns of variation, the decade 1970–1979 stands out, both for having a disproportionately equal number of up versus down market days (compared to the other four decades), as well as for having a higher index for the down versus the up days. This stands in contrast to the clear separations between stability indices in the four other decades, where the up days correspond to higher index values than the down days. (In fact the 1970–1979 decade was also the worst in terms of the overall performance of the S&P500 index.)

6 Discussion

Our purpose, in this paper, was to demonstrate that wavelet-based methods (which hitherto have been used primarily for nonparametric applications) can also be used effectively in certain parametric problems. We have done this by showing how wavelets can be used to obtain estimates of the parameters of the stable laws, and we have seen that wavelet-based methods in this context lead to procedures which are highly efficient, robust, and also competitive with other ad hoc methods that have been devised for this family. There remain substantial avenues for further research.

We have seen that the computational considerations in using wavelets are, in principle, straightforward as well as interesting numerically. A key advantage of wavelet transformation is the manner in which it disbalances data into new 'views' and renders correlation matrices into more quasi-diagonal structure. It appears that wavelet transformation also often results in rearrangement of Fisher information into patterns of more parsimonious concentration. This allows, for example, to carry out inference using reduced numbers of coefficients with less efficiency loss,

to reduce ill-posedness of regression-based methods, and may sometimes even allow us to replace generalized least squares by ordinary least squares. In fact it was often found that little statistical efficiency was lost when only small subsets of wavelet coefficients were used (relative to using all coefficients). Further, ad hoc selection of which coefficients to use in such procedures can often be made in an intuitively more straightforward way using wavelets.

We have also argued that either a wavelet basis or its 'dual' consisting of its Fourier transforms can be used to specify the 'moments' for a generalized regression estimation procedure. While it seems fair to expect that which of these is used will result in different numerical and statistical properties, the two approaches are not equally simple to implement. In particular the computation of empirical wavelet coefficients seems to be most easily carried out by applying wavelet transformation to the empirical characteristic function (unless there is a closed form for the Fourier transform of the wavelet function). Numerically, only one of these approaches has been investigated here. An interesting question is whether an efficient algorithm (such as a pyramid-type algorithm) can be devised for evaluating inner products relative to such Fourier dual bases.

Finally we remark that continued efforts seem warranted to implement reliable MLE algorithms across multiple platforms for estimating the parameters of the stable laws (in both symmetric as well as nonsymmetric cases), especially algorithms that perform reliably for small values of α where MLE procedures have been notoriously difficult to implement. We also found it interesting to observe that the procedure of Koutrouvelis (1980), while not fully efficient, performed resiliently across a wide spectrum of parameter values.

Acknowledgements The research of AF is supported by NSERC, Canada; AF thanks the LMC-IMAG of the Université Joseph Fourier and INRIA, Grenoble, for their hospitality and research facilities provided. PG is currently on leave at IST-ISR, Av. Rovisco Pais, 1, 1049-001 Lisbon, Portugal. The authors thank F. Abramovitch, U. Amato, B. Vidakovics, and the organizers and participants of the workshop on Wavelets and Statistics in Vico Equense for their role in fostering this collaboration.

References

Abramovich, F., Bailey, T. C., Sapatinas, T. (2000). Wavelet analysis and its statistical applications. *The Statistician*, *49*, 1–29.

Antoniadis, A. (1997). Wavelets in statistics: a review (with discussion). *Journal of Italian Statistical Society*, *6*, 97–144.

Antoniadis, A., Grégoire, G., McKeague, I. (1994). Wavelet methods for curve estimation. *Journal of American Statistical Association 89*(428), 1340–1353.

Antoniadis, A., Grégoire, G., Nason, G. (1999). Density and hazard rate estimation for right-censored data using wavelet methods. *Journal of Royal Statistical Society*, *B 61*, 63–84.

Bruce, A. G., Gao, H. -Y. (1994). *S+Wavelets, users manual*. Seattle: StatSci.

Buckheit, J. B., Donoho, D. (1995). Wavelab and reproducible research. In: A. Antoniadis, G. Oppenheim (Eds.) *Wavelets and statistics*. Lecture Notes in Statistics (vol. 103). Berlin Heidelberg New york: Springer.

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

Chen, Y. (1991). *Distributions for Asset Returns.* PhD Thesis, Department of Economics, SUNY-Stony Brook.

Christof, G., Wolf, G. (1992). *Convergence theorems with a stable limit law.* Berlin: Akademie Verlag.

Chui, K. (1992). *Wavelets: A tutorial in theory and applications*. Boston: Academic.
Daubechies, I. (1992). *Ten lectures on wavelets*. CBMS-NSF Regional Conferences Series in Applied Mathematics. Philadelphia: SIAM.
Donoho, D. L., Johnstone, I. M., Kerkyacharian, G., Picard, D. (1996). Density estimation by wavelet thresholding. *The Annals of Statistics 24*(2), 508–539.
DuMouchel, W. H. (1971). *Stable distributions in statistical inference.* PhD Dissertation, Yale University.
DuMouchel, W. H. (1973a). On the asymptotic normality of maximum likelihood estimates when sampling from a stable distribution. *The Annals of Statistics*, *1*, 948–957.
DuMouchel, W. H. (1973b). Stable distributions in statistical inference. I: Symmetric stable distributions compared to other symmetric long-tailed distributions. *Journal of American Statistical Association*, *68*, 469–482.
DuMouchel, W. H. (1975). Stable distributions in statistical inference. II: Information from stably distributed samples. *Journal of American Statistical Association*, *70*, 386–393.
DuMouchel, W. H. (1983). Estimating the stable index α in order to measure tail thickness: a critique. *The Annals of Statistics*, *11*, 1019–1031.
Fama, E. (1965). The behaviour of stock prices. *Journal of Business*, *38*, 34–105.
Fama, E., Roll, R. (1968). Some properties of symmetric stable distributions. *Journal of American Statistical Association*, *63*, 817–836.
Fama, E., Roll, R. (1971). Parameter estimates for symmetric stable distributions. *Journal of American Statistical Association*, *66*, 331–338.
Feller, W. (1971). *An introduction to probability theory and its applications.* (Vol. 2, 2nd ed). New York: Wiley.
Feuerverger, A., McDunnough, P. (1981a). On some Fourier methods for inference. *Journal of American Statistical Association*, *76*, 379–387.
Feuerverger, A., McDunnough, P. (1981b). On efficient inference in symmetric stable laws and processes. In: A. K. Md. E. Saleh, et al. (Eds.), *Proceedings of the International Symposium on Statistics and Related Topics, Ottawa, May 1980*, (pp. 109–122). Amsterdam: North Holland.
Feuerverger, A., McDunnough, P. (1984). On statistical transform methods and their efficiency. *Canadian Journal of Statistics*, *12*, 303–317.
Flandrin, P. (1992). Wavelet analysis and synthesis of fractional Brownian motion. *IEEE Transactions on Information Theory*, *38*, 910–917.
FracLab. (2002). INRIA Project Fractales. Open source freeware distributed by INRIA-Fractales. http://www-rocq.inria.fr/fractales.
Gao, H.-Y. (1993). *Wavelet estimation of spectral densities in time series analysis.* PhD Thesis, University of California, Berkeley.
Kogon, S. M., Williams, D. B. (1998). Characteristic function based estimation of stable distribution parameters. In: R. J. Adler, R. E. Feldman, M. S. Taqqu, (Eds.), *A practical guide to heavy tails.* (pp. 311–335). Boston: Birkhauser.
Gnedenko, V. B., Kolmogorov, A. N. (1954). *Limit distributions for Sums of Independent Random Variables*. Reading: Addison-Wesley.
Hall, P., Patil, P. (1995). On wavelet methods for estimating smooth functions. *Bernoulli 1*, 41–58.
Hill, B. M. (1975). A simple general approach to inference about the tail of a distribution. *The Annals of Statistics*, *3*, 1163–1174.
Holschneider, M. (1995). *Wavelets: An analysis tool.* Oxford: Clarendon Press.
Ibragimov, I. A., Linnik, Yu. V. (1971). *Independent and stationary sequences of random variables*. Groningen: Wolters-Noordhoff.
Janicki, A., Weron, A. (1994). *Simulation and chaotic behaviour of α-stable stochastic processes.* New York: Marcel Dekker.
Johnstone, I. M., Kerkyacharian, G., Picard, D. (1992). Estimation d'une densité de probabilité par méthode d'ondelettes. *Comptes Rendus de l'Academie Sciences, Paris A*, *315*, 211–216.
Koutrouvelis, I. A. (1980). Regression-type estimation of the parameters of stable laws. *Journal of American Statistical Association*, *75*, 918–928.
Koutrouvelis, I. A. (1981). An iterative procedure for the estimation of the parameters of the stable law. *Communications in Statistics Part B–Simulation and Computation*, *10*, 17–28.
Lukacs, E. (1970). *Characteristic functions*. Connecticut: Hafner.
Mallat, S. G. (1989). A theory for multiresolution signal decomposition: the wavelet representation. *IEEE Transactions on Pattern Analysis and Machine Intelligence 11*, 674–693.
Mallat, S. G. (1998). *A wavelet tour of signal processing.* San Diego: Academic.

Mandelbrot, B. B. (1963). The variation of certain speculative prices. *Journal of Business*, *36*, 394–419

Mandelbrot, B. B. (1972). The variation of certain speculative prices. *Journal of Business*, *45*, 542–543

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

McCulloch, J. H. (1998) Linear regression with stable disturbances. In: R. Adler, R. E. Feldman, M. S. Taqqu, (Eds.), *A practical guide to heavy tails: Statistical techniques and applications* (pp. 359–376). Boston: Birkhäuser.

Mittnik, M., Rachev, S. (1993). Modeling asset returns with alternative stable distributions. *Econometric Review*, *12*, 261–330.

Mittnik, M., Rachev, S., eds. (2000). *Stable Paretian models in finance.* New York: Wiley.

Nason, G. J., Silverman, B. W. (1994). The discrete wavelet transform in S. *Journal of Computational and Graphical Statistics 3*, 163–191.

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

Nolan, J. (2001). Maximum likelihood estimation and diagnostics for stable distributions. In: O. E. Barndorff-Nielson, T. Mikosch, S. I. Resnick (Eds.), *Lévy processes* (pp. 379–400). Boston: Birkhauser

Ogden, T. R. (1997). *Essential wavelets for statistical applications and data analysis*. Basel: Birkhäuser.

Paulson, A. S., Halcomb, E. W., Leitch, R. A. (1975). The estimation of the parameters of the stable laws. *Biometrika*, *62*, 163–170.

Press, S. J. (1972). Estimation in univariate and multivariate stable distributions. *Journal of American Statistical Association*, *67*, 842–846.

Samorodnitsky, G., Taqqu, M. (1994). *Stable non-Gaussian random processes: Stochastic models with infinite variance.* New York: Chapman & Hall.

Strang, G. (1986). *Introduction to Applied Mathematics*. Cambridge: Addison-Wellesley.

Uchaikin, V. V., Zolotarev, V. M. (1999). *Chance and Stability: Stable Distributions and their Applications.* Utrecht: VSP.

Vidakovic, B. (1999). *Statistical modeling by wavelets.* New York: Wiley.

Zolotarev, V. M. (1986). *One-dimensional stable distributions.* Translations of Mathematical Monographs (Vol. 65). Providence, Rhode Island: American Mathematical Society (Translation of the original Russian edition of 1983).