Wavelet Based Estimation for Univariate Stable Laws

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

Andrey Feuerverger
Department of Statistics, University of Toronto,
Toronto, Ontario, M5S 3G3 Canada

and

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

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 from stable distributions 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.

Key words and phrases: Wavelets; Stable Laws; Fourier transform; completeness; empirical characteristic function; estimation; Fisher Information; efficiency; generalized least squares; regularization; QR decomposition.

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, approximation, and computational properties which are not only remarkable, but which give rise to applications of wavelet-based methods in many areas of science and engineering. A key underlying characteristic in many of these applications is the ability of wavelet functions to capture effectively the local features of the processes being modeled. Hence, in the field of statistics for example, wavelets have been used primarily to deal with problems of a nonparametric character, such as those arising frequently 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, 1997), Donoho *et al.* (1996), Gao (1993), and Johnstone *et al.* (1992). This leaves open a question which is a natural one 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 the unknown parameter values in a parametric statistical model, given a sample of data generated according to that model? It is to this question that the present paper is addressed, and we shall argue that the answer is in the affirmative.

Specifically, in this paper we will demonstrate the effectiveness of estimation procedures based on wavelet functions to estimate the parameters of the Stable Laws. The Stable Laws are a natural distribution 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 various branches of physics, but also because the Stable Laws do not have closed forms for their densities so that the classical inference procedures are generally quite 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 such parametric inference. These techniques are complementary to maximum likelihood estimation, and are often applicable in situations (e.g. in convolution families) where the density functions may not have a tractable form. These new wavelet-based methods are highly efficient, and inherently robust. In particular, they allow one to deal effectively with distribution families that possess heavy tails. These wavelet transformation methods also have the property of 'disbalancing' certain correlation structures and of rearranging the (Fisher) Information of data in ways that can prove helpful in the context of nearly singular statistical models. A key novelty in this is that wavelets are being used here to solve a *parametric* problem instead of a *nonparametric* one, the latter being the more typical in applications of wavelet methods).

This paper is organized as follows. In Section 2 we give a brief overview of the Stable Laws as well as of some recent methods that have been proposed to estimate the parameters in such families. Section 3 reviews some key properties of wavelets and provides the definitions and background on wavelet analysis required for the proposed methodology. Our main wavelet-based parametric estimation procedures are introduced and developed in Section 4. In particular we explore there 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 Section 5 which deals with the algorithmic details of our estimation procedures. In Section 6 we summarize the results of numerical experimentation and report on a small simulation study intended to assess the performance of the proposed methods in finite sample situations, thus shedding some light on the results obtained in Section 4. These simulations also allow us to compare other different techniques to those

we propose. 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 of the proposed methodology, and concluding remarks are collected together in Section 7.

2 The Stable Laws.

The univariate stable distributions 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, specifically, 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 $\alpha = 2$, the distribution corresponds to a normal distribution with mean μ and variance $2\sigma^2$, while when $\alpha < 2$, the variance of the distribution becomes infinite and the tails of the density then have order $\mathcal{O}(|x|^{-1-\alpha})$ as $|x| \to \infty$. This characteristic of the stable distributions (together with their 'central limit' origin) makes them useful in modeling certain types of data that admit observations of very large magnitude, while the parameter β allows for incorporation of skewness. See, for example, Christof & Wolf (1992), Feller (1971), Gnedenko & Kolmogorov (1954), Ibragimov & Linnik (1971), Lukacs (1970), Samorodnitsky & Taqqu (1994) and Zolotarev (1986). A wide range of applications of the stable laws in various data analytic contexts are discussed, for example, in Cappé *et al.* (2002), Fama (1965), Fama & Roll (1968, 1971), Janicki & Weron (1994), Mandelbrot (1963), Mittnick & Rachev (1993, 2000), and Uchaikin & Zolotarev (1999).

Except in the cases $(\alpha, \beta) = (2, \cdot), (1, 0)$ or $(1/2, \pm 1)$, the probability density function of a stable random variable is not available in closed form, which makes inference by standard statistical techniques, such as maximum likelihood, difficult, if not impossible.

This lack of closed analytic expressions for the probability density functions 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

$$\log \phi(t) = \begin{cases} i\mu t - |\sigma t|^{\alpha} \left\{1 - i \tan\left(\frac{\pi}{2}\alpha\right) \beta' \operatorname{sgn}(t)\right\} & \text{when } \alpha \neq 1 \\ i\mu t - |\sigma t| \left\{1 + i\frac{2}{\pi}\beta' \log|t| \operatorname{sgn}(t)\right\} & \text{when } \alpha = 1 \end{cases} \tag{2.1}$$

where $-1 \leq \beta' \leq 1$ is also a measure of skewness and related to β through

$$\beta' = \cot\left(\frac{\pi}{2}\alpha\right) \tan\left\{\frac{\pi}{2}K(\alpha)\beta\right\}, \tag{2.2}$$

where $K(\alpha) = 1 - |1 - \alpha|$. For $\alpha \neq 1$ the alternative representation

$$\log \phi(t) = -|t|^{\alpha} \exp\left\{-i\frac{\pi}{2}K(\alpha)\beta \operatorname{sgn}(t)\right\}, \tag{2.3}$$

is often more convenient. 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. For convenience of notation we consider only standard stable distributions in the remainder of this section.

For purposes of statistical applications, the canonical representation (2.1) has one disagreeable feature, namely that the $\phi(t)$ are not continuous functions of the parameters, since a discontinuity in the parameter space occurs when α and $\beta \neq 0$ vary continuously across the line $\alpha = 1$. However this discontinuity is of an inessential kind and may be removed. For example, following Feuerverger and McDunnough (1981b), if we write the $\alpha \neq 1$ term of (2.1) as

$$-|t|^{\alpha} - i\beta' t|t|^{\alpha-1} \tan\left(\frac{\pi}{2}\alpha\right)$$

and shift mean by an amount $\beta' \tan\left(\frac{\pi}{2}\alpha\right)$ we then obtain

$$-|t|^{\alpha} - i\beta' t\left(|t|^{\alpha-1} - 1\right) \tan\left(\frac{\pi}{2}\alpha\right)$$

or

$$\log \phi(t) = -|t|^{\alpha} - i\beta^{*} h(t, \alpha) \tag{2.4}$$

where

$$h(t,\alpha) \;=\; \frac{t\,(|t|^{\alpha-1}-1)}{\alpha-1} \tag{2.5}$$

and

$$\beta^* \;=\; \beta'(\alpha-1)\tan\left(\frac{\pi}{2}\alpha\right) \;=\; (1-\alpha)\tan\left(\frac{\pi}{2}K(\alpha)\beta\right)\,. \tag{2.6}$$

The function $h(t,\alpha)$ is defined by continuity as $t\log|t|$ when $\alpha = 1$, and as 0 if also $t = 0$. The shape of this (α,β^*) parameter domain is roughly elliptical with corners at $\alpha = 0$ and $\alpha = 2$, thus better corresponding with the known behaviour of the density functions near these extremities for varying skewness. In this respect, note that for $\alpha = 2$ the density does not vary as β changes, while the extreme values for β^* at $\alpha = 1$ are given by $\pm\lim_{\alpha\to1}(\alpha-1)\tan(\pi\alpha/2) = \pm2/\pi$. The function $h(t,\alpha)$ is jointly continuous in (t,α) on the domain defined by $-\infty < t < \infty$ and $0 < \alpha \leq 2$. (This remains true if we should permit α to exceed 2, but is false if $\alpha = 0$ is included in the domain; in particular, if $\alpha \to 0$ then, at each fixed t, the function $h(t,\alpha)$ tends in limit to $t - \mathrm{sgn}(t)$ which is not continuous at $t = 0$.) The continuous parameterization described above was adopted in Zolotarev (1986); alternative continuous parameterizations were given by DuMouchel (1971) and by Chambers *et al.* (1976).

The probability density functions of stable random 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 the parameters of a stable distribution. In particular, most conventional statistical methods (such as maximum likelihood) can therefore not be used directly in this case. However, several numerical methods for estimating the stable law parameters have been considered in the literature and have been found to be useful in practice. For example, sample quantile methods were developed by Fama (1965) and Fama & Roll (1968, 1971) who 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 later 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, Holcomb & Leitch (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 particular parametric form only for the tails of the distribution function. See also Feuerverger & McDunnough (1981a,b). For a review, see e.g., Mittnik and Rachev (1993).

Using a multinomial approximation to the likelihood function, DuMouchel (1971, 1973) was the first to obtain approximate ML estimates for α and σ (assuming $\mu = 0$); DuMouchel (1973) 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. However, for this to hold it is necessary to restrict $\alpha > \epsilon$ for some fixed $\epsilon > 0$, because the stable likelihood function is unbounded in the neighbourhood of $\alpha = 0$. In accordance with the usual theory for the MLE, it is also necessary that the true value of the parameters be interior to the parameter domain, and in particular away from the extreme β boundaries. Although DuMouchel also requires that the parameter values stay away from $\alpha = 1$, his proof appears to carry over to the continuous reparameterization discussed above. Maximum likelihood procedures for estimating stable parameters were also implemented by Feuerverger & McDunnough (1981b), McCulloch (1986), Chen (1991) and Nolan (1997, 2001).

To our knowledge, until now, no sample characteristic function based method combined with the use of wavelets as a tool for estimating the parameters in various subfamilies of the stable laws, has been explored in the literature and this is the approach adopted in this paper.

3 Some Background on Wavelets.

Since the estimation procedures we shall develop for the parameters of a stable distribution are based on wavelet decompositions, we recall briefly here some relevant facts about wavelets. For detailed expositions of the mathematical aspects of wavelets we refer, for example, to Meyer (1990), Daubechies (1992), Mallat (1998), Chui (1992), Cohen & Ryan (1995) or Holschneider (1995), while comprehensive expositions on the use of wavelets in statistical settings are given, for example, in Abramovich *et al.* (2000), Antoniadis (1997), Antoniadis *et al.* (2001), Ogden (1997) and Vidakovic (1999). To simplify notation, we will consider here 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.

Wavelet analysis thus requires a description of two basic functions, a *scaling function* $\varphi(x)$ and a *wavelet* $\psi(x)$. The function $\varphi(x)$ is a solution of a two-scale difference equation

$$\varphi(x) = \sqrt{2} \sum_{k \in \mathbb{Z}} h_k \, \varphi(2x - k) \tag{3.1}$$

with normalization $\int_{\mathbb{R}} \varphi(x)\, dx = 1$, and the function $\psi(x)$ is then defined by

$$\psi(x) = \sqrt{2} \sum_{k \in \mathbb{Z}} (-1)^k h_{1-k} \, \varphi(2x - k). \tag{3.2}$$

The h_k are called the *filter coefficients*, and it is through careful choice of these that wavelet functions with desirable properties can be constructed.

A wavelet system is the infinite collection of translated and scaled versions of φ and ψ defined by:

$$\varphi_{j,k}(x) = 2^{j/2} \varphi(2^j x - k), \quad j, k \in \mathbb{Z} \tag{3.3}$$

$$\psi_{j,k}(x) = 2^{j/2} \psi(2^j x - k), \quad j, k \in \mathbb{Z}\,. \tag{3.4}$$

Certain 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 that $\{\psi_{j,k},\ j, k \in \mathbb{Z}\}$ is an orthonormal basis of $L^2(\mathbb{R})$. 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, where $\langle \cdot, \cdot \rangle$ is the usual inner product in $L^2(\mathbb{R})$.

A key observation of Daubechies (1992) is that it is possible to construct finite-length sequences of filter coefficients satisfying all the required conditions, and resulting in compactly supported φ and ψ functions that have space-frequency localization. It is this localization that allows parsimonious representations for a wide class of different functions in wavelet series. These derived wavelet bases are well localized in space since the energy of the wavelet is restricted to a finite interval. Frequency localization simply means that the Fourier transform of a wavelet is localized, i.e., mostly contains frequencies from a limited frequency band. The Heisenberg uncertainty principle puts a lower bound on the product of the space and frequency variances.

Decay of its Fourier transform towards high frequencies corresponds to smoothness of the function – the smoother the function, the faster the decay. If the decay is exponential, the function is infinitely differentiable. Decay towards low frequencies corresponds to the number of vanishing moments of the wavelet. A wavelet ψ has N vanishing moments when $\int_{\mathbb{R}} x^p \, \psi(x) \, dx = 0$ for $0 \leq p < N$. The number of vanishing moments relates the order of approximation of a function f to its smoothness. For example, assuming that the wavelet ψ has N vanishing moments, then for all $s \in (0, N]$, the approximation of a C^s function f by its projection onto any approximation space V_M (as defined below) satisfies $\|f - P_{V_M} f\|_\infty \leq C M^{-s}$. Results of this type are detailed in Daubechies (1992).

The wavelet representation of a function $g \in L^2(\mathbb{R})$ is given by

$$g = \sum_{j \in \mathbb{Z}} \sum_{k \in \mathbb{Z}} d_{j,k} \, \psi_{j,k} \tag{3.5}$$

where the wavelet coefficients are given by $d_{j,k} = \int_{\mathbb{R}} g(t) \psi_{j,k}(t) dt$. Typically, we want algorithms with linear or linear-logarithmic complexity to traverse between a function g and its wavelet coefficients. Such an algorithm is referred to as a *fast wavelet transform*. Fast wavelet transforms are often obtained through multiresolution analysis, a framework developed by Mallat (1989), in which the wavelet coefficients $\langle g, \psi_{j,k} \rangle$ of a function g, for a

fixed j, describe the difference between two approximations of g, one having resolution 2^j, and one with the coarser resolution 2^{j-1}. A multiresolution analysis (or approximation) of $L^2(\mathbb{R})$ consists of a nested sequence of closed subspaces V_j, $j \in \mathbb{Z}$, of $L^2(\mathbb{R})$

$$\cdots \subset V_{-2} \subset V_{-1} \subset V_0 \subset V_1 \subset V_2 \subset \cdots \tag{3.6}$$

having the following properties: they have intersection that is trivial and union that is dense in $L^2(\mathbb{R})$:

$$\cap_j V_j = \{0\}, \qquad \overline{\cup_j V_j} = L^2(\mathbb{R})\,; \tag{3.7}$$

they are dyadic dilates of one another:

$$f(x) \in V_j \Leftrightarrow f(2x) \in V_{j+1}\,; \tag{3.8}$$

and there exists a *scaling* function $\phi \in V_0$ whose integer translates span V_0:

$$V_0 = \left\{ f \in L^2(\mathbb{R}) :\ f(x) = \sum_{k\in\mathbb{Z}} \alpha_k \varphi(x-k) \right\}. \tag{3.9}$$

The approximation space V_0 is considered to have resolution 1. An orthonormal basis of V_j, the approximation space with resolution 2^j, is then given by the family $\{\varphi_{j,k} :\ k \in \mathbb{Z}\}$. The orthogonal projection of a function $f \in L^2(\mathbb{R})$ onto V_j is given by

$$P_j f = \sum_{k\in\mathbb{Z}} \langle f, \varphi_{j,k}\rangle\, \varphi_{j,k}\,, \tag{3.10}$$

and can be thought of as an approximation of f having resolution 2^j.

Defining W_j as the orthogonal complement of V_j in V_{j+1}, we obtain another sequence, $\{W_j, j \in \mathbb{Z}\}$, of closed mutually orthogonal subspaces of $L^2(\mathbb{R})$ such that each W_j is a dilate of W_0, and their direct sum is $L^2(\mathbb{R})$. The space W_0 is spanned by the integer translates of the wavelet function ψ which is related to φ through equations (3.1) and (3.2).

According to the above, if g is a square integrable function, then it can also be represented as

$$g = \sum_{k\in\mathbb{Z}} c_{j_0,k}\, \varphi_{j_0,k} + \sum_{j\geq j_0} \sum_{k\in\mathbb{Z}} d_{j,k}\, \psi_{j,k} \tag{3.11}$$

where j_0 represents some "coarse" level of approximation, $c_{j,k} = \langle g, \varphi_{j,k} \rangle$, and $d_{j,k} = \langle g, \psi_{j,k} \rangle$. The first term on the right in (3.11) is the projection $P_{j_0}\, g$ of g onto the coarse approximating space V_{j_0}, and the second term represents the successive details of g at the higher approximation levels. The main properties, and the key to applications of wavelets and multiresolution analyses are their powerful approximating qualities, i.e., they provide accurate approximations of a function using only a relatively small number of coefficients.

Finally, we add a few remarks about computational algorithms and the discrete wavelet transform. If the projection of a square integrable function g onto a fine multiresolution space V_n is known, it can be written as

$$P_n\, g = \sum_{k \in \mathbb{Z}} c_{n,k}\, \varphi_{n,k}\,. \tag{3.12}$$

Given some lower resolution $J_0 < n$, the projection $P_n\, g$ can be decomposed as

$$P_n\, g = \sum_{k \in \mathbb{Z}} c_{J_0,k}\, \varphi_{J_0,k} + \sum_{j=J_0}^{n} \sum_{k \in \mathbb{Z}} d_{j,k}\, \psi_{j,k}\,. \tag{3.13}$$

As a consequence of (3.1), (3.2) and the multiresolution analysis structure, given the V_n coefficients $c_{n,k}$ in (3.12), we may obtain the $c_{J_0,k}$ and $d_{j,k}$ in (3.13) by means of the recursive formulas

$$c_{j-1,k} = \sum_m h_{m-2k}\, c_{j,m}\,, \quad d_{j-1,k} = \sum_m g_{m-2k}\, c_{j,m}\,, \quad j = n, \ldots J_0 + 1, \tag{3.14}$$

where $g_k = (-1)^k h_{1-k}$. The above computations can be summarized in the following way: Let $\mathbf{f} = (f_1, \ldots, f_n)$ be an element of the Hilbert space $\ell_2(n)$ of all square summable sequences of length n. Then the discrete wavelet transform of $\mathbf{f}$ is an $\ell_2(n)$ sequence

$$\beta = (G\mathbf{f}, GH\mathbf{f}, \ldots, GH^{J_0-1}\mathbf{f}, H^J\mathbf{f})$$

where H and G are operators from $\ell_2(2M)$ to $\ell_2(M)$ defined coordinate-wise via

$$\text{for } a \in \ell_2(2M), \quad (Ha)_k = \sum_m h_{m-2k} a_m, \quad (Ga)_k = \sum_m g_{m-2k} a_m\,.$$

(Here M is the length of the filter sequence h.)

The discrete wavelet transformation (DWT) described above is linear and orthonormal and can be represented in matrix form. Given the lowpass filter coefficients $\{h_k\}$, one can thus write the DWT-transformation matrix $\mathcal{W}_{n,J_0}$ explicitly, and in terms of $\mathcal{W}_{n,J_0}$ we have $\beta = \mathcal{W}_{n,J_0}\,\mathbf{f}$ and $\mathbf{f} = \mathcal{W}'_{n,J_0}\,\beta$. More explicitly, assume the convention that the data is given at the highest scale V_J with $(n = 2^J)$ so that $V_J = V_{J_0} \oplus W_{J_0} \oplus \cdots \oplus W_{J-1}$. Then given a vector of sampled values $\mathbf{f}$ of length $n = 2^J$ (this explains the first index n in $\mathcal{W}_{n,J_0}$) the components of $\mathbf{f}$ are thought of as the scaling coefficients of a function in V_J. The matrix $\mathcal{W}_{n,J_0}$ applied to $\mathbf{f}$ then produces the vector β of the same length in such a way that the first 2^{J_0} components of β are the corresponding scaling coefficients of the projection of $\mathbf{f}$ into V_{J_0}, The next block of 2^{J_0} components are the wavelet coefficients of the projection of $\mathbf{f}$ into W_{J_0} and so on. (This explains the second index J_0 in the notation $\mathcal{W}_{n,J_0}$.) Finally the rows of $\mathcal{W}_{n,J_0}$ are constructed via the scaling and wavelet filters. Note that if $n = 2^J$ for some positive J, then both the DWT and the inverse DWT may be performed using Mallat's (1989) fast algorithm which requires only $\mathcal{O}(n)$ operations. This algorithm is available in several standard implementations, for example in the `S-Plus` packages `WaveThresh` (Nason & Silverman, 1994) or `S+Wavelets` (Bruce & Gao, 1994), or in the `MatLab` packages `WaveLab` (Buckheit *et al.*, 1995) or `FracLab` (INRIA Project Fractales). In particular, the computations described in this paper were based primarily on the `FracLab` package.

4 Wavelet Estimation Procedures.

We represent the parameters of the stable laws by means of the column vector $\theta = (\alpha, \beta, \mu, \sigma)'$, and the parameter space by Θ, for some characteristic function representation for the stable laws; as indicated previously, we generally prefer a continuous parameterization of the full stable family. The density, cumulative distribution function, and characteristic function (cf) will be denoted as $f_\theta(x)$, $F_\theta(x)$, and $c_\theta(t)$, respectively. Data will be denoted as $x_1, x_2, \ldots, x_n$, these being realizations of independent and

identically distributed random variables $X_1, X_2, \ldots, X_n$ sampled according to some unknown parameter value $\theta = \theta_0$. The empirical distribution function and empirical characteristic function (ecf) corresponding to this 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

$$\mathcal{E}_\theta\, g_t(X) \equiv G_\theta(t) \equiv \int g_t(x)\, dF_\theta(x)\,, \tag{4.1}$$

and

$$\mathcal{E}_n\, g_t(X) \equiv G_n(t) \equiv \int g_t(x)\, dF_n(x) = \frac{1}{n}\sum_{j=1}^{n} g_t(X_j)\,. \tag{4.2}$$

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 true score function $\partial \log f_\theta(x)/\partial\theta_0$. 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 if in fact 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 (wavelet or otherwise) 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 $\{\hat{\psi}_{j,k}(t);\, j,k \in \mathbb{Z}\}$, respectively. In view of Parseval's identity, both of 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, 1973). 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 the first instance) the nonlinear generalized least squares fitting, in obvious notation, of

$$\langle dF_n, \hat{\psi}_{j,k} \rangle = \langle f_\theta, \hat{\psi}_{j,k} \rangle + \epsilon_{j,k}\,, \tag{4.3}$$

or equivalently of

$$\langle dF_n, \hat{\psi}_{j,k} \rangle = \langle c_\theta, \psi_{j,k} \rangle + \epsilon_{j,k}\,, \tag{4.4}$$

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 (4.4) as

$$Y_{j,k} \equiv \langle dF_n, \hat{\psi}_{j,k} \rangle \equiv \frac{1}{n} \sum_{\ell=1}^{n} \hat{\psi}_{j,k}(X_\ell) \tag{4.5}$$

allows us to exhibit (4.4) in standard least squares format:

$$Y_{j,k} = \langle c_\theta, \psi_{j,k} \rangle + \epsilon_{j,k}\,. \tag{4.6}$$

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 (4.6) is quite simple, evaluation of the empirical wavelet coefficients as in expression (4.5) 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

$$Y_{j,k} = \langle dF_n, \hat{\psi}_{j,k} \rangle = \langle c_n, \psi_{j,k} \rangle\,. \tag{4.7}$$

In (4.7), 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 (4.5), the quantities $\langle dF_n, \hat{\psi}_{j,k} \rangle$ are typically much more difficult to compute.

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

$$Y'_{j,k} \equiv \frac{1}{n} \sum_{\ell} \psi_{j,k}(X_\ell) = \langle dF_n, \psi_{j,k} \rangle \;=\; \langle f_\theta, \psi_{j,k} \rangle + \epsilon'_{j,k} = \langle c_\theta, \hat{\psi}_{j,k} \rangle + \epsilon'_{j,k} \,. \tag{4.8}$$

In accordance with our previous discussions, these procedures will also be asymptotically arbitrarily highly efficient. However the computation of the terms $\langle c_\theta, \hat{\psi}_{j,k} \rangle$ is problematical in that no fast algorithm appears to be 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 means of Fourier inversion of the c_θ which we obviously seek to avoid in the present context. For this reason we will not pursue this second option any further below.

We next discuss the covariance matrix Σ of the error terms $\epsilon_{j,k}$ appearing in (4.3). An initial convention might be that each row (and each column) of Σ corresponds to a particular pair (j, k), i.e. to a particular wavelet function; a typical entry of Σ might therefore (in obvious notation) be denoted as $\Sigma_{(j,k),(j',k')}$. However, quantities such as $\langle dF_n, \hat{\psi}_{j,k} \rangle$ 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 (4.4) in fact represents two equations – one for the real part, and one for the imaginary part. Corresponding to this, the number of rows and columns of the covariance matrix Σ must in fact equal twice the number of wavelet coefficients being used, with the real and imaginary parts of each 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 (i.e., our 'initial convention') the entry of Σ corresponding to the

(j,k) and the (j',k') wavelets is given via

$$n \cdot \Sigma_{(j,k),(j',k')} \equiv n \cdot \mathrm{Cov}\left(Y_{j,k},\, Y_{j',k'}\right) \;=\; \mathrm{Cov}\left(\hat{\psi}_{j,k}(X), \hat{\psi}_{j',k'}(X)\right)$$

$$=\; \mathbb{E}\left(\hat{\psi}_{j,k}(X)\,\overline{\hat{\psi}_{j',k'}(X)}\right) - \mathbb{E}\left(\hat{\psi}_{j,k}(X)\right)\,\overline{\mathbb{E}\left(\hat{\psi}_{j',k'}(X)\right)}\,. \tag{4.9}$$

Now, by Parseval's indentity,

$$\mathbb{E}(Y_{j,k}) = \mathbb{E}(\hat{\psi}_{j,k}(X)) = \int \hat{\psi}_{j,k}(x)\, f_\theta(x)\, dx = \int \psi_{j,k}(x)\, c_\theta(x)\, dx\,; \tag{4.10}$$

therefore, the quantities $E\left(Y_{j,k}\right)$ required in (4.9) are just the wavelet coefficients of c_θ which can readily be computed, and which we will denote here by $\eta_{j,k}(c_\theta)$. Next,

$$E\left(\hat{\psi}_{j,k}(X)\,\overline{\hat{\psi}_{j',k'}(X)}\right) = \int \hat{\psi}_{j,k}(x)\,\overline{\hat{\psi}_{j',k'}(x)}\, f_\theta(x)\, dx$$

$$= \int\int\int \psi_{j,k}(s) e^{-isx}\,\overline{\psi_{j',k'}(t) e^{-itx}}\, f_\theta(x)\, dsdtdx$$

$$= \int\int \psi_{j,k}(s)\,\overline{\psi_{j',k'}(t)}\, c_\theta(t-s)\, dsdt\,. \tag{4.11}$$

Consequently

$$n \cdot \Sigma_{(j,k),(j',k')} = \int\int \psi_{j,k}(s)\,\overline{\psi_{j',k'}(t)}\,\left[c_\theta(t-s) - c_\theta(s)c_\theta(-t)\right]\, dsdt\,. \tag{4.12}$$

The approach taken above 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

$$n \cdot \mathrm{Cov}\left(c_n(s), c_n(t)\right) = c(s-t) - c(s)c(-t)\,. \tag{4.13}$$

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

$$n \cdot \mathrm{Cov}\left(\Re\, c_n(s), \Re\, c_n(t)\right) = \frac{1}{2}\left[\,\Re\, c(s-t) + \Re\, c(s+t)\,\right] - \Re\, c(s)\,\Re\, c(t)\,, \tag{4.14}$$

$$n \cdot \mathrm{Cov}\left(\Re\, c_n(s), \Im\, c_n(t)\right) = \frac{1}{2}\left[\,\Im\, c(s-t) + \Im\, c(s+t)\,\right] - \Re\, c(s)\,\Im\, c(t)\,, \tag{4.15}$$

and

$$n \cdot \mathrm{Cov}\,(\Im\, c_n(s), \Im\, c_n(t)) = \frac{1}{2}\,[\,\Re\, c(s-t) - \Re\, c(s+t)\,] - \Im\, c(s)\,\Im\, c(t)\,. \tag{4.16}$$

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

$$\begin{aligned}\mathrm{Cov}\ (\,\Re Y_{j,k}, \Re Y_{j',k'}\,) &= \mathrm{Cov}\ (\langle \Re c_n, \psi_{j,k}\rangle, \langle \Re c_n, \psi_{j',k'}\rangle) \\ &= \int \psi_{j,k}(s)\,\mathrm{Cov}\ (\Re c_n(s), \Re c_n(t))\,\psi_{j',k'}(t)\,dsdt\,. \end{aligned}\tag{4.17}$$

In this way we obtain the relation

$$\begin{aligned} & n \cdot \mathrm{Cov}\,(\,\Re Y_{j,k}, \Re Y_{j',k'}\,) \\ &= \int\int \psi_{j,k}(s)\,\psi_{j',k'}(t)\left\{\frac{1}{2}\,[\,\Re c(s-t) + \Re c(s+t)\,] - \Re c(s)\,\Re c(t)\right\} dsdt \\ &= \frac{1}{2}\int\int \psi_{j,k}(s)\,\psi_{j',k'}(t)\Re\,[c(s-t) + c(s+t)]\,dsdt - \eta_{j,k}(\Re\,[c_\theta])\,\eta_{j',k'}(\Re[c_\theta])\,, \end{aligned}\tag{4.18}$$

and likewise

$$\begin{aligned} & n \cdot \mathrm{Cov}\,(\,\Re Y_{j,k}, \Im Y_{j',k'}\,) \\ &= \int\int \psi_{j,k}(s)\,\psi_{j',k'}(t)\left\{\frac{1}{2}\,[\,\Im c(s-t) + \Im c(s+t)\,] - \Re c(s)\,\Im c(t)\right\} dsdt \\ &= \frac{1}{2}\int\int \psi_{j,k}(s)\,\psi_{j',k'}(t)\Im\,[c(s-t) + c(s+t)]\,dsdt - \eta_{j,k}(\Re\,[c_\theta])\,\eta_{j',k'}(\Im[c_\theta])\,, \end{aligned}\tag{4.19}$$

and

$$\begin{aligned} & n \cdot \mathrm{Cov}\,(\,\Im Y_{j,k}, \Im Y_{j',k'}\,) \\ &= \int\int \psi_{j,k}(s)\,\psi_{j',k'}(t)\left\{\frac{1}{2}\,[\,\Re c(s-t) - \Re c(s+t)\,] - \Im c(s)\,\Im c(t)\right\} dsdt \\ &= \frac{1}{2}\int\int \psi_{j,k}(s)\,\psi_{j',k'}(t)\Re\,[c(s-t) - c(s+t)]\,dsdt - \eta_{j,k}(\Im\,[c_\theta])\,\eta_{j',k'}(\Im[c_\theta])\,. \end{aligned}\tag{4.20}$$

The entries of Σ corresponding to the real and imaginary parts of the wavelet coefficients are thus given by the expressions (4.18), (4.19) and (4.20).

In the symmetric family $c_\theta(t) = \exp\{i\mu t - |\sigma t|^\alpha\}$, if we work with centered variates such as $\tilde{X}_j = (X_j - \hat{\mu})\,/\hat{\sigma}$, 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 $\mathrm{Cov}\,(\,\Re Y_{j,k}, \Im Y_{j',k'}\,)$ are null, 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 Section 5.

The regression formulation outlined above leads to wavelet based inferential procedures for the stable laws which have high asymptotic efficiency but are considerably simpler and faster to execute than maximum likelihood procedures. More precisely, the weighted least squares wavelet procedure is 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

$$\hat{\theta}_{\mathrm{new}} = \hat{\theta}_{\mathrm{old}} + (G'\,\Sigma^{-1}G)^{-1}G'\,\Sigma^{-1}\eta_{\theta_{\mathrm{old}}}. \tag{4.21}$$

(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 $\Re\,(Y_{j,k} - \eta_{j,k}(c_\theta))$ and $\Im\,(Y_{j,k} - \eta_{j,k}(c_\theta))$, and G is the $2T \times 4$ matrix of partial derivatives of $\boldsymbol{\eta}_\theta$ with respect to the four components of θ, these terms all being evaluated at $\theta = \theta_{\mathrm{old}}$.

We mention here that (4.13), as well as (4.14), (4.15) and (4.16) belong to the class of operators that are made sparse under 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 involved in the iterative updating scheme (4.21). A serious difficulty (which, in fact, occurs in many inverse problems) is that for moderate n and large T, the true and also the 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 average eigenvalue equal to 1) a cutoff value somewhere between 0.01 to 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 done in ridge regression to reduce multicolinearities in the design matrix) before proceeding to the inversion of the matrices involved 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 here about the superiority of one regularization method over another. Whatever method is chosen, we have used a QR decomposition method (see, e.g., Strang, 1986) to iteratively solve equation (4.21); 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 $Q'\,\eta_{\theta_{\rm old}} = R\,(\theta - \theta_{\rm old})$ by substitution.

5 Numerical Algorithms.

In this section we provide details of the numerical algorithms that were used in our experiments with wavelet-based regression for the α-Stable laws. In particular we outline here how the theoretical details of Section 4 (expressed there in continuous form) were implemented by means of discrete algorithms.

To simplify our numerical efforts, we experimented only with symmetric stable families, i.e. skewness parameter $\beta = 0$. An important consideration, when $\beta = 0$, is that the 4×4

Fisher Information matrix then becomes block diagonal, with 2×2 blocks corresponding to the parameter groupings (α, σ) and (β, μ) respectively (see, e.g., DuMouchel, 1975, Thm. 2). This structure allowed us to concentrate our efforts on the key index parameter α, and on the scale parameter σ thereby associated with it. In particular, we saved additional efforts by using only a $\sqrt{n}$-consistent estimator for μ throughout (and not further updating our estimates for μ). The Fisher Information parameter orthogonalities (when $\beta = 0$) imply that our estimators for α and σ may then (asymptotically, at least) be compared fairly to estimators for these two parameters that are obtained from algorithms which fit all four stable law parameters simultaneously.

As before then, denote by $\{x_j\}_{j=1}^n$ the data set to be analyzed; the x_j are independent random variables from a stable distribution f_θ, with $\theta = (\alpha, \beta, \mu, \sigma)'$. As mentioned, in our experiments we always used $\beta = 0$. Furthermore, taking advantage of statistical invariance in the location parameter, we always used $\mu = 0$. Thus, it will now be convenient to change notation so that henceforth $\theta = (\alpha, \sigma)'$ — a column vector of length two. We therefore will speak of the parameter estimates $\hat{\theta} = (\hat{\alpha}, \hat{\sigma})'$, and these we first initialized using the values $\hat{\alpha}_1$ and $\hat{\sigma}_1$ obtained by the four parameter estimation procedure of Koutrouvelis (1980), as implemented in `FracLab`, when applied to the data set $\{x_j\}$. The location parameter μ was estimated by the $\sqrt{n}$-consistent estimator $\hat{\mu}_0$, defined as the empirical mean computed over the subset of the data obtained by removing 28% of the smallest and 28% of the largest values. (See Fama & 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, 2, and 3. Each step, say the kth step, of the estimating loop, thus proceeds as follows. Data are centered (by removing the trimmed mean $\hat{\mu}_0$) and then renormalized using the current estimate $\hat{\sigma}_k$ of the scale parameter. Denote by $\{\tilde{x}_j\}_{j=1}^n$ this normalized data set, where $\tilde{x}_j = (x_j - \hat{\mu}_0)/\hat{\sigma}_k$. We

then compute the empirical characteristic function

$$c_n(s_\ell) = \frac{1}{n}\sum_{j=1}^{n} \exp\{is_\ell \tilde{x}_j\} \tag{5.1}$$

over an interval $[-S, S)$, using the sampled gridpoints $s_\ell = \pm\delta\ell$, where $\ell = 0, \ldots, M$, and the sampling rate is $\delta = S/M$. The ecf c_n needs to be computed only for positive values of s_ℓ, and is completed for negative values using conjugate symmetry,

We next compute the analytic characteristic function c_θ using the currently estimated value of θ, namely $\hat{\theta}_k = (\hat{\alpha}_k, \hat{\sigma}_k)'$; here β is, of course, implicitly equal to 0, while μ is also set equal to 0 in view of the fact that the data have been centered at $\hat{\mu}_0$. Analytic forms of the derivatives of the characteristic function c_θ w.r.t. α, σ and μ are then also computed over the same grids and evaluated at the current parameter values; let us denote these quantities as $\partial_\alpha c_\theta$, $\partial_\sigma c_\theta$ and $\partial_\mu c_\theta$ respectively. These quantities are column vectors of dimension $2M$ Here too, computations only need to be carried out on the positive axis, and then completed by symmetry. All quantities indicated here are also computed over the interval $[-S, S)$, and with gridpoints of form $s_\ell = \delta\ell$ as before, but in order to later obtain the required covariance structures, the values of c_θ must actually be computed over a support $[-2S, 2S]$ that is twice as large, but using the same grid spacing. No tapering of these quantities was required.

We then computed the full tree wavelet decompositions for each of the vectors defined above. Specifically, we computed $W[c_\theta]$, $W[\partial_\alpha c_\theta]$, $W[\partial_\sigma c_\theta]$, and $W[\partial_\mu c_\theta]$ where the wavelet

operator $W[\cdot]$ acts on a finite signal to give

$$\begin{aligned} W[\cdot] \;=\; & \big(d_{-1,0}, d_{-1,1}, d_{-1,2} \ldots, d_{-1,(2M)\cdot 2^{-1}-2}, d_{-1,(2M)\cdot 2^{-1}-1}, \\ & d_{-2,0}, d_{-2,1}, d_{-2,2}, \ldots, d_{-2,(2M)\cdot 2^{-2}-2}, d_{-2,(2M)\cdot 2^{-2}-1}, \\ & \ldots, \\ & d_{j,0}, d_{j,1}, \ldots, d_{j,k} \ \ldots, d_{j,(2M)\cdot 2^{j}-2}, d_{j,(2M)\cdot 2^{j}-1}, \\ & \ldots, \\ & d_{-J+2,0}, d_{-J+2,1}, d_{-J+2,2}, d_{-J+2,3}, \\ & d_{-J+1,0}, d_{-J+1,1}, \\ & d_{-J,0}, \\ & c_{-J,0}\big) \ , \end{aligned}$$

where $[\cdot]$ designates the function (i.e. finite sequence) to be analyzed, and the subscripts of $d_{j,k}$, namely $j = -1, \ldots, -\log_2(2M)$ and $k = 0, \ldots, (2M)2^j - 1$, denote the wavelet scale and location parameters respectively. The "full tree wavelet" decomposition is so-called because its wavelet coefficients can be arranged to naturally lie on the nodes of a binary tree having a single root (which corresponds to the wavelet coefficient at the coarsest possible scale); see e.g. Mallat (1998). Assuming a signal consisting of $N \equiv 2M = 2^J$ points, this decomposition is composed of $N/2$ wavelet coefficients corresponding to the finest scale $j = -1$, $N/4$ wavelet coefficients corresponding to the next finest scale $j = -2$, $N/8$ wavelet coefficients $\cdots\cdots$, 2 wavelet coefficients corresponding to scale $j = -J + 1$, and finally, 1 wavelet coefficient corresponding to 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 with our regression procedure using various numbers of wavelet coefficients. Thus, at each scale of the decomposition, we would retain only the N_ψ wavelet coefficients which lie under (i.e. nearest to) the singularity (i.e.

the origin) of the characteristic function. At the 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

$$D\,\theta^R = \{W[\partial_\alpha c_\theta],\, W[\partial_\sigma c_\theta]\}\,; \quad D\,\theta^I = W[\partial_\mu c_\theta],$$

(remark: when not all wavelet coefficients are used the matrices above will, of course, have fewer than $2M$ rows) as well as the $(2M+1) \times (2M+1)$ real valued Toeplitz structure matrix T whose first row just constitutes the values of the characteristic function $c_\theta(s_\ell)$:

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

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

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

The theoretical covariance matrix of the real part of the empirical cf is then equal to

$$\Sigma^R \;=\; \frac{1}{2}\,(T + \tilde{T}) - R\,, \tag{5.2}$$

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

$$\Sigma^I \;=\; \frac{1}{2}(T - \tilde{T})\,. \tag{5.3}$$

The quantities (5.2) and (5.3) here correspond to the expressions (4.14) and (4.16) of the previous section. The matrix corresponding to (4.15) is, of course, null for the symmetric laws.

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

$$\Psi^R = \mathcal{W}\,\Sigma^R\,\mathcal{W}', \tag{5.4}$$

where $\mathcal{W}$ is the matrix of the 1-D wavelet transform. The (non-standard) 2-D wavelet transform (5.4) thus corresponds to a 1-D orthogonal wavelet transform applied to each column of Σ^R, together with a 1-D 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 for our experiments. (Of course, the cross-covariance matrix here is just 0.) When not all wavelet coefficients are retained in the fitting procedure, then Ψ^R is simply pruned so that only the required covariances are kept.

Finally, the procedure used to update the parameter estimate $\hat{\theta}$ is of the form

$$\hat{\theta}_{k+1} = \hat{\theta}_k + \rho\, d\theta \tag{5.5}$$

where $\rho < 1$ is a *damping parameter*, and $d\theta$ is the weighted least squares solution of the nonlinear regression equation

$$W[\Re\{c_N\}] - W[\Re\{c_\theta\}] = \left(D\,\theta^R\right) d\theta + \mathcal{E}. \tag{5.6}$$

In the above equation, the error vector $\mathcal{E}$ is regarded as being approximately normally distributed with zero mean and covariance matrix Ψ^R, while θ on the left side is at the current estimate. The classical linear algebra solution of this problem (which can be highly singular if large numbers of coefficients are used) was obtained using the Matlab procedure `lscov.m` which is based on QR decomposition, and thus 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. The solution is found without inverting Σ, by computing the QR decomposition of A and modifying its Q component by Σ following a procedure described in Strang(1986, p. 398). Concerning the damping parameter, we found that the value $\rho = 0.9$ resulted in good overall performance in our algorithm. (Small values of ρ require too many loops for satisfactory convergence, while values too large can result in nonconvergence to the correct value.)

Thus ends one iteration of the nonlinear least squares wavelets-based updating algorithm for our estimator. As noted previously, in all our experiments, the total number of iterations was fixed to equal three ($k = 1$, 2, and 3), and this was found, after suitable experimentation, to always be adequate.

Finally, we should remark here that if the location parameter μ were also to be fitted using our wavelet-based regression procedure then the regression (5.6) would need to be augmented by the one-parameter regression problem $W[\Im\{c_N\}] - W[\Im\{c_\theta\}] = \left(D\,\theta^I\right) d\mu + \mathcal{E}'$ for that parameter. In the event that the full four parameter family is to be fitted, the covariance structure of the real and imaginary terms would not separate (i.e. we would not have the parameter orthogonalities), so that 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 non-zero cross terms off the diagonal blocks. Put otherwise, in the four parameter case the covariance matrix of the real and imaginary components of the empirical cf and of their wavelet coefficients would then not be block diagonal; this corresponds to the fact that the terms (4.15), and hence (4.19), are not null in nonsymmetric families. We remark that the quantities $\partial_\mu c_\theta$ and $D\,\theta^I$, whose computations have been indicated above, 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.

6 Summary of Numerical Experiments.

In this section we describe the numerical experiments carried out using the algorithms described in Section 5. Firstly, however, we make two general observations. The first is that wavelet transformation tends to 'disbalance' or to 'alter the presentation' of data; it also tends to diagonalize, or at least to make sparse, the covariance and correlation structures of the signals to which wavelet transformation is applied. (See, e.g., Vidakovic, 1999.) To illustrate this point, Figures 6.1 and 6.2 show the covariance and correlation

functions (i.e. matrices), multiplied by n, which are associated with the real part of the ecf on the interval $t \in [-1, 1]$ for the case of the standardized stable distribution with $\alpha = 0.5$, while Figures 6.3 and 6.4 show the corresponding functions for the imaginary part of the ecf; these plots are based on the linear colour-code scales which are shown at the right in these figures. (We chose a low value for α in order to exhibit high levels of correlation here; as α increases the illustrated effects only become more striking.) The values of covariance here 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 the covariance matrices in both Figures 6.1 and 6.3. 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; however these correlations are evident in Figures 6.2 and 6.4. Figures 6.5 and 6.6 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. (Ten db equals a 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 trace 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 wavelets functions but centered at (nearly) identical locations; the (dark blue) remaining correlations (away from the topmost and rightmost edge) are all essentially null. The plots for the covariance and correlation matrices for the wavelet coefficients of the imaginary part of the ecf are very similar to those in Figures 6.5 and 6.6 and therefore

are not given here.

Our second observation is that wavelet transformation tends to redistribute as well as to 'concentrate' the (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 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 present in individual terms. For example, if all parameters are known except for one — let us say α — and if this parameter 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 will be

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

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

$$[\, n \cdot \operatorname{Var} \Re\, c_n(t) \,]^{-1} \;\times \left[\frac{d\, \Re\, c_\theta(t)}{d\alpha} \right]^{2} . \tag{6.1}$$

(Alternately, we could refer to this as the 'statistical sensitivity' of $\Re\, c_n(t)$ to this parameter.) Formulas analogous to (6.1) also hold for the other stable law parameters, as well as for $\Im\, c_n(t)$, and also for the individual wavelet coefficients.

Figure 6.7 shows the information per observation for α, σ, and μ – in the sense of (6.1) – at each point of the ecf for the case of a stable distribution with $\alpha = 1.75$, $\sigma = 1$ and $\mu = 0$; the curves shown for α and σ are computed with respect to the real part of the cf while that for μ is computed with respect to the imaginary part, these being the relevant information components for each parameter. The same plot is also given in Figure 6.8 but using a logarithmic vertical scale. In these plots, the ecf is viewed on the domain $t \in [-4, 4]$ with a gridding of 32 points per unit. Figure 6.9 shows the information per observation for the corresponding individual wavelet coefficients (with coarsest scales located leftmost, and only wavelets centered on the nonnegative axis included.) As compared to Figure

6.7, it is seen that only a very few of the wavelet coefficients contain any appreciable amount of information; to show this clearly, the plots for α, σ and μ have not been superimposed here. The same plots are also given in Figure 6.10 using a logarithmic vertical scale. To further illustrate the above points, Figures 6.11, 6.12 and 6.13 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 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 the rearrangement and concentration of Fisher's Information under wavelet transformation that these plots illustrate, is that typically fewer wavelet coefficients are required (than, say, ecf coefficients) in order to obtain the same level of statistical performance for purposes of estimation and inference. The use of fewer and less correlated coefficients often also carries numerical advantages.

We turn now to the results of a simulation study summarized compactly here in Figures 6.14 and 6.15. 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, and for estimates obtained from the four parameter regression procedure of Koutrouvelis are included within each array. Both the McCulloch and Koutrouvelis estimators that are used here are based on algorithms that are implemented in `FracLab`. (See McCulloch, 1986, and Koutrouvelis, 1980.) In all of the results given here, we used the compactly supported Daubechies wavelet of regularity 2 (i.e. 4 nonzero filter coefficients and 2 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 6.14 summarizes the performance of the estimators for α. The six rows of Figure 6.14 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 the interval $[0, 2]$. 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 1600 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 numbers of wavelet coefficients (32, 8, 4, 2) as indicated on the x-axis. When not all wavelet coefficients were retained, the indicated, equal number of coefficients was selected from each scale by selecting those centered most closely to the origin. (At coarser scales having fewer than this number of coefficients, all coefficients were included.) Algorithm parameters that are 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 within these displays, a full horizontal line is drawn at the true parameter value across each of the box-plot arrays.

Figures 6.15 is the same as Figure 6.14 except it summarizes the performance of the

estimators for the scaling parameter σ. In particular, however, 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 include summaries here for estimates of the location parameter μ and only note that estimates of μ obtained from fitting the symmetric three parameter stable 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 (1975, Table 1) does provide some asymptotic correlations $\rho_{\beta,\mu}$ for these correlated parameters for several values of α; 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 the parameters α and σ we do not require similar asymptotic corrections. However note in passing that DuMouchel (1975) uses a discontinuous representation for the stable family so use of his Table 1 for values of β other than 0 may require adjustment of the value of β.

Examining Figures 6.14 and 6.15 we see that for most combinations of algorithm parameters our wavelet based estimation procedure somewhat outperform (or at least match) 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 we summarize these results briefly here. The data set, which was downloaded from the `http://chart.yahoo.com` website, 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 to 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 each of the decades 1950-59, 1960-69, 1970-79, 1980-89, and 1990-99 respectively. 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 take note of this disparity here, we do not take it further into account below.

Table 1: **Sample sizes for the S&P500 data sets.**

S&P500 Data	Positive values	Negative Values
1950-1959	1407	1119
1960-1969	1320	1139
1970-1979	1277	1230
1980-1989	1334	1187
1990-1999	1354	1171

Because our interest focussed on the tails of the distribution, we applied our wavelet based estimation procedures only to selected segments of the empirical characteristic functions near the origin. The resulting estimates for α and σ using cf supports $[-a, a]$ with $a = 0.5, 1.0$ and 2.0 are given in Tables 2 and 3; the cf scaling here involved first standardizing the returns by dividing by their interquartile ranges. We note that (with this scaling) the ecf functions appeared to follow the cf's for stable distributions very closely up to approximately $t = \pm 1.5$ only. The estimates shown in Tables 2 and 3 are each based on 64 sample points and 64 wavelet coefficients using the Haar wavelet. As a general rule, it may be seen that both the stability index and scale parameter 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. In terms of discernable patterns of variation, the decade 1970-1979 stands out, both for having a disproportionately equal number of up versus down market days (as compared with 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 up the 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.)

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

7 Discussion.

Our purpose, in this paper, was to demonstrate that wavelet-base methods (which hitherto have typically been used primarily in nonparametric applications) can also be used effectively in statistics to carry out estimation (and inference) 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, a problem to which the theory of maximum likelihood is applicable, but where standard statistical methods are nevertheless quite difficult to implement. We have seen that wavelet-based methods in this context lead to procedures

Table 3: **Scale parameter σ estimates for the S&P500 data sets.**

S&P500 Data	Koutrouvelis	McCulloch	Wavelets 0.5	Wavelets 1.0	Wavelets 2.0
1950-1959 pos	0.5215	0.5217	0.5335	0.5243	0.5232
1950-1959 neg	0.5251	0.5163	0.5719	0.5322	0.5224
1960-1969 pos	0.5208	0.5208	0.5226	0.5211	0.5208
1960-1969 neg	0.5460	0.5158	0.5987	0.5565	0.5433
1970-1979 pos	0.5319	0.5189	0.5641	0.5375	0.5308
1970-1979 neg	0.5454	0.5209	0.5736	0.5547	0.5534
1980-1989 pos	0.5377	0.5152	0.5747	0.5511	0.5419
1980-1989 neg	0.5398	0.5157	0.5613	0.5457	0.5357
1990-1999 pos	0.5516	0.5158	0.5877	0.5624	0.5555
1990-1999 neg	0.5469	0.5115	0.6373	0.5780	0.5557

which are highly efficient and robust and also are competitive with other ad-hoc methods that have been devised for this family. There also remain substantial avenues for further research.

We have seen that the computational considerations in using wavelets in the way considered here 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 rearrangements of the Fisher information in data into patterns which give 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 estimation methods, and may sometimes allow us to replace generalized least squares by ordinary least squares. Further, ad-hoc selection of which coefficients to use in such procedures can often be made in an intuitively more straightforward way when using wavelets. In fact it was often found to be the case that little statistical efficiency was lost when only small subsets of wavelet coefficients were used (relative to using all of the coefficients).

We have argued that either a wavelet basis or the 'dual' basis consisting of its Fourier

transforms can be used to determine the 'moments' used in a generalized regression estimation procedure. Although 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 typically 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 open question is whether an efficient algorithm (such as a pyramid-type algorithm) can be devised for evaluating inner products relative to the Fourier dual bases.

In addition to inheriting the robustness properties associated with the fact that the ecf is not greatly altered when outliers are present, wavelets also have some advantages in dealing with heavy tailed families of distributions. This is because wavelet transformation allows us to 'zoom in' to singularities at the origin of a cf which is where information about the tails of the distribution is located. (The degree of singularity at the origin of the cf is related directly to the heaviness of the tail.) DuMouchel (1975), for example, has noted that stable samples contain a substantial amount of information for the parameter α in the extreme order statistics.

Finally we remark that, in our view (and in spite of the considerable efforts already spent in this direction) continued efforts seem warranted to implement reliable MLE algorithms (across multiple platforms) for estimating the parameters of the stable laws (in both the symmetric as well as nonsymmetric cases) especially algorithms that perform reliably for smaller values of α where MLE procedures have been notoriously difficult to implement. We found it interesting to observe that the procedure of Koutrouvelis (1980), while not fully efficient, nevertheless performed effectively and 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. The authors thank F. Abramovitch, U. Amato, B. Vidakovics and the organisers and participants of the workshop on Wavelets and Statistics in Vico Equense for their role in fostering this collaboration.

Headings and Captions for Figures.

Figure 6.1: 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.

Figure 6.2: Correlation Matrix of $\Re\, c_n$.

The correlation matrix corresponding to Figure 6.1.

Figure 6.3: 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.

Figure 6.4: Correlation Matrix of $\Im\, c_n$.

The correlation matrix corresponding to Figure 6.3.

Figure 6.5: Covariance Matrix of $\mathcal{W}\, \Re\, c_n$.

The covariance matrix of the wavelet coefficients $\mathcal{W}\, \Re\, c_n$ corresponding to Figure 6.1, with logarithmic colour scale.

Figure 6.6: Correlation Matrix of $\mathcal{W}\, \Re\, c_n$.

The correlation matrix corresponding to Figure 6.5, with logarithmic colour scale.

Figure 6.7: 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.

Figure 6.8: Information for α, σ, and μ in $\Re\, c_n$.

Same as Figure 6.7, but using a logarithmic scale for information (vertical axis).

Figure 6.9: 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.

Figure 6.10: Information for α, σ, and μ in $\mathcal{W}\Re c_n$.

Same as Figure 6.9, but using logarithmic scales for information.

Figure 6.11: 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.

Figure 6.12: 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.

Figure 6.13: 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.

Figure 6.14: 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 1600. 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 effects of using fewer than all wavelet coefficients. (See text for further details.)

Figure 6.15: Performance of the Wavelet Estimators for σ.

Same as Figure 6.14 except for estimators of the scale parameter σ. Rows correspond to

values of α as in Figure 6.14, with columns as before. True value of $\alpha = 1$. (See text for details.)

References

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

[2] Antoniadis, A., Grégoire, G. and McKeague, I. (1994). Wavelet methods for curve estimation. *J. Amer. Statist. Assoc.* **89**(428), 1340–1353.

[3] Antoniadis, A. (1997). Wavelets in statistics: A review (with discussion). *J. Italian Statist. Soc.*, **6**, 97-144.

[4] Antoniadis, A., Grégoire, G. and Nason, G. (1999). Density and hazard rate estimation for right-censored data using wavelet methods. *J. Royal Statist. Soc.*, **B 61**, 63-84.

[5] Bruce, A. G. and Gao, H.-Y. (1994). *S+Wavelets, Users manual*. StatSci, Seattle.

[6] Buckheit, J. B. and Donoho, D. (1995). Wavelab and Reproducible research. In A. Antoniadis and G. Oppenheim (eds.), *Wavelets and Statistics*, Lecture Notes in Statistics, **103**, Springer-Verlag.

[7] Cappé, O., Moulines, E., Pesquet, J.-C., Petropulu, A. and Yang, X. (2002). Long-range dependence and heavy-tail modeling for teletraffic data. *IEEE Signal Processing Magazine*, Vol. 19, May 2002, p. 14-27

[8] Chambers, J.M., Mallows, C.L. and Stuck, B.W. (1976). A method for simulating stable random variables. *J. Amer. Statist. Assoc.*, **71**, 340-344.

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

[10] Christof, G. and Wolf, G. (1992). *Convergence Theorems with a Stable Limit Law.* Akademie Verlag, Berlin.

[11] Chui, K. (1992). *Wavelets: A Tutorial in Theory and Applications.* Academic Press, Boston.

[12] Cohen, A. and Ryan, R. D (1995). *Wavelets and Multiscale Signal Processing.* Chapman & Hall, London.

[13] Daubechies, I. (1992). *Ten Lectures on Wavelets.* CBMS-NSF Regional Conferences Series in Applied Mathematics. SIAM, Philadelphia.

[14] Donoho, D. L., Johnstone, I. M., Kerkyacharian, G. and Picard, D. (1996). Density estimation by wavelet thresholding. *Ann. Statist.* **24**(2), 508–539.

[15] DuMouchel, W.H. (1971). *Stable Distributions in Statistical Inference.* Ph.D. Dissertation, Yale University.

[16] DuMouchel, W.H. (1973a). On the asymptotic normality of maximum likelihood estimates when sampling from a stable distribution. *Ann. Statist.*, **1**, 948-957.

[17] DuMouchel, W.H. (1973b). Stable distributions in statistical inference. I: Symmetric stable distributions compared to other symmetric long-tailed distributions. *J. Amer. Statist. Assoc.*, **68**, 469-482.

[18] DuMouchel, W.H. (1975). Stable distributions in statistical inference. II: Information from stably distributed samples. *J. Amer. Statist. Assoc.*, **70**, 386-393.

[19] DuMouchel, W.H. (1983). Estimating the stable index α in order to measure tail thickness: A critique. *Ann. Statist.*, **11**, 1019-1031.

[20] Fama, E. (1965). The behaviour of stock prices. *J. Business*, **38**, 34-105.

[21] Fama, E. and Roll, R. (1968). Some properties of symmetric stable distributions. *J. Amer. Statist. Assoc.*, **63**, 817-836.

[22] Fama, E. and Roll, R. (1971). Parameter estimates for symmetric stable distributions. *J. Amer. Statist. Assoc.*, **66**, 331-338.

[23] Feller, W. (1971). *An Introduction to Probability Theory and its Applications.* Vol. 2, 2nd ed. Wiley, New York.

[24] Feuerverger, A. and McDunnough, P. (1981a). On some Fourier methods for inference. *J. Amer. Statist. Assoc.*, **76**, 379-387.

[25] Feuerverger, A. and McDunnough, P. (1981b). On efficient inference in symmetric stable laws and processes. In: *Proceedings of the International Symposium on Statistics and Related Topics, Ottawa, May 1980*, A.K.Md.E. Saleh et al., eds., North Holland, Amsterdam, 109-122.

[26] Feuerverger, A. and McDunnough, P. (1984). On statistical transform methods and their efficiency. *Canadian Journal of Statistics*, **12**, 303-317.

[27] Flandrin, P. (1992). Wavelet analysis and synthesis of fractional Brownian motion. *IEEE Trans. Info. Theory*, **38**, 910–917.

[28] FracLab. (2002). INRIA Project Fractales. Open source freeware distributed by INRIA-Fractales: `http://www-rocq.inria.fr/fractales`.

[29] Gao, H.-Y. (1993). *Wavelet estimation of Spectral densities in Time series analysis.* Ph. D. Thesis, University of California, Berkeley.

[30] Gnedenko, V.B. and Kolmogorov, A.N. (1954). *Limit distributions for Sums of Independent Random Variables.* Addison-Wesley, Reading.

[31] Hall, P.G. (1981). A comedy of errors: the canonical form for a stable characteristic function. *Bull. London Math. Soc.*, **13**, 23-27.

[32] Hall, P. and Patil, P. (1995). On wavelet methods for estimating smooth functions. *Bernoulli* **1**, 41–58.

[33] Hill, B.M. (1975). A simple general approach to inference about the tail of a distribution. *Ann. Statist.*, **3**, 1163-1174.

[34] Holschneider, M. (1995). *Wavelets: An Analysis Tool.* Clarendon Press, Oxford.

[35] Holt, D.R. and Crow, E.L. (1973). Tables and graphs of the stable probability density functions. *J. Res. Nat. Bur. Standards*, 77B, Vols. 3 & 4.

[36] Ibragimov, I.A. and Linnik, Yu.V. (1971). *Independent and Stationary Sequences of Random Variables.* Wolters-Noordhoff, Groningen.

[37] Janicki, A. and Weron, A. (1994). *Simulation and Chaotic Behaviour of α-stable Stochastic Processes.* Marcel Dekker, New York.

[38] Johnstone, I. M., Kerkyacharian, G. and Picard, D. (1992). Estimation d'une densité de probabilité par méthode d'ondelettes. *Compt. Rend. Acad. Sci. Paris A* **315**, 211–216.

[39] Koutrouvelis, I.A. (1980). Regression-type estimation of the parameters of stable laws. *J. Amer. Statist. Assoc.*, **75**, 918-928.

[40] Koutrouvelis, I.A. (1981). An iterative procedure for the estimation of the parameters of the stable law. *Commun. Statist. – Simulation & Computation*, **10**, 17-28.

[41] Lukacs, E. (1970). *Characteristic Functions.* Hafner, Connecticut.

[42] Mallat, S. G. (1989). A theory for multiresolution signal decomposition: the wavelet representation. *IEEE Trans. on Pattern Analysis and Machine Intelligence* **11**, 674–693.

[43] Mallat, S.G. (1998). *A Wavelet Tour of Signal Processing.* Academic Press, San Diego.

[44] Mandelbrot, B.B. (1963, 1967, 1972). The variation of certain speculative prices. *J. Business*, **36**, 394-419; **40**, 393-413; and **45**, 542-543

[45] McCulloch, J.H. (1986.) Simple consistent estimators of stable distribution parameters. *Commun. Statist. – Computation & Simulation*, **15**, 1109-1136.

[46] Meyer, Y. (1990). *Ondelettes et Opérateurs I: Ondelettes.* Hermann, Paris.

[47] Mittnik, M. and Rachev, S. (1993). Modeling asset returns with alternative stable distributions. *Econometric Rev.*, **12**, 261-330.

[48] Mittnik, M. and Rachev, S., eds. (2000). *Stable Paretian Models in Finance.* Wiley, New York.

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

[50] Nolan, J. (1997). Numerical calculation of stable densities and distribution functions. *Commun. in Statist. – Stochastic Models*, **13**, 759-774.

[51] Nolan, J. (2001). Maximum likelihood estimation and diagnostics for stable distributions. In: *Lévy Processes*, O.E. Barndorff-Nielson, T. Mikosch and S.I. Resnick, eds., 379-400. Birkhauser, Boston.

[52] Ogden, T. R. (1997). *Essential Wavelets for Statistical Applications and Data Analysis.* Birkhäuser, Basel.

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

[54] Press, S.J. (1972). Estimation in univariate and multivariate stable distributions. *J. Amer. Statist. Assoc.*, **67**, 842-846.

[55] Samorodnitsky, G. and Taqqu, M. (1994). *Stable non-Gaussian Random Processes: Stochastic Models with Infinite Variance.* Chapman and Hall, New York.

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

[57] Tribouley, K. (1995). Practical estimation of multivariate densities using wavelet methods. *Statistica Neerlandica* **49**, 41–62.

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

[59] Vidakovic, B. (1999). *Statistical Modeling by Wavelets.* Wiley, New York.

[60] Wahba, G. (1990). *Spline models for observational data.* CBMS-NSF regional conferences series in applied mathematics. SIAM, Philadelphia.

[61] Zolotarev, V.M. (1986). *One-Dimensional Stable Distributions.* Translations of Mathematical Monographs, Vol. 65, American Mathematical Society, Providence, Rhode Island. (Translation of the original Russian edition of 1983.)