The Report Committee for Novin Ghaffari
Certifies that this is the approved version of the following report:

Estimation with Stable Disturbances

APPROVED BY
SUPERVISING COMMITTEE:

Supervisor:

____
Carlos Carvalho

____
Peter Mueller

Estimation with Stable Disturbances

by

Novin Ghaffari, B.A.; B.B.A.; B.S.Math.

Report

Presented to the Faculty of the Graduate School of

The University of Texas at Austin

in Partial Fulfillment

of the Requirements

for the Degree of

Master of Science in Statistics

The University of Texas at Austin

May 2014

Dedication

To those who have supported and raised me from infancy, I would not be where I am, were it not for your patience and devotion.

Acknowledgements

I would like to acknowledge the support and direction I received from my advisor, Carlos Carvalho, and my second reader Peter Mueller. Their guidance and expertise in the statistical sciences has been invaluable.

Abstract

Estimation with Stable Disturbances

Novin Ghaffari, M.S. Stat.
The University of Texas at Austin, 2014

Supervisor: Carlos Carvalho

The family of stable distributions represents an important generalization of the Gaussian family; stable random variables obey a generalized central limit theorem where the assumption of finite variance is replaced with one of power law decay in the tails. Possessing heavy tails, asymmetry, and infinite variance, non-Gaussian stable distributions can be suitable for inference in settings featuring impulsive, possibly skewed noise. A general lack of analytical form for the densities and distributions of stable laws has prompted research into computational methods of estimation. This report introduces stable distributions through a discussion of their basic properties and definitions in chapter 1. Chapter 2 surveys applications, and chapter 3 discusses a number of procedures for inference, with particular attention to time series models in the ARMA setting. Further details and an application can be found in the appendices.

Table of Contents

List of Figures

Chapter 1 Stable Distributions

INTRODUCTION

During the last half century, stable distributions have been increasingly exploited to characterize the dynamics of certain stochastic models. The appeal of using stable distributions stems from a number of convenient theoretical properties. The stable family of distributions may be thought of as a generalization of Gaussian distributions, retaining the Gaussian family as special subset. As might be expected, stable distributions observe a number of the key features familiar to the Gaussian case. Like the Gaussian, stable distributions are limiting distributions for sums of iid random variables, characterizing a generalized central limit theorem. Additionally, stable random variables are closed, with respect to their underlying distribution, under the summation of iid copies. However, stable distributions (except for the Gaussian), are leptokurtic and heavy-tailed. The stable family can also accommodate asymmetry. Inference with these distributions can capture impulsive and skewed patterns of variation better than traditional Gaussian methods.

Inference with stable distributions is hampered by a few major inconveniences. All non-Gaussian stable distributions have infinite second moments and hence infinite variances; some even exhibit infinite first moments. This effectively rules out variance-based estimation techniques customary to the Gaussian setting. In fact, barring a few exceptions, stable distributions do not have known closed-form densities; the foundational probability theory behind stable laws was largely accomplished in the frequency domain using characteristic functions. Given these difficulties, early work in statistical inference was hindered. However, advances in computer hardware and statistical computing have enabled new methods for inference with stable laws. The development of novel techniques has been coupled with an increasing breadth of application.

The rest of this report ensues as follows. The next section of Chapter 1 will formally introduce stable measures and random variables through a brief discussion of their basic properties. The last section of Chapter 1 will extend the univariate case to stable vectors and processes, with particular attention to stable ARMA processes. Chapter 2 will motivate the use of stable laws in inference with a survey of applications, with particular attention to finance and economics. Chapter 3 will provide an overview of methods for inference. Of the various procedures, this report highlights two methods: a Bayesian/MCMC approach for inference with symmetric stable noise and a fast Fourier transform (FFT) method for calculating the likelihood and conducting MLE. Detailed information and application of these methods is to be found in Appendices B, C, and D.

DEFINING PROPERTIES

There are a number of equivalent definitions for stable distributions. Samorodnitsky and Taqqu provide four equivalent definitions; these may be placed into three categories: sum stability, domains of attraction, and characteristic function (1994). These definitions will be reproduced and discussed here. Sum stability and domains of attraction highlight mathematically significant features of stable distributions that are defining properties. Characteristic functions are important for representing stable distributions parametrically and conducting statistical analysis with stable laws. Thus these defining properties also embody important characteristics for inference and application.

Sum Stability

Broadly speaking, stability is a property of closure on a class of distributions with respect to a binary operation on its random variables. Formally, let X_i denote a random variable, F_i its distribution function, $\mathcal{F}$ a class of distributions and $B(\cdot,\cdot)$ a binary operation. Then class $\mathcal{F}$ is said to be stable with respect to operation $B(\cdot)$ if $\forall X_i, X_j \sim F_{ij}$ such that $F_{ij} \in \mathcal{F}$, we have $B(X_i, X_j) = X_k$ where $X_k \sim F_k$ and $F_k \in \mathcal{F}$.

There are a number of categories of stability that have been studied, including geometric and min/max stability. But by far the most studied form of stability is sum stability, where $B(\cdot,\cdot)$ corresponds to the summation of random variables. Random variables that are sum stable under iid summation and their corresponding distributions are referred to as just "stable." This is done for simplicity's sake without implying they constitute the only type of stability. In some literature this family is referred to as α-stable or Lévy stable. This report keeps to the term stable for convenience and brevity.

Sum stability is a defining characteristic of stable distributions. Two equivalent definitions of stable distributions via sum stability can be formulated. Below, the d above the equality sign denotes equality in distribution.

Definition 1: A random variable X is stable if, given independent copies $X_1, X_2 \overset{\text{iid}}{\sim} X$, and any positive constants $a, b \in \mathbb{R}^+$, $\exists$ constants $c \in \mathbb{R}^+$ and $e \in \mathbb{R}$ such that,

$$aX_1 + bX_2 \overset{d}{=} cX + e \tag{1.1}$$

Definition 2: A random variable X is stable if for iid replications $X_1, X_2, \ldots, X_n \overset{\text{iid}}{\sim} X$ with $n \geq 2$, $\exists$ constants $c_n \in \mathbb{R}^+$ and $e_n \in \mathbb{R}$ such that,

$$X_1 + X_2 + \cdots + X_n \overset{d}{=} c_n X + e_n \tag{1.2}$$

Definitions 1 and 2 define stable random variables. If $e = 0\ \forall A, B$ in (1.1) the random variable is said to be *strictly stable*[1]. Likewise if $e_n = 0\ \ \forall n$ in (1.2) (Samorodnitsky & Taqqu, 1994). The value of the constant c in (1.1) satisfies the equation: $c^\alpha = a^\alpha + b^\alpha$ for some $\alpha \in (0,2]$. For a proof see Section VI.1 (Feller, 1971). Similarly in (1.2), c_n satisfies the relation: $c_n = n^{1/\alpha}$, again with $\alpha \in (0,2]$ (Feller, 1971). Note under definition 2, the requirement that (1.2) hold for $n = 2$ is not sufficient to determine stability. A random variable is necessarily stable only if (1.2) holds for $n = 2$ and $n = 3$ (Feller, 1971) (Zolotarev, 1986).

Domains of Attraction and Characteristic Functions

A random variable X has a domain of attraction if $\exists$ a sequence of iid random variables $Y_1, Y_2, \ldots$ and a sequences of constants $\{g_n\} \in \mathbb{R}^+$ and $\{h_n\} \in \mathbb{R}$ such that

$$\frac{\sum_{i=1}^{\infty} Y_i}{g_n} + h_n \overset{d.}{\rightarrow} X \tag{1.3}$$

Where $\overset{d.}{\rightarrow}$ denotes convergence in distribution. From definition 2, it is clear that every stable distribution admits a domain of attraction. In fact, any non-degenerate distribution will necessarily be in the domain of attraction of a stable law (Gnedenko & Kolmogorov, 1954). Accordingly, this sets an alternative definition of stable distributions:

[1] Beware, Lévy (1954) and Feller (1966) use ‘stable’ and ‘quasi-stable’ for ‘strictly stable’ and ‘stable’. Feller (1971) uses the terminology ‘stable in the narrow/strict sense’ and ‘stable in the broad sense.’

Definition 3: A random variable X has a stable distribution $\Longleftrightarrow$ it has a domain of attraction.

Feller, Khinchin, and Lévy showed in 1935 that a random variable X with distribution function F is in the normal domain of attraction if and only if, $x^2 \frac{\int_{|t|>x} dF(t)}{\int_{|t|<x} t^2 dF(t)} \to 0$ as $x \to \infty$. Gnedenko and Kolmogorov define a generalized central limit theorem, the counterpart to the classical central limit theorem without the condition of finite variance (1954). Stable laws, like the Gaussian distribution, arise as the sum of many individual noise components, an important consideration for statistical modeling.

Gnedenko and Kolmogorov prove the generalized theorem using the characteristic function to represent stable laws. Lacking a general closed form density or distribution, most early work with stable distributions was carried out using the characteristic function. The characteristic function can be expressed through several readily interpretable parameterizations of stable laws. These parameterizations are vital to estimation.

The characteristic function of univariate stable distributions is typically parameterized by four parameters. The shape parameter and the skew parameter are denoted α and β respectively. The parameter α is also referred to as the index of stability or the characteristic exponent; it determines the thickness of the tails. In regard to sum stability, the α parameter is the most important. The sum of independent stable random variables with equivalent α values, regardless of other parameter values, will yield a stable random variable with the same α. In general, the sum of stable random variables with differing α values will not yield a stable distribution. The skew parameter β is an indicator of asymmetry. Note, this parameter does *not* correspond to the classical notion of skewness. In fact classical skewness is zero in the Gaussian case and undefined for all other cases where higher order moments are infinite. The location and scale parameters will be denoted δ and γ in this report. This is in contrast to the typical choice of μ and σ, reflecting that for all non-Gaussian stable random variables the scale γ is not equivalent to the standard deviation (which does not exist when variance is infinite), and for some non-Gaussian stable variables, even the mean μ does not exist. There are several alternative parameterizations under which some of the above parameters take on slightly different interpretations. However, α's value and interpretation remain the same in each parameterization presented here. This report presents the three most commonly encountered parameterizations. Abusing the conditional notation, $S_\alpha(\beta, \delta, \gamma|k)$ signifies a stable distribution with parameters $(\alpha, \beta, \delta, \gamma)$ given parameterization k. For our parameterizations we will use Nolan's custom of labeling them as $k = 0,1,2$ (1998). General stable distributions, when referenced without regard to the specific parameterization or parameter values, will be denoted $S_\alpha(\beta, \delta, \gamma)$.

The most widely used parameterization is what Nolan terms the 1-parameterization (Nolan, 1998). The characteristic function has one of the simplest forms making it the parameterization of choice for algebraic manipulation. A random variable X is said to be a stable random variable if its characteristic function can be expressed in the form,

$$E[\exp(i\tau X)] = \begin{cases} \exp\left(i\delta\tau - \gamma^\alpha|\tau|^\alpha\left[1 - i\beta\tan\left(\frac{\pi\alpha}{2}\right)\operatorname{sgn}(\tau)\right]\right) & \alpha \neq 1 \\ \exp\left(i\delta\tau - \gamma|\tau|\left[1 + i\beta\frac{2}{\pi}\ln(|\tau|)\operatorname{sgn}(\tau)\right]\right) & \alpha = 1 \end{cases} \tag{1.4}$$

Where $\operatorname{sgn}(\cdot)$ is the sign function returning the sign of τ, with the specification that $\operatorname{sgn}(0) = 0$.

While the 1-parameterization is convenient in theoretical settings, it presents a few drawbacks for numerical procedures and statistical estimation. The exceptional case for $\alpha = 1$ makes it discontinuous as $\alpha \to 1$ for $\beta \neq 0$. Nolan's 0-parameterization, based on Zolotarev's polar (M)-parameterization, corrects for this, taking the form,

$$E[\exp(i\tau X)] = \begin{cases} \exp\left(i\delta\tau - \gamma^{\alpha}|\tau|^{\alpha}\left[1 + i\beta\tan\left(\frac{\pi\alpha}{2}\right)(|\gamma\tau|^{1-\alpha} - 1)\operatorname{sgn}(\tau)\right]\right) & \alpha \neq 1 \\ \exp\left(i\delta\tau - \gamma|\tau|\left[1 + i\beta\frac{2}{\pi}\ln(\gamma|\tau|)\operatorname{sgn}(\tau)\right]\right) & \alpha = 1 \end{cases} \tag{1.5}$$

The limit $\lim_{\alpha\to 1}\tan\left(\frac{\pi\alpha}{2}\right)(|\gamma\tau|^{1-\alpha} - 1) = \frac{2}{\pi}\ln(\gamma|\tau|)$ ensures the continuity of this form (Nolan, 1998). This makes the 0-parameterization preferable for numerical applications.

Other features of this parameterization support its use in statistical application. If $X \sim S_{\alpha}(\beta, \delta, \gamma|0)$ then for $Z \sim S_{\alpha}(\beta, \delta, \gamma|0)$ we have $X = \gamma Z + \delta$, i.e. the parameters (δ, γ) represent the traditional scale and shift of a location-scale family. This is not the case with the 1-parameterization, where $X = \gamma Z + \gamma\beta\tan\frac{\pi\alpha}{2} + \delta$ when $\alpha \neq 1$ and $X = \gamma Z + \gamma\beta\frac{2}{\pi}\ln\gamma$ when $\alpha = 1$. Hence the 0-parameterization is favored for likelihood estimation (Nolan, 1998). From these location-scale representations we can determine the relation between δ_0 and δ_1,

$$\begin{aligned} \delta_1 &= \delta_0 - \gamma\beta\tan\tfrac{\pi\alpha}{2} \quad \alpha \neq 1 \\ \delta_1 &= \delta_0 - \gamma\beta\tfrac{2}{\pi}\ln\gamma \quad \alpha = 1 \end{aligned} \tag{1.6}$$

The other parameters are the same between these two parameterizations. The 0-

parameterization admits a simpler interpretation of the mode, making it also the favored form for maximum a posteriori estimation (Nolan, 1998).

The 2-parameterization, constitutes yet another form for the characteristic function,

$$E\left[e^{i\tau X}\right] = \begin{cases} \exp\left(i\delta\tau - \gamma^{\alpha}|\tau|^{\alpha}\exp\left[-i\frac{\pi}{2}\beta\min(\alpha,2-\alpha)\,\mathrm{sgn}(\tau)\right]\right) & \alpha \neq 1 \\ \exp\left(i\delta\tau - \gamma|\tau|\left[1 + i\beta\min(\alpha,2-\alpha)\frac{2}{\pi}\ln(\gamma|\tau|)\,\mathrm{sgn}(\tau)\right]\right) & \alpha = 1 \end{cases} \tag{1.7}$$

This parameterization is generally undesirable. It has the numerical issues as the 1-parameterization, does not lend itself to easy manipulation, and the beta parameter exhibits the peculiarity that a negative value corresponds to negative skew for $\alpha \in (0,1)$ and positive skew for $\alpha \in (1,2)$. Nonetheless this form is mentioned in Samorodnitsky and Taqqu (1994), used in DuMouchel's paper on maximum likelihood estimation (1973b), and features in Buckle's MCMC scheme (Buckle, 1995), so it is presented here for convenience.

BASIC PROPERTIES

Despite the fact that no general closed-form expression exists for the density or distribution function of a stable random variable, many properties of these random variables and their associated density and distribution have been revealed through the characteristic function. For instance, it is known that the density function, f, of stable random variables, is unimodal and smooth, i.e. $f \in C^{\infty}$ (Yamazato, 1978). Here this report will cite five characteristics of stable laws that facilitate a general understanding of this family of distributions and are important for inference.

Property 1 (Gaussian): When $\alpha = 2$, $X \sim \mathrm{N}(\delta, 2\gamma^2)$. As $\alpha \uparrow 2$, $\tan \pi\alpha/2 \rightarrow 0$ and the β parameter effectively drops out of the characteristic function leaving $\phi_x(\tau) = \exp(i\delta\tau - \gamma^2\tau^2) \equiv \exp\left(i\delta - \frac{1}{2}\sigma^2\tau^2\right)$ for $\sigma^2 = 2\gamma^2$.

Property 2 (Symmetry): $X \sim S_\alpha(\beta, \delta, \gamma)$ is symmetric $\Leftrightarrow$ $\beta = 0$ and $\delta = 0$. It is symmetric about δ $\Leftrightarrow$ $\beta = 0$ (Samorodnitsky & Taqqu, 1994). As a corollary, $\sim S_\alpha(\beta, 0, \gamma) \Leftrightarrow -X \sim S_\alpha(-\beta, 0, \gamma)$, known as the reflection property (Samorodnitsky & Taqqu, 1994).

Property 3 (Totally Skewed Stable Laws): a random variable $X \sim S_\alpha(\beta, \delta, \gamma)$ with $\alpha < 1$ and $|\beta| = 1$ is a so-called totally skewed stable distribution with support only over the half real line. Under the 1-parameterization, the support is restricted to set $[\delta, \infty)$ and $(-\infty, \delta]$ when $\beta = 1$ and $\beta = -1$ respectively.

Property 4 (Product Property): Owing to Feller (1971), we have a product property stating that any symmetric stable random variable, Z , may be represented as the product of a symmetric and a positively skewed random variable. Let $X \sim S_{\alpha 1}(0,0,\gamma|1)$ for $0 < \alpha_1 \leq 2$ and let $0 < \alpha_2 < \alpha_1$. Then define the skewed positive stable random variable

$$A \sim S_{a2/a1}\left(1,0,\left(\cos\left(\frac{\pi\alpha_2}{2\alpha_1}\right)\right)^{\alpha_1/\alpha_2}\right) \tag{1.8}$$

and assume X, A are independent. Then,

$$Z = A^{1/\alpha_1} X \sim S_{\alpha 2}(0,0,\gamma) \tag{1.9}$$

The proof may be carried out using the Laplace transform, for more information see Samorodnitsky and Taqqu (1994) or Feller (1971).

Property 5 (Power Law Decay): The tails of stable distributions are "heavy" and follow a power law decay when $\alpha < 2$. That is for a non-Gaussian stable X, $P(|X| > x) \propto x^{-\alpha}$. Specifically,

$$\lim_{x\to\infty} P(X > x) = C_\alpha (1+\beta)\gamma^\alpha x^{-\alpha}$$
$$\lim_{x\to\infty} f(x|\alpha,\beta,\delta,\gamma) = \alpha C_\alpha (1+\beta)\gamma^\alpha x^{-(\alpha+1)}$$
$$\lim_{x\to\infty} P(X > -x) = C_\alpha (1-\beta)\gamma^\alpha x^{-\alpha}$$
$$\lim_{x\to\infty} f(-x|\alpha,\beta,\delta,\gamma) = \alpha C_\alpha (1-\beta)\gamma^\alpha x^{-(\alpha+1)} \tag{1.10}$$

When in the 1-parameterization $C_\alpha = (1-\alpha)/\Gamma(2-\alpha)\cos(\pi\alpha/2)$ for $\alpha \neq 1$ and $2/\pi$ for $\alpha = 1$ (Samorodnitsky & Taqqu, 1994). For the 0-parameterization $C_\alpha = \sin(\pi\alpha/2)\,\Gamma(\alpha)/\pi$ (Nolan, 1998).

Property 1 demonstrates under what conditions the stable parameterization recovers the Gaussian distribution. It highlights three important considerations for estimation and inference: 1) the scale parameter in the given parameterizations do not match the standard deviation of the classical Gaussian parameterization, accordingly caution must be exercised, 2) the beta parameter vanishes when $\alpha = 2$, becoming less meaningful as $\alpha \uparrow 2$, and will be insignificant for α in a neighborhood of 2, 3) for α close to 2 the practitioner may want to consider swapping a stable setting for the Gaussian assumption to take advantage of the established repertoire of estimation procedures. Properties 2 and 3 define symmetric and totally skewed stable distributions. These are important cases of stable distributions that will feature in inference methods presented in this report. Property 4 relates symmetric and positively skewed stable laws through the product property. This property features in application and inference; it permits any symmetric stable law (Z) to be defined as the product of a Gaussian (X) and a positively skewed stable distribution (A). This may be interpreted as a scale mixture of normals representation (SMiN) of Z, making this distribution Gaussian conditional on the positively skewed stable law A. Finally, property 5 expresses the power law decay in the tails of stable distributions. This is a distinguishing feature of stable laws that can play an important part in identifying relevant applications. For plots of stable laws, see Appendix A.

EXTENSIONS

As the theory and application of univariate stable laws were developed, increasing attention was given to multivariate extensions of univariate stable distributions. These include objects like stable random vectors, stable integrals, self-similar processes, stable ARMA processes and stable FARIMA processes. This report presents a brief summary of stable ARMA processes to facilitate later discussions on time series inferences. More information on stable ARMA processes in particular, and stable random vectors and processes in general, can be found in Samorodnitsky and Taqqu, chapters 2, 3, and 6-13 (1994).

A sequence $\{X_t : t = \cdots, -1, 0, 1, \ldots\}$ is a stable ARMA(p,q) process, for $p, q \geq 0$ if it satisfies the following,

$$X_t = \sum_{i=1}^{p} \phi_i X_{t-i} + \varepsilon_t + \sum_{i=0}^{q} \theta_i \varepsilon_{t-i} \quad \equiv \quad X_t - \sum_{i=1}^{p} \phi_i X_{t-i} = \sum_{i=0}^{q} \theta_i \varepsilon_{t-i} \tag{1.11}$$

With the disturbances distributed, $\epsilon_t \overset{iid}{\sim} s_\alpha(\beta, \gamma, \delta)$. Defining the backshift operator B, such that $B(X_t) = X_{t-1}$, $B^2(X_t) = X_{t-2}$,… and the polynomials $\Phi(z)$ and $\Theta(z)$ as,

$$\Phi(z) = 1 - \phi_1 z - \cdots - \phi_p z^p$$
$$\Theta(z) = 1 + \theta_1 z + \cdots + \theta_q z^q \tag{1.12}$$

We may express the ARMA model as $\Phi(B)X_t = \Theta(B)\varepsilon_t$ (Samorodnitsky & Taqqu, 1994).

As in the Gaussian case, it is customary to assume $\Phi(z)$ and $\Theta(z)$ share no common roots, else the model may be rewritten as a lower order ARMA process. Here we present two theorems that are also germaine to the strictly Gaussian case. For proofs of these see Samorodnitsky and Taqqu, section 7.12 (1994).

Theorem 1 (causality): The ARMA model (1.8) is causal with a unique solution $\Leftrightarrow \Phi(z)$ has no roots in the closed unit disk $\{z: |z| \leq 1\}$

Theorem 2 (invertibility): If $\Theta(z)$ has no roots in the closed unit disk $\{z: |z| \leq 1\}$, then $ARMA(p, q)$ is invertible.

Chapter 2 Applications

Here we survey some of the applied literature to illustrate the applicability of stable laws to different disciplines and to motivate a discussion of inference methods in chapter 3. Due to the lack of analytic form for the density or distribution function, stable distributions were rarely pursued in applied disciplines. After their full introduction by Lévy and Khinchin in the 1920's and 1930's, stable laws were mainly studied within probability theory. Some of the earliest applications include hitting times of random processes, in describing certain branching processes, and in the theory of random determinants. Examples of all three may be found in chapter 1 of Zolotarev (1986). Starting in the 1960's, with increased accessibility to computing power, researchers began studying statistical inference with stable distributions and their application to practical disciplines.

In signal processing stable distributions are useful for modeling impulsive noise. Stable laws have successfully been employed to model degraded audio samples with significant jumps in the noise signature (Godsill & Lombardi, 2004), impulsive random fields in image processing, and a generalization of the Rayleigh distribution for radar tracking (Kuruoglu & Zerubia, 2004). In physics, stable laws arise when describing light reflecting off of a rotating convex mirror, certain dynamics in quantum mechanics and statistical physics, and certain stochastic differential equations. One of the earliest physics applications, discovered before the full explication of stable laws by Lévy, is the Holtsmark distribution; it characterizes the chaotic electromagnetic fields of plasma and the gravitational fields of stars (Zolotarev, 1986). In biology the so-called Lévy flight foraging hypothesis has been used to describe animal movement patterns (Zolotarev, 1986). The Zipf-Pareto law, a generalized power law distribution, appears in a number of practical settings, for example: the sizes of cities in a nation, the frequencies of words in a language, the prevalence of animals by species in a habitat. Sums of such random variables will converge to stable limiting random variables. In fact, because of this relation, totally skewed stable distributions can be utilized in place of Zipf-Pareto distributions in a number of applications (Zolotarev, 1986). Zipf-Pareto laws also occur in the distribution of incomes, as well as in the description other economic and financial variables. Indeed some of the earliest applications of stable laws were in these fields. Today a large body of literature exists on the application of stable distributions to economics and finance. The next section of this chapter will explore these applications in more detail to provide some idea of the scope of applicability.

ECONOMICS AND FINANCE

Mandelbrot was the earliest pioneer in applying stable laws to financial and economic data. In the early 1960's Mandelbrot examined stable laws in connection with the distribution of income (Mandelbrot, 1961). However, it was his seminal paper *The Variation of Certain Speculative Prices* that gained the most attention (Mandelbrot, 1963). Mandelbrot initiates the discussion by showing wool price to be too peaked and leptokurtic for the traditional Gaussian hypothesis. He also remarks that the second moment varies erratically when calculated for nested samples, despite the large sample size, suggesting this as a possible indicator of infinite variance. Later, Mandelbrot demonstrates that non-Gaussian stable distributions appear to characterize cotton prices better than Gaussian distributions; he produces log plots of the empirical distribution functions and compares them to log plots of theoretical distribution functions. Inspired by Mandelbrot, Fama analyzed the log differences of stock prices. Here he finds evidence for non-Gaussianity in the tendency of the stock market to exhibit extreme deviations. Under an assumption of strict Gaussianity, an observation four standard deviations or more from the mean would occur about once every fifty year period. By contrast, "observations this extreme are observed about four times in every five-year period" (Fama, 1965, p. 50). Fama notes while typically such observations would be discarded as outliers to enforce normality, these extreme observations may in fact be representative of the underlying stock price dynamic. Indeed Fama observes the practical implications of such data trimming, "Unlike the statistician, however, the investor cannot ignore the possibility of large price changes before committing his funds, and once he has made his decision to invest, he must consider their effects on his wealth" (1965, p. 42). Additionally he notes that financial data seem to be best fit by stable distributions with $\alpha \in [1.6,1.9]$ (Fama, 1965). After the initial interest of Mandelbrot and Fama in describing financial asset prices with stable distributions, a number of reports began examining this idea in more detail, developing models and inference techniques.

Mandelbrot and Howard observe that stock price differences are Gaussian-distributed, when considered over a *fixed number of transactions*, but follow a stable distribution when examined over a *fixed time period*. On this basis they formulate a model of stock price differences; stock price differences are driven by an underlying Gaussian process, $X(n)$, on the number of transactions, n, itself a positively skewed stable process with respect to time, $n = T(t)$. Then the distribution of stock price differences over time, $Z(t)$ can be formulated $X\big(T(t)\big)$. Here $T(t)$ is a so-called directing process, subordinated to $X(n)$ (Mandelbrot & Taylor, 1967). This may be seen as an extension of the product property of univariate stable laws to stable processes.

Fama and Roll apply symmetric stable distributions to stock market data (1968) (1971). Overall they encounter better results than when restricted to Gaussian distributions, and they develop some of the first widely used estimators for stable distributions. However, Fielitz and Smith, in subsequent studies, find that asymmetric stable laws may better characterize stock data (Fielitz & Smith, 1972) (Fielitz, 1976). Examining 199 stocks, Fielitz finds that all pass Fama and Roll's hypothesis test for asymmetry at a significance level of $.05$ and 198 pass at a level of $.005$. He also mentions that Roll found a similar result when analyzing Treasury bill interest rates (Fielitz, 1976).

Bartels justifies the use of stable laws in financial and economic applications by combining theoretical and empirical reasoning (1977). He cites two theorems. The first, owing to Feller, states that a distribution exists in the domain of attraction of a non-Gaussian stable law $\Longleftrightarrow$ the distribution observes a power law decay in the tails, with power law rate $\alpha \in (0,2)$. Pareto distributions with $\alpha < 2$ fit this description, and there is considerable empirical evidence for the presence of Pareto distributions in economic settings. The second theorem, owing to Tucker, considers the iid sum of random variables, where each variable is itself the sum of several random variables in the domain of attraction of stable random variables with *different* α values. In this case the composite random variables will be in the domain of attraction of a stable law with characteristic exponent equal to the minimum α of the constituent summands. Hence, even if an economic variable is the sum of random variables, themselves the sum of several Pareto distributions, they will converge to a stable distribution, though convergence can be slow in some instances (Bartels, 1977).

Of course, the assumption of stability has its detractions. One complaint is that while stocks and other financial asset prices may observe large deviations uncharacteristic of Gaussian distributions, in all likelihood they are not true infinite variance phenomena. There is likely some value, say k, such that the absolute value of an asset's price changes will never realistically exceed k. And any bounded distribution must have a finite variance (Granger & Orr, 1972). Nevertheless if the bounds are "large" and the distribution is heavy-tailed, stable distributions may still be useful and may still provide insight into the underlying distribution. In fact, Fama and Roll note that, owing to the bounded computational resources at hand, any simulated stable variates essentially come from a truncated distribution. Nevertheless, stable distributions still provide a good fit for such pseudorandom data and, with sufficiently large bounds, sums of such data do not tend to a Gaussian distribution over any plausible quantity of summations (Fama & Roll, 1971). It should also be noted that stable distributions not only characterize income distributions and asset prices, but also many other economic variables. Recent study published in the PNAS found that several credit ratios used in financial accounting, including the Altman Z-score appear to follow an asymmetric stable distribution (Horvatić, Podobnik, Stanley, & Valentinčič, 2011).

Non-Gaussian stable distributions have been demonstrated to characterize financial data more effectively than Gaussian distributions. Still, they rarely provide a perfect description of the data. Officer finds that stock returns tend to be heavy-tailed compared to Gaussian distributions. However, his analysis found that the standard deviation is a "well-behaved" measure of scale (Officer, 1972). This may indicate that stable distributions are a bit too heavy-tailed relative to stock returns. These drawbacks considered stable distributions can still be valuable tools in financial and econometric modeling. Contemporary research in this field is focused on expounding more complex models, such as multivariate stable models, fractional Lévy motion, harmonizable processes, stable GARCH models, and stable vector autoregression.

Chapter 3: Inference

Here we outline four major approaches to inference: quantile estimators, empirical characteristic function (ECF) methods, ML estimation, and Bayesian MCMC procedures. Without claiming to exhaustively cover every category, we survey some of the major methods in each case. In Appendix B and C, two methods, the SMiN Gibbs sampler and FFT method for ML estimation, are examined in more detail and an example using both methods is presented in Appendix D. Finally it should be noted that there are several approaches other than those mentioned here. As an example, see Calzolari and Lombardi for an indirect estimation approach that uses skewed-t distributions as an auxiliary model (2008).

Since simulation of stable laws is required of some inference procedures and handy for testing any estimation method, the most common method for generating stable pseudorandom variates is presented before beginning a discussion on estimators. The algorithm for generating stable variates was proposed by Chambers, Mallows, and Stuck while working at Bell Laboratories. Using integral representations proposed by Zolotarev and earlier work by Kanter, Chambers, Mallows, and Stuck developed the algorithm which requires as a (pseudo)-random input, an exponential and a uniform variate, $W \sim \exp(1)$ and $\Theta \sim \mathrm{U}\left(-\frac{\pi}{2}, \frac{\pi}{2}\right)$ respectively. These should be independent of one another (Chambers, Mallows, & Stuck, 1976). Then, with the parameters (α, β), standard stable variates may be generated,

$$Z = \begin{cases} \dfrac{\sin \alpha(\eta+\Theta)}{(\cos \alpha\eta \cos \Theta)^{1/\alpha}} \left[\dfrac{\cos(\alpha\eta+(\alpha-1)\Theta)}{W}\right]^{1-\alpha/\alpha} & \alpha \neq 1 \\ \left(1+\frac{2}{\pi}\beta\Theta\right) \tan \Theta - \frac{2}{\pi}\beta \ln\left(\dfrac{\frac{\pi}{2} W \cos \Theta}{\frac{\pi}{2}+\beta\Theta}\right) & \alpha = 1 \end{cases} \tag{3.1}$$

Where, $\eta = \arctan\left(\beta \tan\frac{\pi\alpha}{2}\right)/\alpha$. This produces variates from the 1-parameterization. Variates from the 0-parameterizations may be obtained by shifting and scaling appropriately. Interestingly, when $\alpha = 2$, it can be shown that the algorithm recovers the familiar Box-Mueller algorithm; the CMS algorithm may be thought of as a generalization of the Box-Mueller algorithm. For a discussion on implementation and numerical issues (Chambers, Mallows, & Stuck, 1976) (Weron, 1996).

QUANTILE ESTIMATORS

Quantile estimators were among the earliest estimation procedures. The first quantile estimators were proposed by Fama and Roll for symmetric distributions (1968). Their method relies on a series expansion for the densities of symmetric stable distributions, derived a decade and half earlier by Bergström (1952). Integrating the density expansions, they obtain series expansions for the distribution functions. Evaluating the distributions for different parameter values, Fama and Roll, use the results to identify quantile-based estimators of parameter values, given a set of observations. The method is best-suited to $\alpha \in (1,2]$, which cover the range of α values typically seen in financial data. The authors suggest methods for checking the assumption of symmetry, and they note that truncated mean estimators of the location parameter tend toward lower dispersion than full mean estimators, except in the Gaussian case (Fama & Roll, 1968). Fama and Roll update their early quantile estimators in another paper, exhibiting new estimators, demonstrating goodness-of-fit tests for the non-Gaussianity assumption, and proposing a resampling method for assessing stability against alternative non-Gaussian hypotheses (e.g. a mixture of normals) (1971).

Quantile estimators have largely given way to more accurate, albeit more intensive estimation procedures. Still, quantile estimators can be utilized in situations requiring high efficiency or when very precise results are not necessary. Quantile estimators also serve as convenient first guesses for estimation procedures that require reasonable initial approximations. By far the most popular and widely used quantile estimator is that of McCulloch (1986). This method is effective for $\alpha \in [.6,2]$ and does not present any discontinuities as $\alpha \to 1$ (regardless of skew). Unlike the Fama and Roll estimator it accommodates asymmetry, supporting estimation of β in its full parameter range. Furthermore the slight asymptotic bias of the Fama and Roll estimators for α and the scale γ are corrected in the McCulloch estimator. However, the method does lose some efficiency relative to the Fama and Roll estimator in calculating the location parameter δ. McCulloch derives asymptotic variances and covariances and demonstrates asymptotic normality. It should be noted that the estimates for some parameters do exhibit correlations (McCulloch, 1986). Nevertheless the relative accuracy, ability to accommodate skew, joint continuity of parameters, and the efficiency of the McCulloch estimator make it a viable, efficient approximation method.

EMPIRICAL CHARACTERISTIC FUNCTION METHODS

While stable densities and distributions do not generally exhibit tractable forms, their characteristic functions do. This has prompted some to examine empirical characteristic function (ECF) methods of estimation. The ECF for a set of iid observations $(x_1, x_2, \dots., x_n)$ is expressed,

$$\varphi_n(\tau) = \frac{1}{n}\sum_{k=1}^{n} \exp(i\tau x_k) \tag{3.2}$$

Press seems to have been the first to examine ECF methods (1972). Press' proposed estimator is contingent on the choice of τ values used to evaluate the ECF. The selection of τ is important for convergence to the population parameter values. Lacking a principled manner for choosing τ, this method can return poor results. Nonetheless, Press did introduce concepts that would play an important role in the development of other ECF methods. He defines the following metric, the minimum weighted r-th mean, for assessing the fit of different parameter values given the ECF (Press, 1972),

$$h(\alpha, \beta, \delta, \gamma) = \int_{-\infty}^{\infty} |\varphi(\tau) - \varphi_n(\tau)|^r W(\tau) d\tau \qquad r \geq 1 \tag{3.3}$$

Where $\varphi(\tau)$ is the characteristic function for the input values $(\alpha, \beta, \delta, \gamma)$. So $h(\cdot,\cdot,\cdot,\cdot)$ is a measure of the closeness of the estimated characteristic function, $\varphi(\tau)$ and the ECF, $\varphi_n(\tau)$. Paulson, Holcomb, and Leitch consider the case where $r = 2$ and $W(\tau) = e^{-\tau^2}$. They estimate $(\alpha, \beta, \delta, \gamma)$ by minimizing $h(\cdot)$ through iterative renormalization (Koutrouvelis, 1980). Wiener proposes an iterative regression-based method for calculating the parameter estimates in the symmetric case (Koutrouvelis, 1980). Koutrovelis extends this approach to accommodate asymmetry. The basic idea of Koutrouvelis' approach is to use mathematical manipulations of the ECF to write the four parameters in terms of two regression equations. Koutrouvelis uses these regression equations in an iterative scheme to identify parameter estimates. He shows that his parameter estimates are consistent and approximately unbiased, even for moderately large sample sizes (Koutrouvelis, 1980). Kogon and Williams present another extension of the regression-based ECF method. They substitute the 0-parameterization for the 1-parameterization used by Koutrouvelis. This eliminates the discontinuity as $\alpha \to 1$ when

$\beta \neq 0$. The result is a tradeoff with significantly improved performance for α in a neighborhood of 1, and slightly worse performance at other values of α. Kogon and Williams' procedure is significantly faster than Koutrouvelis' method though still significantly slower than the McCulloch's estimator. In fact, Kogon and Williams employ McCulloch's estimator as an initial estimate (Weron, 1996).

Comparing effective ECF methods to the most effective quantile methods, there is a typical tradeoff between the increased accuracy of the ECF methods and the greater efficiency of the quantile methods. An additional advantage of ECF methods over quantile methods is that they may be generalized for inference on time series whereas quantile estimators are only suited to location-scale models. Knight and Yu explore ECF methods for inference in stable ARMA models (Knight & Yu, 2002). Their approach is to group a set of T time series observations $\{y_j\}_{j=1}^{T}$ into bins of size p. These bins are defined $x_j = \{y_j, \ldots, y_{j+p}\}$ for $j = 1, \ldots, T-p$. The result is $T - p$ overlapping bins of size p. These are used to calculate the ECF over the bins, $\varphi_n(\vec{\tau}, \vec{x}_j) = E\left[\exp\left(i\vec{\tau}^{\mathrm{T}}\vec{x}_j\right)\right] = \frac{1}{n}\sum_{j=1}^{n} \exp(i\vec{\tau}^{\mathrm{T}}\vec{x}_j)$ where $\vec{\tau} = (\tau^1, \ldots, \tau^{p+1})$. From here the method is a multivariate extension of the Press' minimum weighted r-th mean, for $r = 2$ (Knight & Yu, 2002),

$$I_n(\theta) = \int \ldots \int |\phi(\vec{\tau}|\theta) - \varphi_n(\vec{\tau})|^2 W(\vec{\tau}) d\tau^1 \ldots d\tau^{p+1} \tag{3.4}$$

Where $W(\tau)$ is a weight function. Depending on the choice of weighting, such an integral may have a closed-form solution or, in most cases, will require numerical integration. Knight and Yu include a discussion on suitable weighting functions and optimal bin size before applying the method to stable ARMA models. They also prove the strong consistency and asymptotic normality of the ECF estimators, given appropriate regularity conditions (Knight & Yu, 2002).

MAXIMUM LIKELIHOOD ESTIMATION

Compared with ECF methods and quantile estimators, maximum likelihood estimators (MLE) have the greatest accuracy. They are also the most computationally intensive, requiring numerical integration. There are two general approaches, direct numerical integration (DNI) and a fast Fourier transform (FFT) method. This report will briefly introduce DNI, before delving into greater depth with the FFT method. However, before discussing either, relevant theory on ML estimation with stable distributions will be presented.

Most of the theory of ML estimation for stable distributions was established by DuMouchel. Here we present relevant facts and theorems. For greater discussion and proofs see (DuMouchel, 1973b).

Theorem (non-existence of maximum): For any observations $x_1, x_2, \dots, x_n$, if α and δ are to be estimated, then $\mathcal{L}(\alpha, \delta)$ has no maximum for $\alpha \in (0,2]$, $\delta \in (-\infty, \infty)$, rather $\mathcal{L}(\alpha, \delta) \to \infty$ as $(\alpha, \delta) \to (0, x_i)$, for $i = 1, \dots, n$.

While this may seem troubling, the centroid of $\mathcal{L}$ is not affected by this phenomenon, and the undesirable behavior of $\mathcal{L}$ vanishes when restricting α to be greater than some $\varepsilon > 0$. Then we may present the following theorem,

Theorem (consistency and asymptotic normality): Let $\hat{\theta}_n$ denote the MLE for $\theta = (\alpha, \beta, \delta, \gamma)$, $\hat{\alpha}_n$ the MLE for α, and θ_0 the true parameter value. If $\hat{\alpha}_n$ is restricted to be greater than some $\varepsilon > 0$, then the MLE $\hat{\theta}_n$ is consistent and asymptotically normal, with variance-covariance matrix $\Im^{-1}(\theta_0)$ where $\Im$ is Fisher information matrix, so long as θ_0 exists in the interior of the parameter space. Thus the cases $\alpha \leq \varepsilon$, $\alpha = 2$, and $\beta = \pm 1$ are excluded. Additionally, if the 0-parameterization is not used, the case of $\alpha = 1$, $\beta \neq 1$ is also excluded.

Note that for true parameter values on the boundary of the parameter space, the limiting distribution tends to a degenerate and ML estimation is superefficient. DuMouchel provides a discussion of the necessary conditions for the last theorem to hold, and he provides proofs that they are met (1973b).

Nolan introduces DNI as a procedure for numerically approximating the density of a stable random variable (1999). Based on similar integral representations by Zolotarev, Nolan derives an integral representation for the 0-parameterization; other parameterizations encounter numerical issues in a neighborhood of α. Integrating these representations for each point yields the desired density value. Nolan's paper discusses adaptively selecting integration bounds and numerical issues. For accurate tail density estimation quadrature may be too computationally expensive. For this situation, the series expansions of Bergström can offer more accurate and less computationally cumbersome density estimates (Nolan, 1999). Based on this approach, Nolan created the program 'STABLE,' elements of which are the basis for the R code package 'stabledist.' DNI may be used to obtain density estimates for data points thought to come from a stable distribution. This enables numerical calculation of the likelihood or log likelihood of the parameters given the data. ML estimation is carried out through numerical maximization.

Chenyao, Doganoglu, and Mittnik introduce an FFT algorithm for calculating estimates of stable densities through inversion of the characteristic function. They compare their method to DNI (Chenyao, Doganoglu, & Mittnik, 1999). Chenyao, Doganoglu, Mittnik, and Rachev introduce ML estimation through the FFT method (Chenyao, Doganoglu, Mittnik, & Rachev, 1999).[2] The FFT method works best for $\alpha \in [1,2]$. This covers the range of values typically encountered in financial and econometric applications (Chenyao, Doganoglu, & Mittnik, 1999).

The characteristic function may be inverted through an application of the Fourier transform to the characteristic function, $f(x|\alpha,\beta,\gamma,\delta) = \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-ixt}\varphi(t|\alpha,\beta,\gamma,\delta)dt$. The density is approximated by inverting the characteristic function over a grid of points on a bounded interval of values for x, centered at 0. In turn, the Fourier transform integral is approximated by limiting the bounds of integration to a bounded interval centered at 0, i.e. from $(-\infty,\infty)$ to $(-c,c)$ for $c < \infty$. Then the integral may be approximated by a quadrature rule over a grid of points on these new bounds of integration. Thus the final approximation to the density is discretized into the form of a discrete Fourier transform (DFT). Rearranging this DFT one obtains a representation of the density function that can be computed as an FFT of the characteristic function multiplied by a normalizing constant. The exact expression of the DFT and the layout of the grid points will depend on the quadrature rule used. Chenyao, Doganoglu, and Mittnik introduce a left point rule (1999). However, Menn and Rachev note that the midpoint rule yields better accuracy; in fact they find surprisingly large gains in accuracy for similar computational burden (Menn & Rachev, 2006). We exhibit the derivation of the DFT for the midpoint rule in Appendix B and summarize Menn and Rachev's discussion on the numerical error. The density for a sample point x, in the 0-parameterization, may be expressed as a function of the density of the standardized sample point $z = \frac{x-\delta}{\gamma}$ as $f(x|\alpha,\beta,\delta,\gamma) = \frac{1}{\gamma}f(z|\alpha,\beta,0,1)$.

[2] Ironically both appear in the same issue as Nolan's introduction of DNI.

Due to numerical considerations, it is better to calculate the density values for a standard variable and rescale accordingly (Chenyao, Doganoglu, & Mittnik, 1999).

The FFT method allows for the numerical definition of the density function, enabling ML estimation. For a set of observations $(x_1, \dots, x_n)$ the density is approximated by the FFT algorithm and the likelihood at each x_i is either matched or interpolated given the grid of points obtained. Linear interpolation is typically accurate enough for application (Chenyao, Doganoglu, & Mittnik, 1999). However, if greater accuracy is needed, a spline may be used near the mode, and the Bergström series expansion may be employed at the tails (Menn & Rachev, 2006). With obtained density approximations for the data points, the likelihood or log likelihood given the dataset can be constructed. Finally the MLE is estimated by numerical maximization of the likelihood function. A grid search method may be used on a constrained parameter space, with the unbounded location and scale parameter restricted to a "reasonable" range. Alternatively a gradient-based search might be employed. Chenyao, Doganoglu, Mittnik and Rachev opt for an unconstrained maximization method, where the parameters are transformed so their ranges are unbounded on $\mathbb{R}$. For further details see (Chenyao, Doganoglu, Mittnik, & Rachev, 1999).

An advantage of the ML methods is that they can be used to fit a broad set of models. Chenyao, Doganoglu, Mittnik, and Rachev extend their FFT ML estimation procedure to regression, ARMA, ARMAX, and GARCH models. The setup for ARMA models is presented here for reference to the application example in Appendix D. For an $ARMA(p, q)$ model,

$$y_t = \mu + \sum_{i=1}^{p} \phi_i y_{t-i} + \sum_{i=1}^{q} \vartheta_i \varepsilon_{t-i} + \varepsilon_t \tag{3.5}$$

Where $\varepsilon_t \overset{iid}{\sim} S_\alpha(\beta, \gamma, 0)$. The conditional likelihood may then be stated,

$$\ln \mathcal{L}(\mu, \phi, \vartheta, \theta|\varepsilon) = \sum_{i=1}^{T} \ln(\varepsilon_i|\mu, a, b, \theta) \tag{3.6}$$

Where $\varepsilon_t = y_t - \mu - \sum_{i=1}^{p} \phi_i y_{t-i} - \sum_{i=1}^{q} \vartheta_i \varepsilon_{t-i}$, $\phi = (\phi, \dots, \phi_p)$, $\vartheta = (\vartheta_1, \dots, \vartheta_q)$, and $\theta = (\alpha, \beta, \gamma)$. we condition on the first p realizations $y_1, \dots, y_p$ and set the corresponding disturbances $\varepsilon_p, \dots, \varepsilon_{p-q+1}$ to 0. Restrictions on the possible values of ϕ and ϑ will insure stationarity and invertibility (Chenyao, Doganoglu, Mittnik, & Rachev, 1999).

Bayesian MCMC Methods

Markov Chain Monte Carlo (MCMC) methods for inference in Bayesian models have enjoyed increasing popularity in the last couple decades. Accordingly, some researchers have focused their efforts on fitting stable distributions to such Bayesian MCMC methods.

The first effort in this area was by Buckle (1995). He notes that the lack of a general closed-form density had hindered Bayesian application of stable distributions and recognizes the potential for MCMC computation to accommodate stable distributions in a Bayesian framework. His method relies on another of Zolotarev's integral representations. Rearranging the formula, Zolotarev shows that a closed form expression for the stable density may be obtained, conditional on a set of auxiliary variables. By sampling these auxiliary variables, one for each observation, the joint density can be expressed analytically. Buckle then uses a Gibbs sampler to obtain draws from the four parameters. The parameters (α, β, δ) have posterior conditional densities that are undulating and multimodal. Buckle corrects for this through transformations of the auxiliary variables that produce unimodal conditional densities. This process necessitates an application of the Newton-Raphson method for each auxiliary variable on each parameter draw. For sampling these parameters, Buckle proposes embedding a Metropolis-Hastings step, as direct sampling for the Gibbs step is not possible. However, Buckle also entertains the possibility of using adaptive rejection sampling schemes to sample the parameters. The scale parameter can be transformed as $\gamma^{\alpha/(\alpha-1)} = \theta$; then an inverse gamma prior on θ yields an inverse gamma posterior that may be readily sampled and transformed to obtain γ draws. The entire procedure is very computationally intensive. The need to sample an auxiliary variable for each observation on each iteration and the need to apply the Newton-Raphson method to each auxiliary variable on each iteration when obtaining draws for each of the parameters (α, β, δ) rapidly scales the computational burden. On top of this, the method requires function evaluations from a couple of complicated functions involving sines and cosines for each parameter draw. Another drawback is that some parameters show consistent correlation. In particular, α and γ both affect the general spread of the distribution, albeit in different ways, and hence tend to exhibit negative correlation (Buckle, 1995). Finally the parameter estimates returned are for the 2-parameterization, an undesirable form. Estimates for other parameterizations can be found through applying the appropriate relations. Still, the 2-parameterization leads to discontinuity and numerical instability in a neighborhood of $\alpha = 1$. Nonetheless, the method provides a method for introducing stable distributions into a Bayesian framework. Buckle demonstrates that for a set of stock price differences, his model provides a better fit than a comparable Gaussian model (1995). Buckle introduces his method for location-scale models. Qiou and Ravishanker extend Buckle's method to ARMA and VARMA models (1998a) (1998b). They obtain reasonable results. However, they initialize their method with the maximum likelihood estimate and scale the variance matrix of the normal distribution in the Metropolis-Hastings step with the inverse of the Fisher information matrix (Qiou & Ravishanker, 1998a). It is not clear how the method would perform in the absence of this information, especially since obtaining it can be a numerically intensive procedure itself.

A few researchers have examined an alternative MCMC approach for inference on time series in the presence of symmetric stable innovations (Godsill S. J., 1999) (Godsill S. J., 2000) (Godsill & Kuruoglu, 1999) (Tsionas, 1999). The approach takes advantage of the product property to rewrite the symmetric stable disturbances ν_t, with characteristic exponent α, as Gaussian random variables, conditioned on positively skewed stable random variables λ_t with characteristic exponent $\alpha/2$. Such an approach can be viewed as a scale mixture of normal (SMiN) that allows for inference in the familiar Gaussian setting. Godsill and Kuruoglu consider MCMC inference for an $AR(p)$ model, though they note the applicability of the method is far broader, including the non-linear case (1999). The main difference between a model with Gaussian errors and the proposed model with stable innovations, is that on top of the usual Gibbs steps another step is added to sample the positively skewed λ_t. This sampling may be carried out with rejection sampling or a Metropolis-Hastings step. Using power law expansions for the tail of appositively skewed distribution, it can be shown that an inverted gamma distribution may be used to generate more accurate tail approximations for extreme values (Godsill S. J., 1999) (Godsill & Kuruoglu, 1999). Later, Godsill introduces a slice sampler as yet another method for sampling the λ_t (Godsill S. J., 2000). Since the positively skewed stable variates are sampled in this procedure, there is no need to directly evaluate the density of any stable laws; this the chief advantage of this method. Tsionas presents a similar sampling scheme for econometrics, albeit in a more formal setting. He includes a discussion on extensions to GARCH models and proposes a method for estimating the α

parameter via an application of the FFT on each iteration (Tsionas, 1999). All other papers cited here omit a discussion of estimating α, assuming its value is already known. In practice, its value may be estimated with a reasonable and simple estimator such as the McCulloch estimator. For more details of SMiN method and the details of the $AR(1)$ model case, see Appendix C. For an example application, see Appendix D.

Godsill and Lombardi extend the SMiN method to a TVAR Markovian state space model with Gaussian process noise and symmetric stable observation noise (Godsill & Lombardi, 2004). They use particle filtering and smoothing for inference. Of particular note, their method accommodates inference on α, a static parameter that is not easily amenable to particle filtering analysis. They test their method on old audio data from gramophone disk recordings that have been degraded by non-Gaussian clicks. The results are found to be effective as the assumption of stable observation noise accounts for the non-Gaussianity induced by the degraded recording (Godsill & Lombardi, 2004).

CONCLUDING REMARKS

Stable distributions are an important class in the theory of probability. They are sum stable, and they are the limiting distributions in a generalized central limit theorem. Stable distributions constitute a generalization of the Gaussian family that can accommodate heavy-tails and skew. Theoretical and empirical justifications exist for their use in a series of practical settings. Despite the difficulties presented by the lack of closed-form densities, the recent development of estimation methods has made stable distributions viable and valuable tools in statistical modeling.

Appendix A: Stable Plots

Here we present standardized stable density plots for different values of α and β.

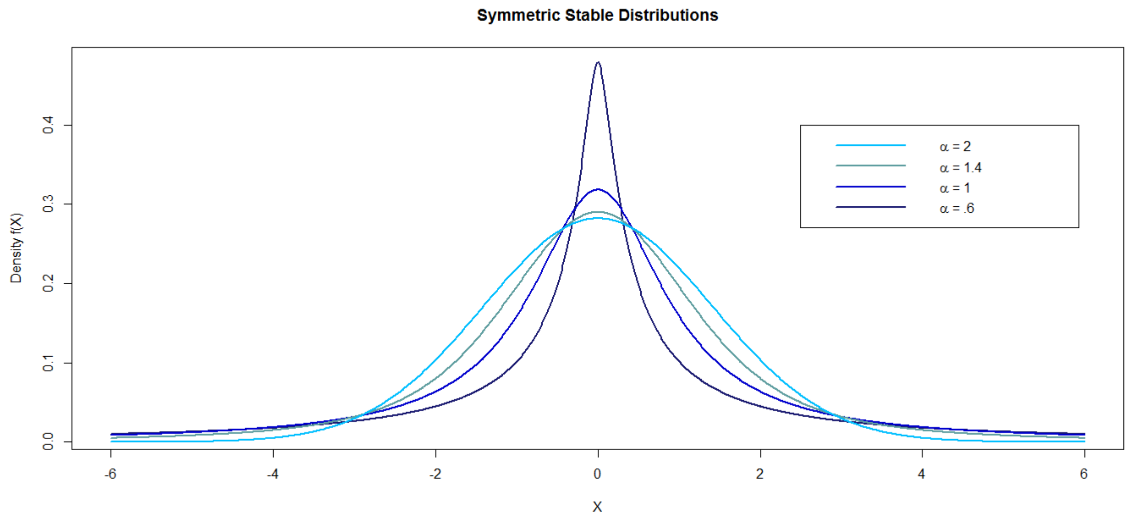

Figure 1: a plot of symmetric stable densities over arrange of α parameter values. Note the increased peakedness and heavier tails as $\alpha \to 0$. This plot is the same for the 0 and the 1-parameterization.

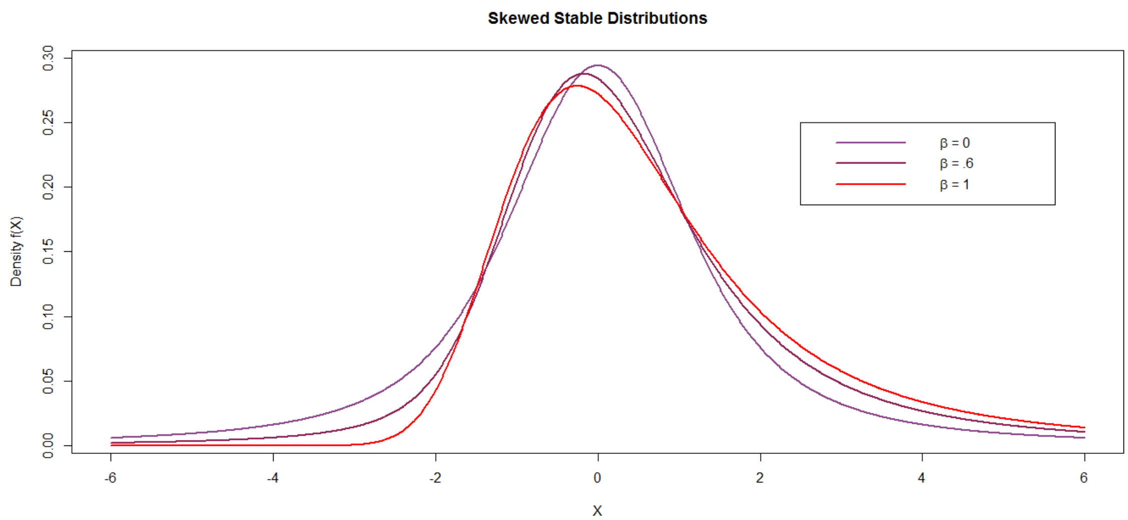

Figure 2: a plot of skewed stable densities over non-negative β values, with α fixed at 1.3. The plots represent the 0-parameterization.

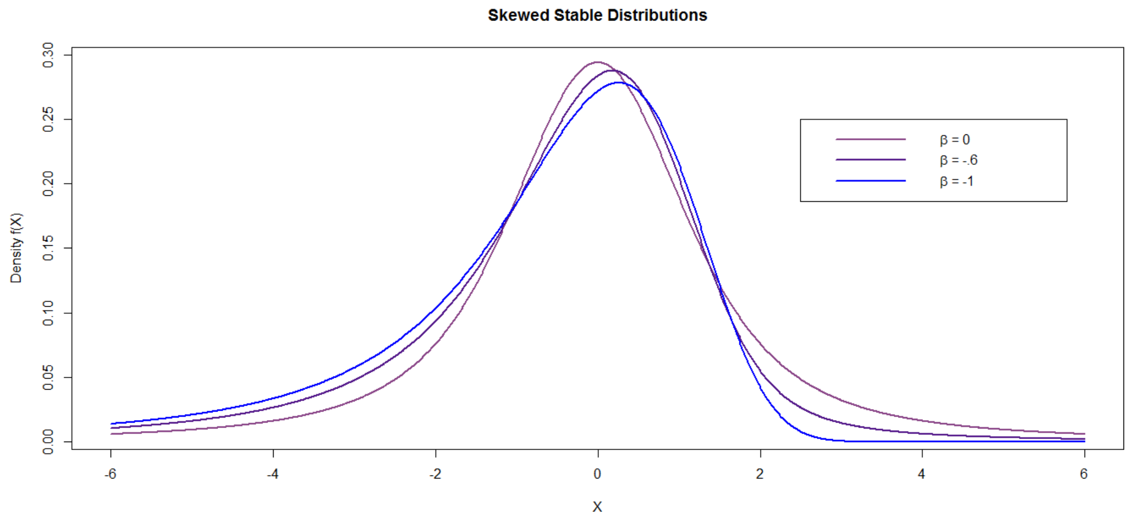

Figure 3: a plot of skewed stable densities over non-negative β values, with α fixed at 1.3. The plots represent the 0-parameterization.

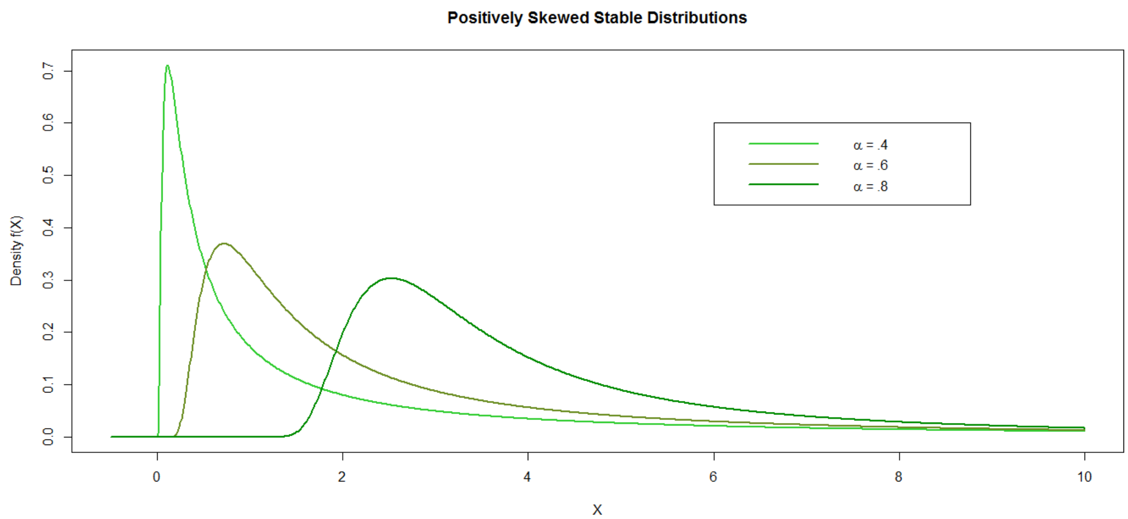

Figure 4: Plots of positively skewed stable distributions over several α values. The plots are in the 1-parameterization.

Appendix B: Midpoint Rule FFT and Errors

Menn and Rachev give a formulation of the FFT applied to the mid-point rule, noting that usually yields higher accuracy than the left or right point rules (Menn & Rachev, 2006). Let $N = 2^m \in \mathbb{N}$ and $a \in \mathbb{R}^+$ with $h = \frac{2a}{N}$. We define an equidistant grid over the interval $[-a, a]$ via $t_j = -a + jh,\ j = 0,1, \dots, N$. The midpoints are given $t_j^* = .5(t_j + t_{j+1}),\ j = 0,1, \dots, N-1$.

We then define vectors $\vec{y} \in \mathbb{C}^N$ and $\vec{c} \in \mathbb{C}^N$, as

$$y_j = (-1)^j \varphi(t_j^*), \qquad j = 0, \dots, N-1$$
$$c_k = h(-1)^k \exp\left(-i\left(\frac{\pi}{N}\right)k\right) i$$

For $x_k = -\frac{N\pi}{2a} + \frac{\pi}{a}k,\ \ k = 0,1, \dots, N-1 \qquad x_0 = -\frac{N}{2a}\pi$ and $x_{N-1} = \frac{N\pi-2}{2a}$

We have,

$$\sum_{j=0}^{N-1} \varphi\left(t_j^*\right) \exp\left(it_j^* x_j\right) h = c_k \mathrm{DFT}(y)_k$$

Proof:

$$t_j^* x_j = \left(-a + \frac{h}{2} + hj\right)\left(-\frac{N\pi}{2a} + \frac{\pi}{a}k\right) = \left(-a + \frac{a}{N} + \frac{2a}{N}j\right)\left(-\frac{N\pi}{2a} + \frac{\pi}{a}k\right)$$

$$= \frac{N\pi}{2} - \pi k - \frac{\pi}{2} + \frac{\pi k}{N} - \pi j + \frac{2\pi k}{N} j$$

$$\int_{-a}^{a} \varphi(t) \exp(-itx_k)\, dt =$$

$$\overset{MPR}{\approx} \sum_{j=0}^{N-1} \varphi(t_j^*) \exp(it_j^* x_k) (t_{j+1} - t_j)$$

$$= \frac{2a}{N} \sum_{j=0}^{N-1} \varphi(t_j^*) \exp\left(i\left(\frac{N\pi}{2} - \pi k - \frac{\pi}{2} + \frac{\pi k}{N} - \pi j + \frac{2\pi k}{N} j\right)\right)$$

$$= \frac{2a}{N} \underbrace{\exp\left(i\frac{N\pi}{2}\right)}_{=1} \underbrace{\exp(i\pi k)}_{=(-1)^k} \underbrace{\exp\left(\frac{i\pi}{2}\right)}_{=i} \exp\left(-i\left(\frac{\pi}{N}\right)k\right) \sum_{j=0}^{N-1} \underbrace{\exp(i\pi j)}_{=(-1)^j} \varphi(t_j^*) \exp\left(-i\frac{2\pi k}{N} j\right)$$

$$= c_k \mathrm{DFT}(y)_k$$

The quantity $\mathrm{DFT}(y)_k$ may be tabulated by applying an FFT algorithm. Multiplying by the normalizing constant c_k yields the density. The values of h and m, where $N = 2^m$ are chosen by the programmer. The values $h = .01$ and $m = 13$ deliver a sufficiently accurate result for many numerical applications. For a discussion on h and m, and for a comparison with DNI, see (Chenyao, Doganoglu, & Mittnik, 1999).

Menn and Rachev identify three sources of error (2006),

1. The first source of error, denoted $\varepsilon_1(h, N)$, is the reduction of the integral bounds to the compact set $(-a, a) = \left(-\left(\frac{N}{2}\right)h, \left(\frac{N}{2}\right)h\right)$.

2. The second error source, denoted $\varepsilon_2(h, N)$, occurs from applying an approximation rule (left point, mid-point, etc).

3. The third source of error is interpolation error, denoted $\varepsilon_3(h, N)$

For an arbitrary point x, with $x \in \left(-\frac{\pi}{h}, \frac{\pi}{h}\right)$, we have the following error equation,

$$f(x) - \hat{f}(x) = \varepsilon_1(h, N) + \varepsilon_2(h, N) + \varepsilon_3(h, N)$$

The sum of the imaginary parts will vanish in the limit. For most applications it is sufficient to evaluate only the real parts of the errors. For $\alpha > 1$, $\varepsilon_1 \leq 2e^{-a}$. So for a sufficiently large, this need not be of concern for most application (Menn & Rachev, 2006).

In assessing the quadrature error, Menn and Rachev note that in the absence of a general closed form density for stable distributions, an analytical derivation is seemingly impossible. Instead they resort to carrying out the algorithm for a range of parameter values with a suitable choice of grid points. All points were evaluated at the grid points to remove any interpolation error. They summarize their conclusions as follows (Menn & Rachev, 2006),

1. The relative error is only acceptable in a narrow region about the origin. For large $|x|$ the relative error is unacceptable, judging by the graphs this acceptable region seems range from $(-8,8)$ to $(-10,10)$, depending on specific parameters.

2. Given they require the same computational effort, it is surprising that the midpoint rule relative errors are consistently about twice as good as the left point rule relative errors. The two errors always have opposite signs

3. As α decreases, the relative error increases

4. In the case of skewed stable distributions, the error for the lighter tail is greater. This effect increases as skewness increases.

For assessing points in the tails of the distributions (i.e. $|x|$ large), Menn and Rachev, explore using Bergstrom's series expansion for greater accuracy. Interpolation error, by contrast, decreases in the tails, but play a significant role close to the mode. The interpolation error may on the order of two magnitudes of the sum of ε_1 and ε_2.If high accuracy is desired, Menn and Rachev suggest using a cubic spline to interpolate in-between grid points, particularly close to the mode. For further discussion, results, and sources, see (Chenyao, Doganoglu, & Mittnik, 1999), (Chenyao, Doganoglu, & Mittnik, 1999), (Menn & Rachev, 2006).

Appendix C: SMiN Gibbs Sampler

Applying the product property to a noise process with iid symmetric stable disturbances, $\nu_t \sim S_\alpha(0,0,\gamma)$ with $\alpha \in (0,2)$ coerces the ν_t to be conditionally Gaussian given the appropriate positively skewed stable scale variables,

$$\nu_t \sim \mathrm{N}(0,2\gamma^2\lambda_t) \qquad \lambda_t \sim S_{\alpha/2}(1,0,\xi) \quad \xi = \left(\cos\frac{\pi\alpha}{4}\right)^{2/\alpha}$$

This can be interpreted as a SMiN representation of the original noise process,

$$f_V(v) = \int_0^\infty N(v|0,2\gamma^2\lambda) f_\Lambda(\lambda) d\lambda$$

Here, $p(\nu|\lambda) = \mathrm{N}(v|0,2\gamma^2\lambda)$ and $p(\lambda|v) \propto p_\Lambda(\lambda)\mathrm{N}(v|0,2\gamma^2\lambda)$. Godsill and Kuruoglu, while stressing the applicability of the present method to nonlinear cases, develop a general linear model. Let $\vec{\theta}$ be vector of observations, G a fixed-basis matrix tying the observations $\vec{x}$ to the parameters, and $\vec{\nu}$ a random vector such that each component $\nu_i \sim S_\alpha(0,0,\gamma)$. Then the model,

$$\vec{y} = G\vec{\theta} + \vec{\nu}$$

May be expressed in a conditionally Gaussian framework, where Λ is diagonal matrix with diagonal values $\left(\lambda_1^{-1/2}, \lambda_2^{-1/2}, \dots, \lambda_n^{-1/2}\right)$,

$$f\left(\vec{y}\middle|\vec{\theta},\Lambda,\gamma\right) \propto \exp\left(-\frac{1}{4\gamma^2}\left\|\Lambda\left(\vec{y} - G\vec{\theta}\right)\right\|_2^2\right)$$

In appendix D the preceding model is implemented for an $AR(1)$ process. Here we exhibit the details for the parameterization of this specific linear model. Let $\sigma^2 = 2\gamma^2$. Then we may denote the $AR(1)$ model

$$y_t = a + by_{t-1} + \nu_t$$

With the following priors on each parameter,

$$a \sim \mathrm{N}(a_0, \tau_a) \qquad b \sim \mathrm{N}(b_0, \tau_b) \qquad \sigma^2 \sim \mathrm{IG}(c, d)$$

And finding the following posterior densities for these parameters

$$a \sim N(a_1, \varsigma_a) \qquad \varsigma_a = \left(\frac{1}{\tau_a} + \frac{n}{\sigma^2}\right)^{-1} \qquad a_1 = \varsigma_a\left(\frac{a_0}{\tau_a} + \frac{\sum_{i=2}^{n} y_i - by_{i-1}}{\sigma^2}\right)$$

$$b \sim N(b_1, \varsigma_b) \qquad \varsigma_b = \left(\frac{1}{\tau_b} + \frac{\sum_{i=1}^{n-1} y_i^2}{\sigma^2}\right)^{-1} \qquad b_1 = \varsigma_b\left(\frac{b_0}{\tau_b} + \frac{\sum_{i=2}^{n}(y_i - a)y_{i-1}}{\sigma^2}\right)$$

$$\sigma^2 \sim IG(c_1, d_1) \qquad c_1 = c + \frac{n}{2} \qquad d_1 = d + \frac{1}{2}\sum_{i=2}^{n}\left(\lambda_{i-1} * (y_i - a - b * y_{i-1})\right)^2$$

All that remains is to sample the auxiliary variables λ_t. Here we present Godsill's methods, a rejection sampler, a Metropolis-Hastings algorithm, and a slice sampler (Godsill S. J., 1999).

Rejection Sampler

The likelihood $\mathrm{N}(\nu_t | 0, \lambda_t\sigma^2) \leq \frac{1}{\sqrt{2\pi\nu_t^2}} \exp\left(-\frac{1}{2}\right)$. Starting with $t = 1$,

1. Sample $\lambda_t \sim S_{\alpha/2}(0,0,\xi)$

2. Sample $u \sim U\left(0, \frac{1}{\sqrt{2\pi\nu_t^2}}\exp\left(-\frac{1}{2}\right)\right)$

3. If $u > N(\nu_t|0, \lambda_t\sigma^2)$ go to 1. Otherwise increment t by 1. If $t = n + 1$, stop, else go back to 1

Metropolis-Hastings

Using $S_{\alpha/2}(0,0,\xi)$ as the proposal distribution for the target $f(\lambda_t|\nu_t, \sigma^2)$, the acceptance probability is,

$$\alpha(\lambda_t'|\lambda_t) = \min\left(1, \frac{N(\nu_t|0, \lambda_t'\sigma^2)}{N(\nu_t|0, \lambda_t\sigma^2)}\right)$$

Slice Sampler

The slice sampler procedure for this SMiN is introduced in Bayesian format. For a greater discussion of the slice sampler see (Godsill S. J., 2000).

The basic idea is to introduce auxiliary uniform variables u_t. The joint density for the auxiliary variables λ_t and the auxiliary uniforms u_t is,

$$f(\lambda_t, u_t|\nu_t, \sigma^2) \propto N(\nu_t|0, \lambda_t\sigma^2) \times S_{\alpha/2}(1,0,\xi) \times U\big(u_t\big|0, N(\nu_t|0, \lambda_t\sigma^2)\big)$$

Sampling from the joint density can be carried out in two steps,

1. $f(\lambda_t|u_t, \nu_t, \sigma) \propto \begin{cases} S_{\alpha/2}(1,0,\xi), & u_t < N(\nu_t|0, \lambda_t\sigma^2) \\ 0, & \text{elsewhere} \end{cases}$

2. $f(u_t|\lambda_t, \nu_t, \sigma) = U\big(0, N(\nu_t|0, \lambda_t\sigma^2)\big)$

For more information and discussion on the SMiN MCMC method, see (Godsill S. J., 1999) (Godsill & Kuruoglu, 1999) (Godsill S. J., 2000) (Tsionas, 1999).

Appendix D: SMiN and FFT MLE Example

Here we apply the SMiN Gibbs sampling method, detailed in Appendix C. Using the output from the Gibbs sampler, we then apply the midpoint rule FFT, in Appendix B, for ML estimation. Taking 331 days of S&P 500 index prices, we fit an $AR(1)$ model with stable innovations as an illustration, without intending to suggest the model is the most suitable for the given data. We begin by obtaining a first estimate for α using the Fama and Roll estimators. Using the least squares method (Knight & Yu, 2002) crude estimates of the AR parameters are obtained, from which residuals are calculated. The Fama and Roll estimator then returns an estimate of $\alpha = 1.7$. This value of α initializes the SMiN Gibbs sampler. The procedure is set to run 4000 iterations. Since starting values were arbitrary, we throw away the first 2000 samples, largely based on a conservative ad-hoc analysis of convergence from plots. The parameter values returned by the SMiN Gibbs sampler are $\alpha = 1.7$ (pre-determined), $\beta = 0$ (fixed), $\delta = 8.6572$, $\gamma = 7.0483$, $\phi = .9956$.

Next the FFT midpoint rule algorithm is used for ML estimation. Of particular interest is the fact that this procedure, unlike the SMiN Gibbs sampler, can accommodate asymmetry. We define a function to calculate the negative log likelihood (NLL), given the parameter values and the dataset. To maximize the likelihood, we minimize the aforementioned function using a gradient-based, constrained optimizer, restricting the parameter ranges accordingly: $\alpha \in (1,2)$, $\beta \in (-1,1)$, $\delta \in (0,25)$, $\gamma \in (0,15)$, and $\phi \in (0,1)$ To initiate, we use the estimates from the SMiN sampler, which returns a negative log likelihood 1284.04 . After 40 iterations, the optimizer reaches a (possibly local)[3] minimum for parameter values $\alpha = 1.78$, $\beta = -.77$, $\delta = 8.652$, $\gamma = 7$, $\phi =$

[3] To insure a global maximum is attained, we might want to use a grid-based optimizer instead, but we avoid delving into details here.

.9966, with a modestly improved NLL of 1269.381. Indeed examining the estimated disturbance histogram plots, fitted with the estimated density, the FFT method appears to correct for some of the skew seen in the SMiN plot. To some degree, this better fits can be explained by a couple factors: 1) The FFT method directly estimates α, whereas for the SMiN method we used unsophisticated estimation procedures to initialize the procedure with an appropriate, approximate, static α value 2) The FFT method allows α and β to vary; with more parameters free to vary, a better fit is expected. Nonetheless, the results are in agreement with studies that find empirical evidence for asymmetry in financial data, based on hypothesis testing (Fielitz & Smith, 1972) (Fielitz, 1976).

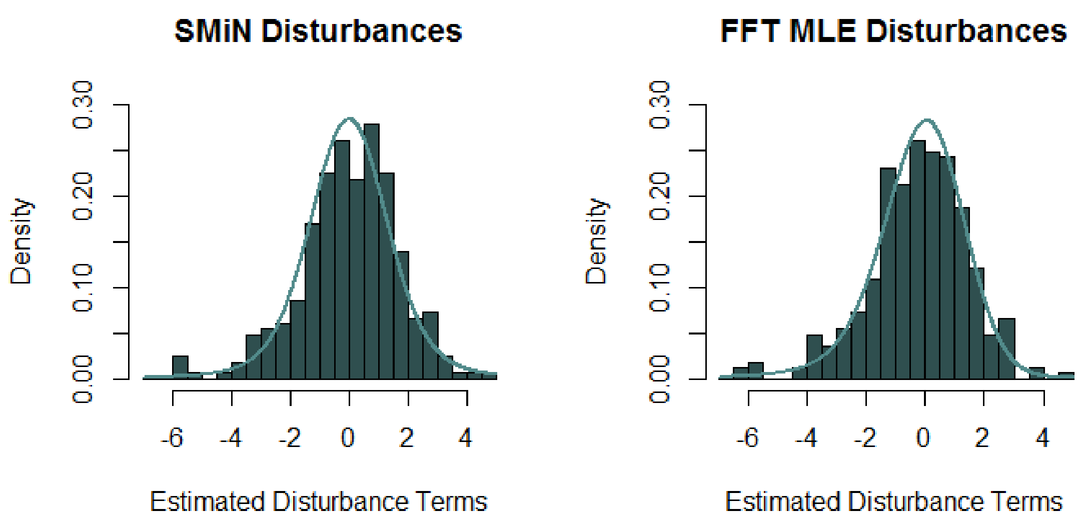

Figure 5: The estimated disturbance term and fitted stable densities for the residuals of the $AR(1)$ process under the SMiN and FFT approaches.

References

Bartels, R. (1977, May). On the Use of Limit Theorem Arguments in Economic Statistics. *The American Statistician*(31), 85-87.

Bergström, H. (1952). On Some Expansions of Stable Distributions. *Arkiv for Matematik, II*, 375-378.

Brockwell, P. J., & Cline, D. B. (1985). Linear Prediction of ARMA Processes with Infinite Variance. *Stochastic Processes and their Applications*(19), 281-296.

Buckle, D. J. (1995, June). Bayesian Inference for Stable Distributions. *Journal of the American Statistical Association*(90), 605-613.

Calzolari, G., & Lombardi, M. (2008). Indirect Estimation of Alpha-Stable Distributions and Processes. *Econometrics Journal, 11*, 193-208.

Chambers, J. M., Mallows, C. L., & Stuck, B. W. (1976, June). A Method for Simulating Stable Random Variables. *Journal of the American Statistical Association*(71), 340-344.

Chenyao, D., Doganoglu, T., & Mittnik, S. (1999). Computing the Probability Density Function of the Stable Paretian Distribution. *Mathematical and Computer Modelling*(29), 235-240.

Chenyao, D., Doganoglu, T., Mittnik, S., & Rachev, S. T. (1999). Maximum Likelihood Estimation of Stable Paretian Models. *Mathematical and Computer Modelling*(29), 275-293.

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

DuMouchel, W. H. (1973, June). Stable Distributions in Statistical Inference. *Journal of the American Statistical Association*(68), 469-477.

Fama, E. F. (1965, January). The Behavior of Stock-Market Prices. *Journal of Business*(38), 34-105.

Fama, E. F., & Roll, R. (1968, September). Some Properties of Symmetric Stable Distributions. *Journal of the American Statistical Association*(63), 817-836.

Fama, E. F., & Roll, R. (1971, June). Parameter Estimates for Symmetric Stable Distributions. *Journal of the American Statistical Association*(66), 331-338.

Feller, W. (1971). *An Introduction to Probability Theory and Its Applications* (2nd ed., Vol. II). New York: John Wiley & Sons, Inc.

Feller, W. (1971). *An Introduction to Probability Theory and Its Applications* (2nd ed., Vol. II). New York: John Wiley & Sons, Inc.

Fielitz, B. D. (1976). Further Results on Asymmetric Distributions of Stock Prices. *The Journal of Financial and Quantitative Analysis, 11*, 39-55.

Fielitz, B. D., & Smith, E. W. (1972). Asymmetric Distributions of Stock Price Changes. *Journal of the American Statistical Association, 67*, 813-814.

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

Godsill, S. J. (1999). MCMC and EM-based Methods for Inference in Heavy Tailed Processes with Stable Innovations. *Proceedings of the IEEE Signal Processing Workshop on Higher-Order Statstics*.

Godsill, S. J. (2000). Inference in Symmetric Alpha-Stable Noise Using MCMC and the Slice Sampler. *Proceedings of the IEEE on Signal Processing*.

Godsill, S. J., & Kuruoglu, E. E. (1999, May). Bayesian Inference for Time Series with Heavy-Tailed Symmetric alpha-Stable Noise Process. *Process Applications of Heavy-Tailed Distributions in Economics, Engineering, and Statistics*.

Godsill, S. J., & Lombardi, M. J. (2004). On-line Bayesian Estimation of AR Signals in Symmetric Alpha-Stable Noise. *Technical Report*.

Granger, C. W., & Orr, D. (1972). Infinite Variance and Research Strategies in Time Series Analysis. *Journal of the American Statistical Association, 67*, 275-285.

Horvatić, D., Podobnik, B., Stanley, H. E., & Valentinčič, A. (2011). Asymmetric Lévy Flight in Financial Ratios. *Proceedings of the National Academy of Sciences of the United States of America, 108*, 17883-17888.

Knight, J. L., & Yu, J. (2002). Empirical Characteristic Function in Time Series. *Econometric Theory, 18*, 691-721.

Koutrouvelis, I. (1980). Regression-Type Estimation of the Parameters of Stable Laws. *Journal of the American Statistical Association, 75*, 918-928.

Kuruoglu, E. E., & Zerubia, J. (2004). Modeling SAR Images with a Generalization of the Rayleigh Distribution. *IEEE Transactions on image processing, 13*, 527-533.

Mandelbrot, B. (1961). Stable Paretian Random Functions and the Multiplicative Variation of Income. *Econometrica, 29*, 517-543.

Mandelbrot, B. (1963, October). The Variation of Certain Speculative Prices. *Journal of Business*(36), 394-419.

Mandelbrot, B., & Taylor, H. M. (1967). On the Distribution of Stock Price Differences. *Operations Research, 15*(6), 1057-1062.

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

Menn, C., & Rachev, S. T. (2006). Calibrated FFT-based Density Approximations for Stable Distributions. *Computational Statistics & Data Analysis, 50*, 1891-1904.

Nolan, J. P. (1998). Parameterizations and Modes of Stable Distributions. *Statistics and Probability Letters*(38), 187-195.

Nolan, J. P. (1999). An algorithm for evaluating stable densities in Zolotarev's (M) parametrization. *Mathematical and Computer Modelling, 29*, 229-233.

Officer, R. R. (1972). The Distribution of Stock Returns. *Journal of the American Statistical Association, 67*, 807-812.

Press, S. J. (1972). Estimation in Univariate and Multivariate Stable Distributions. *Journal of the American Statistical Association*(67), 842-846.

Qiou, Z., & Ravishanker, N. (1998, January). Bayesian Inference for Time Series with Stable Innovations. *Journal of Time Series Analysis*(19), 235-250.

Qiou, Z., & Ravishanker, N. (1998). Bayesian Inference for Vector Arma Models with Stable Innovations. *Sankhyā: The Indian Journal of Statistics, 60*, 459-475.

Rosenfeld, G. (1976). Identification of Time Series with Infinite Variance. *Journal of the Rotyal Statistical Society, Series C (Applied Statistics)*(25), 147-153.

Samorodnitsky, G., & Taqqu, M. S. (1994). *Stable Non-Gaussian Random Processes: Stochastic Models with Infinite Variance.* New York: Chapman & Hall.

Tsionas, E. G. (1999). Monte Carlo Inference in Econometric Models with Symmetric Stable Disturbances. *The Journal of Econometrics*(88), 365-401.

Weron, R. (1996). On the Chambers, Mallow, Stuck Method for Simulating Skewed Stable Random Variables. *Technical Report for the Hugo Steinhaus Center for Stochastic Methods*.

Yamazato, M. (1978). Unimodality of Infinitely Divisible Distribution Functions of Class. *The Annals of Probability, 6*, 523-531.

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