DO FINANCIAL RETURNS HAVE FINITE OR INFINITE VARIANCE? A PARADOX AND AN EXPLANATION

MICHAEL GRABCHAK AND GENNADY SAMORODNITSKY

Abstract. One of the major points of contention in studying and modeling financial returns is whether or not the variance of the returns is finite or infinite (sometimes referred to as the Bachelier-Samuelson Gaussian world versus the Mandelbrot stable world). A different formulation of the question asks how heavy the tails of the financial returns are. The available empirical evidence can be, and has been, interpreted in more than one way. The apparent paradox, which has puzzled many a researcher, is that the tails appear to become less heavy for less frequent (e.g. monthly) returns than for more frequent (e.g. daily) returns, a phenomenon not easily explainable by the standard models. Inspired by the prelimit theorems of Klebanov et al. (1999) and Klebanov et al. (2000) we provide an explanation to this paradox. We show that, for financial returns, a natural family of models are those with tempered heavy tails. These models can generate observations that appear heavy tailed for a wide range of aggregation levels before becoming clearly light tailed at even larger aggregation scales. Important examples demonstrate the existence of a natural scale associated with the model at which such an apparent shift in the tails occurs.

1. Introduction

Do financial returns have a finite variance or not? This \$10,000 question has been made much more valuable by the ongoing – at the time of this writing – worldwide stock market, housing market, and general financial crisis. The question can be traced back to the pioneering paper of Mandelbrot (1963), which suggested that infinite variance stable models provide a better fit for certain financial returns than the more classic Bachelier-Samuelson Gaussian models. These models where further used to model stock returns in the influential paper of Fama (1965). The answer to this question is not of

Key words and phrases. financial returns, finite or infinite variance, heavy tails, Bachelier-Samuelson model, Mandelbrot model.

Grabchak's research was supported in part by the NSF grant EMSW21-RTG at Cornell University. Samorodnitsky's research was partially supported by the ARO grant W911NF-07-1-0078 at Cornell University.

purely academic interest; it crucially affects risk calculation, portfolio selection, and option pricing. The question can be formulated in several related, but not equivalent, ways.

(1) Do financial returns have a finite variance or not?
(2) Do financial returns follow a Gaussian law or an infinite variance stable law?
(3) In what range is the tail index of financial returns?

This last formulation refers to the parameter of a certain class of semi-parametric models, which are often used to model returns. This is the class of distributions with regularly varying tails. A random variable X (here representing a return) is said to have a regularly varying *right* tail with tail index $\alpha > 0$ if

$$P(X > x) = x^{-\alpha} L(x), \ x > 0 \,, \tag{1.1}$$

where L is a slowly varying at infinity function (see Embrechts et al. (1997) for information on regular variation). A similar definition applies to the regular variation of the left tail, which may have a different tail index. The smaller of the two tail indices (we will often refer to it simply as the tail index) controls which finite moments the random variable has. If the smaller tail index is greater than 2, then the random variable has a finite variance, when it is between 1 and 2, the random variable has a finite mean but an infinite variance, and when it is less than 1, the random variable does not have a finite mean.

Sometimes a more specific assumption of a power tail is used. Here one assumes that $P(X > x) \sim cx^{-\alpha}$ as $x \to \infty$ for some $c > 0$, and similarly with the left tail. Many well-known probability models have power tails. For example, an α-stable random variable with $0 < \alpha < 2$ (its tail index is equal to the index of stability α, see Samorodnitsky and Taqqu (1994)) and a Student t-random variable (its tail index is equal to the number of its degrees of freedom).

The α-stable distributions form a relatively narrow family of models, and it is accepted in the literature that for most financial returns an exactly stable model does not provide a good statistical fit (and an exactly Gaussian model does not provide a good fit either). The assumption of regular variation of the tails is, on the other hand, not very controversial (even though certain option pricing techniques require tails lighter than those given by (1.1)). What is controversial is the range of the tail index of financial returns.

Approximate stability of the returns and infinite variance indicate a tail index of less than 2. Infinite variance models are advocated in Mittnik and Rachev (2000), but other authors, going back to Blattberg and Gonedes (1974), or Lau et al. (1990) report statistical evidence of finite variance in financial returns.

The most confusing issue of all, and this is where the apparent paradox arises, is that the empirically measured tail index appears to be lower for more frequent returns and higher for less frequent returns; rich evidence is in Gencay et al. (2001). In particular, one could find, for example, that daily or more frequent returns had infinite variance, but weekly or less frequent returns had finite variance. Not everyone agrees that this is really important in practice (see e.g. Taleb (2009)), but the paradoxical nature of this phenomenon was quickly realized. Thus, Akgirav and Booth (1988), upon reporting that the estimated index of stability increases from daily to weekly to monthly returns, point out that this is inconsistent with a stable model of the returns. Even more generally, it is difficult to explain increasing tail indices for returns with power tails, or regular varying tails. This cannot occur if the returns are independent and identically distributed, and most known models of dependent returns rule out this phenomenon as well (see e.g. Mikosch and Samorodnitsky (2000)).

In this work we propose a resolution of "the tail paradox" by showing that the empirical findings of the increasing tail index as the aggregation of the returns increases is consistent with models possessing what we call "tempered heavy tails". A random variable with tempered heavy tails does not, strictly speaking, have power or regular varying tails, because of the tempering. Nevertheless, this random variable can be similar, in important respects, to a random variable with power tails, or even to a stable random variable.

Recall that a random variable with regularly varying (and balanced) tails of tail index $0 < \alpha < 2$ is in the domain of attraction of an α-stable distribution. This means that the distribution of properly shifted and normalized sums of independent and identically distributed copies of this random variable converge in distribution, as the number of observations increases, to an α-stable distribution. Informally, the sums $X_1 + \ldots + X_n$ of independent observations distributed as X will, while having the same tail index as the original X, look more and more like an α-stable random variable as n increases. Similarly, if X has a tail index $\alpha > 2$ or, more generally, has a finite variance, then $X_1 + \ldots + X_n$ will, as n increases, look more and more like a Gaussian random variable.

A random variable with tempered heavy tails has a finite variance. However, empirically, its tail index may appear to be less than 2. Moreover, the distribution of the sum $X_1 + \ldots + X_n$ of independent observations distributed as X may be well approximated by an infinite variance stable distribution for a wide range of values of n, before (necessarily) converging (after a proper shift and scaling) to a Gaussian distribution as n becomes very large. Furthermore, it is possible that in the intermediate range of n, the distribution of the sum $X_1 + \ldots + X_n$ becomes "more stable-like" as n increases. If (relatively) high frequency returns have tempered heavy tails, then, empirically, these returns and certain lower frequency (aggregated) returns will appear to have a low tail index and may, even, appear to have an approximately stable distribution, while even lower frequency returns will look more and more Gaussian-like and, hence, empirically, the tail index! will appear to be increasing. The resulting picture is in excellent correspondence with the discussed above paradoxically increasing with aggregation tail index of actual financial returns.

Financial markets have built-in mechanisms designed to limit the fluctuations of the prices. This means that one would not expect to see the returns exhibit regular variation in their entire unlimited range. Models with tempered heavy tails may, appropriately, exhibit power-like tail behavior in a large part, but not the whole, of their range.

In summary, we suggest that using models with tempered heavy tails is an attractive option that accounts for the otherwise difficult to explain phenomenon of the tail index increasing with the aggregation of the returns and, more generally, reconciles the otherwise irreconcilable views for and against infinite variance models of financial returns. In fact, this class of models explains why, at certain return frequencies (but not at other return frequencies) the empirical distribution of financial returns may look, approximately, stable.

It is important to mention, at this point, that random variables with tempered heavy tails retain many of the properties of random variables with the "usual" heavy tails. In particular, risk calculations with tempered heavy tails will have much in common with risk calculations with the "usual" heavy tails, but they will not be identical to the latter calculations. We will address this issue in a future work.

This paper is organized as follows. In Section 2 we introduce the general idea of tempered heavy tails and concentrate on two specific models: the truncated Pareto distribution and the smoothly truncated Lévy flights. We present a small empirical study of the behavior of the sums $X_1 + \ldots + X_n$ of independent identically distributed random variables generated from these models. In Section 3 we introduce several distances between probability distributions and state a theorem showing that, under certain conditions, sums of many independent identically distributed random variables may have an approximately α-stable distribution even though the random variables are not in the domain of attraction of an α-stable distribution. The proof of the main result is postponed until Section 5. Our presentation here is heavily influenced by the work of Klebanov et al. (1999, 2000). We present a clear and self-contained proof (this appears to be somewhat lacking in the above references); furthermore, we extend the main result to the important multivariate case. In Section 4 we show that the distributions of the two families of models with tempered heavy tails that we are considering in this paper have a natural scale associated with them. It determines when sums $X_1 + \ldots + X_n$ of independent identically distributed random variables from these distributions look “stable-like”, and when “Gaussian-like” behavior sets in.

2. Distributions with tempered heavy tails

We say that the distribution of a random variable X has *tempered heavy tails* if it can be obtained by modifying a random variable with power, or regularly varying, tails, via tempering the tails of the latter. This tempering restricts the range where the power, or regularly varying, tails, apply. Depending on the kind and the extent of the tempering of the tails, sums of the type $X_1 + \ldots + X_n$ of independent identically distributed random variables with such tails can have, for a large number of terms n, a distribution that is very close to an infinite variance α-stable distribution even though the random variable itself is not in the domain of attraction of an α-stable distribution. Random variables with tempered heavy tails have a finite variance and, hence, must be in the domain of attraction of a Gaussian distribution.

The tails of a random variable can be tempered in different ways. In this paper we consider two ways of tempering heavy tails, leading to two different families of distributions with tempered heavy tails.

The first example is based on the idea of *tail truncation*. Let Z be a random variable, which we assume to be in the domain of normal attraction of some α-stable distribution, $0 < \alpha < 2$. That is, the normalized partial sums $S'_n = n^{-1/\alpha} \sum_{i=1}^n Z_i$ of independent copies of Z converge in distribution, as $n \to \infty$, to a non-degenerate α-stable random variable. This means, that $|Z|$ has a power tail with exponent α, and for some $0 \le p \le 1$,

$$\lim_{x\to\infty} \frac{P\big(Z > x\big)}{P\big(|Z| > x\big)} = p \, ,$$

see e.g. Feller (1971). For a large $T > 0$, consider a random variable X obtained by “rejecting” the values of Z whose magnitude is larger than T. Formally,the distribution of X is the conditional distribution of Z given that $|Z| \le T$. That is, for any measurable set $A \subset [-T, T]$, $P(X \in A) = P(Z \in A)/P(|Z| \le T)$. We view X as Z with truncated tails. Clearly, X has a finite variance and, hence, is in the domain of attraction of a Gaussian distribution. However, compare the sums $S'_n = n^{-1/\alpha} \sum_{i=1}^n Z_i$ and $S_n =$

$n^{-1/\alpha}\sum_{i=1}^{n} X_i$, where $X_1, \ldots, X_n$ are independent copies of X, obtained from the sequence $Z_1, Z_2, \ldots,$ by retaining the first n entries (returns) in that sequence of magnitude not exceeding T. It is clear that, if T is "very large" and n is "small in comparison with T", then, very likely, all of the X_i will coincide with the corresponding Z_i, and so the normalized sums S_n and S'_n will have have very similar distributions. This suggests that if, for such values of n, the distribution of S'_n is well approximated by an infinite variance α-stable law, then this law should be a good approximation for the distribution of S_n for such values of n and T as well.

Later in this paper, we will see that the truncation level T provides, in a formal sense, a natural time scale for the number of terms n for which the above approximation of the distribution of S_n by an α-stable law is valid.

A simple, but illustrative, example is that of *a symmetric Pareto distribution.* Let $0 < \alpha < 2$. For $b > 0$ consider a distribution with a symmetric density given by

$$f(x) = \frac{\alpha b^\alpha}{2} |x|^{-1-\alpha} 1_{|x|>b} \,. \tag{2.1}$$

This distribution is in the domain of normal attraction of a symmetric α-stable law. For a small numerical study, we choose $\alpha = 1.5$, $b = .5$, and we truncate this Pareto distribution at the value $T = 70$. Figure 1 shows plots of the estimated densities of S_n for several values of n with the overlayed density of a symmetric α-stable random variable; the scale of the latter that provided the best fit (and which we used in the plots) was somewhat smaller than the scale of the limiting distribution to which the normalized sum of non-truncated Pareto random variables converges. The density of S_n was calculated approximately, using kernel density estimators based on a sample of size $100,000$.

In the figure, we see that by the time n reaches 50 the approximation by the stable law appears to be quite good. By the time the sample size reaches 150, we start seeing divergence both in the center and in the tails, and once the sample size n is at 300, the approximation by the stable law is visibly bad.

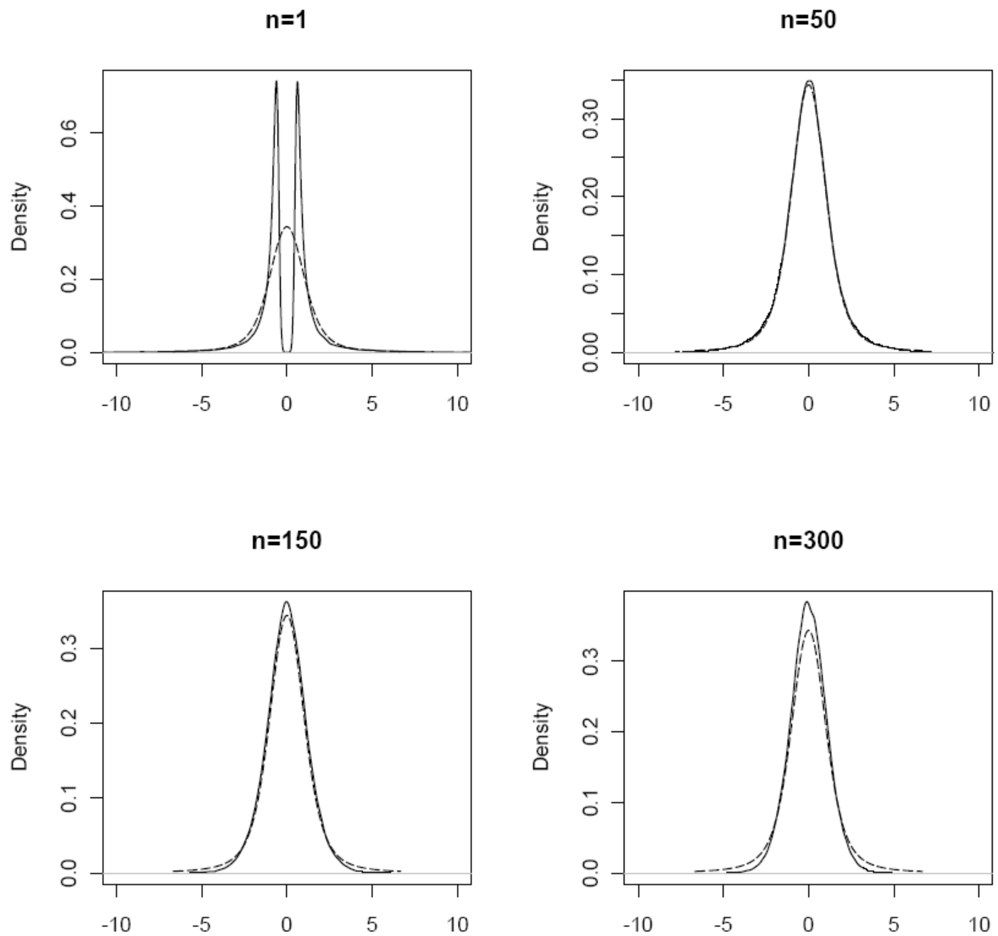

FIGURE 1. Plots of the densities of sums of truncated symmetric Pareto random variables (solid lines) with the approximating stable densities (dashed lines) overlayed.

On the other hand, suppose we use the normalization dictated by the finite variance of truncated Pareto random variables, i.e. we consider the distribution of $n^{-1/2}\sum_{i=1}^{n} X_i$. Figure 2 gives plots of the approximated finite sample densities with the limiting normal distribution overlayed. We see that the normal approximation is quite unsatisfactory at $n = 100$, is improving when the sample size reaches 500, and appears to be very good at the sample size $n = 1,000$.

The tails of a random variable can be tempered in "more delicate" ways than tail truncation. For example, the so-called *tempered stable random*

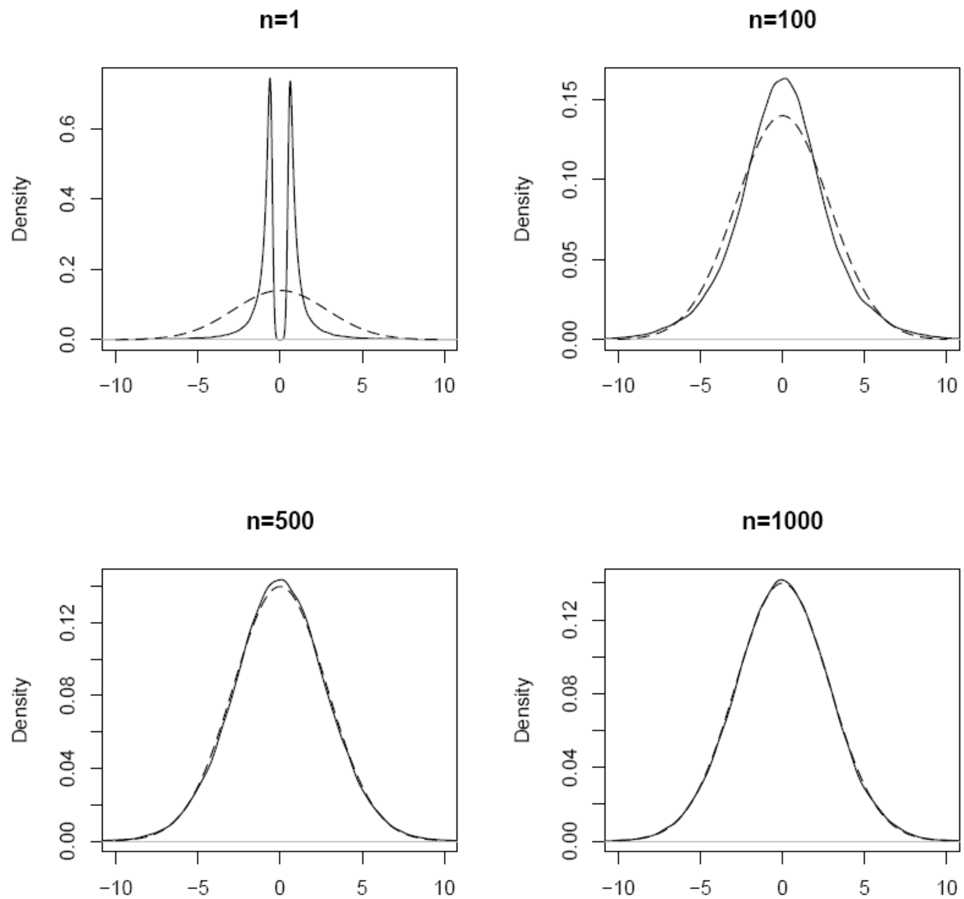

FIGURE 2. Plots of the densities of sums of truncated symmetric Pareto random variables (solid lines) with the approximating normal densities (dashed lines) overlayed.

variables have infinitely divisible distributions related to the infinite variance α-stable distributions, where the tempering is done at the level of Lévy measures; see Rosiński (2007). For our second example of distributions with tempered heavy tails we choose the *smoothly truncated Lévy flights* (STLFs), a subclass of the tempered stable distributions; see Koponen (1995). These models have already been successfully used in financial applications, see, for example, Carr et al. (2002, 2003); Cont and Tankov (2004). The usefulness of general tempered stable models comes from the fact that they are so similar to the stable models that a Lévy process with tempered stable marginal distributions behaves like a stable motion at small time scales. On the other

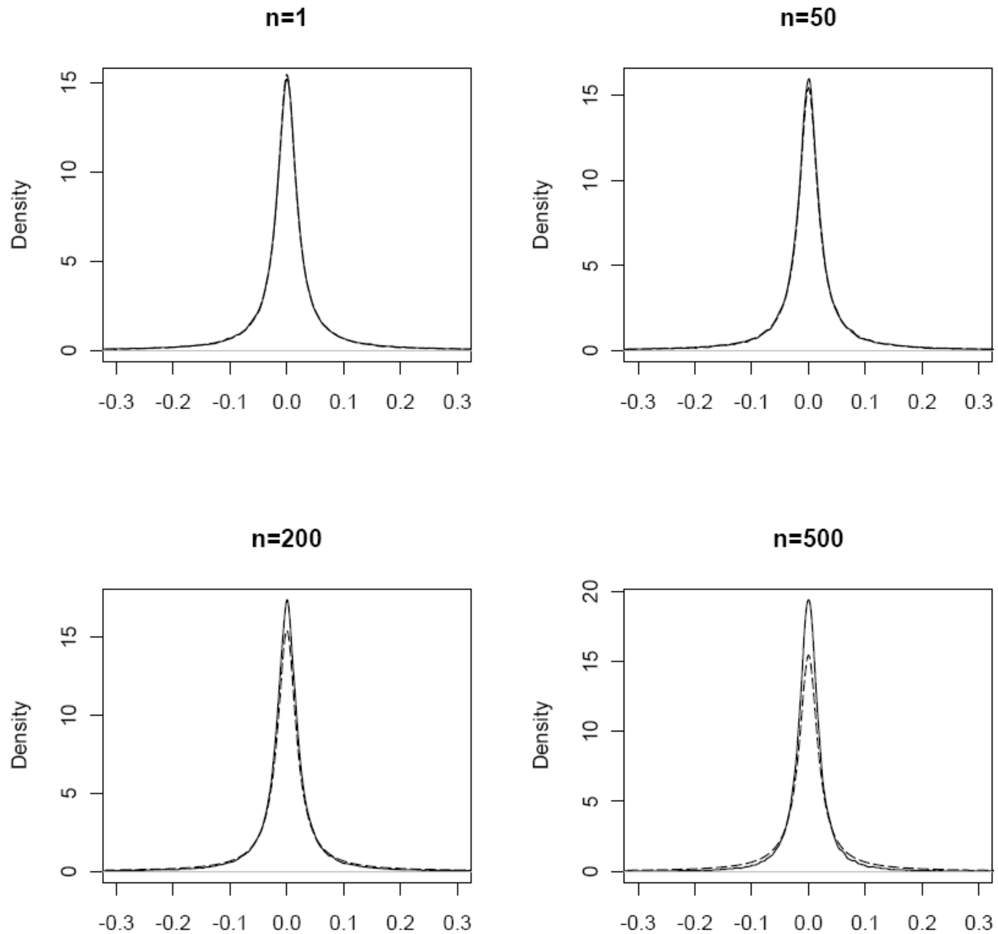

FIGURE 3. Plots of the densities of scaled sums of iid STLFs (solid lines) with the approximating stable density (dashed lines) overlayed.

hand, for certain values of the parameters, a tempered stable random variable has a finite variance; see Rosiński (2007). This is the case for the smoothly truncated Lévy flights. Despite this, we will show that, a sum of smoothly truncated Lévy flights can be well approximated by an infinite variance stable distribution even when the sum has many terms.

For an illustrative example we consider symmetric smoothly truncated Lévy flights that are tempered α-stable distributions with $\alpha \in (0, 1)$. Beside the exponent α, such distributions are characterized by a scale $\sigma > 0$ and the level of tempering $\ell > 0$. The characteristic function of such a smoothly

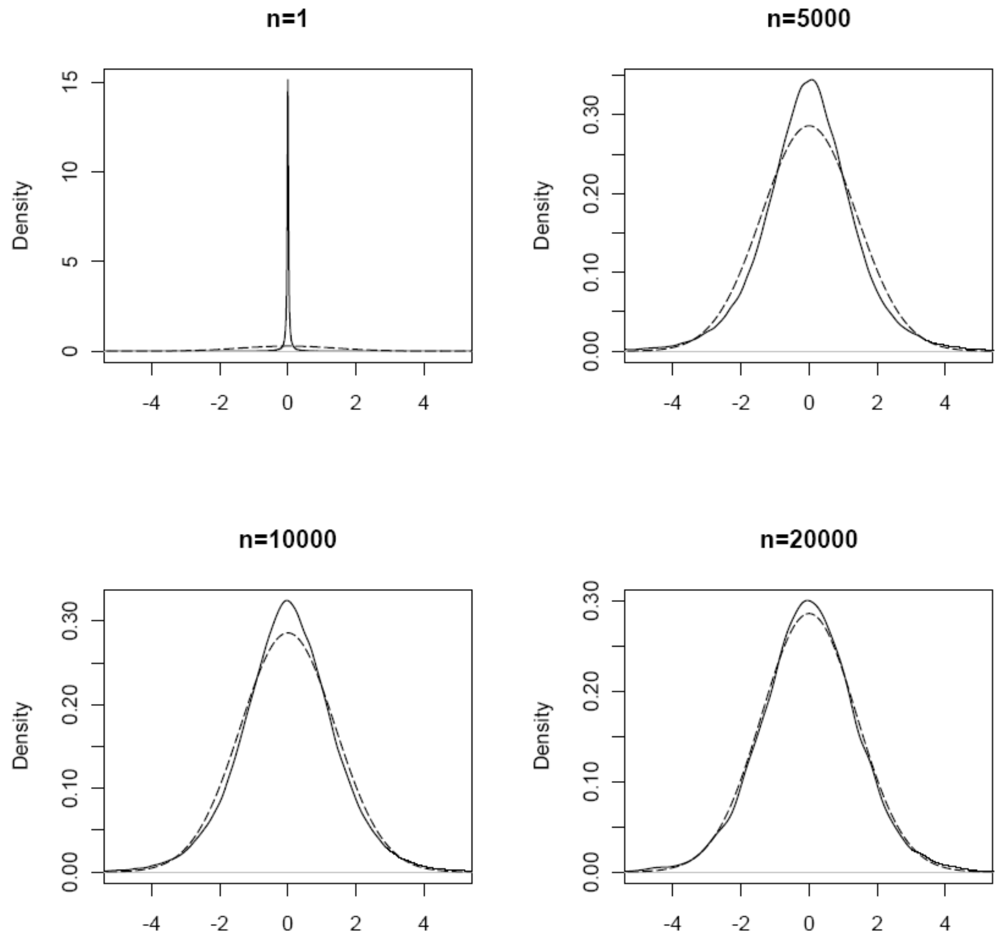

FIGURE 4. Plots of the densities of scaled sums of STLFs (solid lines) with the approximating normal densities (dashed lines) overlayed.

truncated Lévy flight is given by

$$\varphi(\lambda) = \exp\left\{-\sigma^\alpha \ell^{-\alpha}\left[\left(1+\lambda^2\ell^2\right)^{\alpha/2}\cos\left(\alpha\arctan(\lambda\ell)\right)-1\right]\right\}, \tag{2.2}$$

$\lambda \in \mathbb{R}$. In Section 4.1 we will see that the level of tempering ℓ provides, in this model, a natural time scale for the number of terms n for which the above approximation of the distribution of S_n by an α-stable law is valid.

For a small numerical study, we choose $\alpha = .95$, $\sigma^\alpha = .1628$, and $\ell = 100$. Figure 3 shows plots of the estimated densities of scaled sums of iid STLFs with the density of the corresponding α-stable distribution overlayed. When $n = 1$ they appear to be almost identical, and they are still very similar when $n = 50$. By $n = 500$, however, the densities are quite different in both the peaks and the tails. Ultimately the distribution of the normalized sum of iid smoothly truncated Lévy flights converges to a normal limit, but the convergence is very slow. This slow convergence is illustrated in Figure 4.

It is interesting and important to note that in both of the numerical studies presented above a very large sample size is needed to get normal-like behavior of the sum. This is at odds with the common rule of thumb that convergence in the Gaussian Central Limit Theorem is practically attained at $n = 30$.

We conclude this section by noting that the above discussion of tempered heavy tails fully applies to random vectors in $\mathbb{R}^d$ as well. In fact, the main quantitative result of the next section will be stated and proved in the multivariate case.

3. The Main Result

In this section we state a theorem, which shows that, under appropriate conditions, the distribution of the sum of many independent identically distributed random variables with tempered heavy tails can be well approximated by an infinite variance α-stable distribution even though these random variables are not in the domain of attraction of the α-stable distribution. The presentation in this section is inspired by the work of Klebanov et al. (1999) and Klebanov et al. (2000). Our main approximation theorem (Theorem 1 below) applies to random vectors (i.e. to entire portfolios of returns).

We begin by setting up the notation. Let X be a d-dimensional random vector. We will denote its characteristic function by $\hat{\mu}_X$ and its probability law and distribution function by F_X. If X has a density with respect to the d-dimensional Lebesgue measure, we will denote it by f_X.

The convolution of two measurable functions f and g is defined by

$$f * g(x) = \int_{\mathbb{R}^d} g(x-y) f(y) dy$$

at any point x where the integral exists. If F is a (signed) measure and g is a measurable function, then the convolution of F and g is defined by

$$F \star g(x) = \int_{\mathbb{R}^d} g(x-y) F(dy)$$

at every x for which the integral exists. If we choose $g = H$, a cdf, then $F \star H$ can also be viewed as a convolution of two measures. Clearly, if F has a density f with respect to the d-dimensional Lebesgue measure, then $F \star g = f * g$.

A function f on $\mathbb{R}^d$ is said to satisfy the Lipschitz condition with coefficient M if

$$|f(x) - f(y)| \leq M|x - y|$$

for every $x, y \in \mathbb{R}^d$. We will use the notation $f \in Lip_M$. Note that if h is differentiable and if $M := \sup_{x \in \mathbb{R}^d} |\nabla h(x)| < \infty$, then $h \in Lip_M$.

A useful distance on the space of probability laws on $\mathbb{R}^d$ can be defined as follows. Let $c, \gamma \geq 0$. For d-dimensional random vectors X and Y we set

$$d_{c,\gamma}(X, Y) \Big(= d_{c,\gamma}(F_X, F_Y) \Big) = \sup_{|z| \geq c} \frac{|\hat{\mu}_X(z) - \hat{\mu}_Y(z)|}{|z|^\gamma}. \tag{3.1}$$

Note that this measures the distance between two probability laws and not two random variables. Thus, the notation $d_{c,\gamma}(F_X, F_Y)$ is more precise than $d_{c,\gamma}(X, Y)$. However, we will use, as is common, the latter notation. It is well known, and easy to check, that if Y is a strictly α-stable random vector (see Samorodnitsky and Taqqu (1994)), and $X_1, X_2, \ldots$ are iid copies of a random vector X, such that $d_{0,\gamma}(X, Y) < \infty$ for some $\gamma > \alpha$ then X is in the domain of normal attraction of Y. See Klebanov et al. (1999).

Let h be a probability density on $\mathbb{R}^d$. We define another distance on the space of probability laws on $\mathbb{R}^d$ by setting, for two d-dimensional random vectors X and Y,

$$K_h(X, Y) \Big(= K_h(F_X, F_Y) \Big) = \sup_{x \in \mathbb{R}^d} |F_X \star h(x) - F_Y \star h(x)|. \tag{3.2}$$

It is easy to check that if h satisfies a Lipschitz condition and the corresponding characteristic function does not vanish, then K_h metrizes weak convergence on $\mathbb{R}^d$.

We can now state our main theorem.

Theorem 1. *Fix $\alpha \in (0, 2]$. Let h be a probability density on $\mathbb{R}^d$. Assume that $h \in Lip_{M_h}$ for some $M_h > 0$. Let $X_1, X_2, \ldots$ be iid d-dimensional random vectors, and let $S_n = n^{-1/\alpha} \sum_{j=1}^n X_j$. Let Y be a strictly α-stable* *d-dimensional random vector. For any $\gamma > \alpha$, we have*

$$K_h(F_{S_n}, F_Y) \quad \leq \quad \inf_{a,\Delta>0} \left\{ \frac{d_{\Delta n^{-1/\alpha},\gamma}(X,Y)}{n^{\gamma/\alpha-1}} \frac{2^{\gamma+1}(a\sqrt{d})^{\gamma+d}}{\pi^{d/2}\Gamma(d/2)(\gamma+d)} + \frac{2}{\pi^d}[\Delta \wedge (2a)]^d + M_h \frac{12d}{\pi a} \right\}.$$

Remark 2. Clearly, it is possible to optimize the upper bound in Theorem 1 over $a > 0$ for a fixed $\Delta > 0$. The resulting bound is difficult to interpret, and we do not present it here. Nonetheless, such bounds are useful in numerical work.

Remark 3. We can think of the quantitative bounds presented in Theorem 1 in the following way. Suppose that the random vector X is such that, for some $\gamma > \alpha$, the distance $d_{c,\gamma}(X,Y)$ remains "not too big" even for certain reasonably small values of $c > 0$. In that case one can choose $a > 0$ large, $\Delta > 0$ small, and have the upper bound on the distance between the distribution of S_n and that of the α-stable random vector small for fairly large values of n.

Remark 4. How does one interpret the distance between the smoothed densities in Theorem 1? The easiest way to interpret this distance is that, in practice, one always performs smoothing while estimating the density by using kernel density estimation. Therefore, the theorem simply gives an upper bound on the distance between such smoothed densities. Technically, the smoothing operation puts an absolute upper bound on the Lipschitz coefficient of the densities being compared. If one performs smoothing with the density of a random variable that is nearly concentrated at zero, the smoothed density will be very similar to the non-smoothed density. As an illustration, we present, in Figure 5 the densities of the symmetric Pareto distribution of Section 2 above smoothed with various Gaussian densities. The case $\sigma = 0$ corresponds to absence of smoothing.

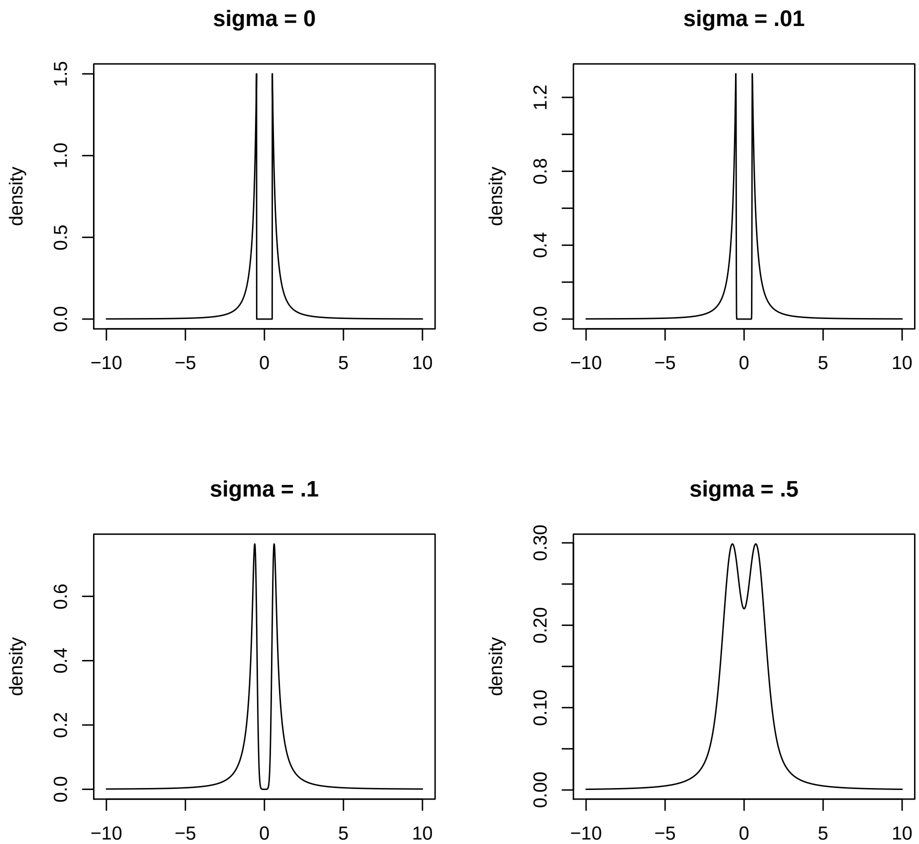

FIGURE 5. Plots of a symmetric Pareto density smoothed with a centered Gaussian density with varying standard deviations.

4. NATURAL SCALE OF CERTAIN DISTRIBUTIONS WITH TEMPERED HEAVY TAILS

Let $X_1, X_2, \ldots$ be a random sample from some tempered heavy tailed distribution, and consider sums of the form $X_1 + \ldots + X_n$. In this section we demonstrate that certain families of distributions with tempered heavy tails have a natural scale, which determines for what number n, such sums can be well approximated by an infinite variance α-stable distribution. In this section we will only consider the one-dimensional case, specifically the two examples discussed in Section 2 above: the symmetric Pareto distributions with truncated tails and smoothly truncated Lévy flights. We begin with the latter.

4.1. **Natural scale for the smoothly truncated Lévy flights.** Consider a smoothly truncated Lévy flight with $0 < \alpha < 1$ and the characteristic function given by (2.2). For this discussion we will denote this characteristic function by φ_ℓ, to emphasize the dependence on the tempering level ℓ. It is elementary that, as the tempering level $\ell \to \infty$, for any $\lambda \in \mathbb{R}$,

$$\varphi_\ell(\lambda) \to \varphi_\infty(\lambda) = \exp\left\{-\sigma^\alpha \cos\left(\frac{\pi\alpha}{2}\right)|\lambda|^\alpha\right\}, \tag{4.1}$$

the characteristic function of a symmetric α-stable distribution. Note that for every $\lambda > 0$ (say),

$$\big|\varphi_\infty(\lambda) - \varphi_\ell(\lambda)\big| \le \min\big(1, \sigma^\alpha \ell^{-\alpha} g(\lambda\ell)\big),$$

where for $x > 0$,

$$g(x) = \left|(1+x^2)^{\alpha/2}\cos(\alpha \arctan x) - 1 - x^\alpha \cos\left(\frac{\pi\alpha}{2}\right)\right|. \tag{4.2}$$

Clearly, g is continuous on $(0,\infty)$, $g(x) \to 1$ as $x \to \infty$, and $g(x) \sim x^\alpha \cos(\pi\alpha/2)$ as $x \to 0$ (recall that $0 < \alpha < 1$). Therefore, there is a finite positive constant A_α such that $g(x) \le A_\alpha \min\big(1, x^\alpha\big)$ for all $x > 0$.

Let X be a random variable whose distribution is a smoothly truncated Lévy flight with the characteristic function φ_ℓ, and Y a symmetric α-stable random variable with the characteristic function φ_∞. We conclude that for $c > 0$ and $\gamma > \alpha$, the distance (3.1) satisfies

$$d_{c,\gamma}(X,Y) \le A_\alpha \sigma^\alpha \ell^{-\alpha} c^{-\gamma} \min\big(1, (c\ell)^\alpha\big). \tag{4.3}$$

Write (a weaker version of) the bound given in Theorem 1 in the form

$$K_h(F_{S_n}, F_Y) \le \inf_{a,\Delta>0}\left\{C_1(\gamma) a^{\gamma+1} \frac{d_{\Delta n^{-1/\alpha},\gamma}(X,Y)}{n^{\gamma/\alpha - 1}} + C_2\Delta + C_3 M_h a^{-1}\right\}$$

$$:= \inf_{a,\Delta>0} B_\gamma(a,\Delta), \tag{4.4}$$

with

$$C_1(\gamma) = \frac{2^{\gamma+1}}{\pi(\gamma+1)}, \quad C_2 = \frac{2}{\pi}, \quad C_3 = \frac{12}{\pi}.$$

We can use (4.3) to obtain

$$B_\gamma(a,\Delta) \le C_1(\gamma) A_\alpha a^{\gamma+1} \Delta^{-\gamma} n \ell^{-\alpha} \sigma^\alpha \min\Big(1, \big(\Delta n^{-1/\alpha}\ell\big)^\alpha\Big) + C_2\Delta + C_3 M_h a^{-1}.$$

Consider now the range $n \leq \ell^\alpha$. Selecting

$$\tilde{\Delta} = \big(n\ell^{-\alpha}\big)^{1/(\gamma+1)}\tilde{a}, \quad \tilde{a} = \left(\frac{C_3 M_h}{\big(C_1(\gamma)A_\alpha\sigma^\alpha + C_2\big)\big(n\ell^{-\alpha}\big)^{1/(\gamma+1)}}\right)^{1/2}$$

we obtain an upper bound on the distance between the density of the smoothed distribution of S_n, and that of the smoothed distribution of Y, given by

$$K_h(F_{S_n}, F_Y) \leq 2\big(C_3 M_h\big)^{1/2}\big(C_1(\gamma)A_\alpha\sigma^\alpha + C_2\big)^{1/2}\big(n\ell^{-\alpha}\big)^{1/(2(\gamma+1))}. \tag{4.5}$$

This bound (4.5) shows that, if the level of tempering ℓ is large and the number of terms n in the sum $X_1 + \ldots + X_n$ is such that $n\ell^{-\alpha}$ is small, then the distance $K_h(F_{S_n}, F_Y)$ will be small.

Therefore, for smoothly truncated Lévy flights with $0 < \alpha < 1$ and the characteristic function given by (2.2), ℓ^α provides the natural scale: if the number n is much less than this natural scale, then the distribution of S_n is well approximated by the distribution of the infinite variance symmetric α-stable random variable with the characteristic function given by (4.2).

4.2. **Natural scale for the symmetric Pareto distributions with truncated tails.** Let Z come from the symmetric Pareto distribution given in (2.1). We choose $b = 1$ for simplicity of notation. The characteristic function of the distribution attained by truncating this distribution at $T > 0$ is given by

$$\varphi_T(\lambda) = \frac{\alpha}{1 - T^{-\alpha}} \int_1^T x^{-(1+\alpha)} \cos \lambda x \, dx, \quad \lambda \in \mathbb{R}. \tag{4.6}$$

Note that independent and identically distributed random variables from the non-truncated symmetric Pareto distribution ($T = \infty$) satisfy

$$n^{-1/\alpha} \sum_{j=1}^{n} Z_j \Rightarrow Y,$$

where Y is a symmetric α-stable random variable with characteristic function

$$\psi(\lambda) = e^{-C_\alpha |\lambda|^\alpha}, \quad \lambda \in \mathbb{R} \tag{4.7}$$

with

$$C_\alpha = \alpha \int_0^\infty \frac{1 - \cos y}{y^{1+\alpha}} \, dy;$$

see Feller (1971). For every $\lambda > 0$ (say),

$$\big|\psi(\lambda)-\varphi_T(\lambda)\big| \le \left|\frac{\alpha}{1-T^{-\alpha}}\int_1^T x^{-(1+\alpha)}\cos\lambda x\,dx - \alpha\int_1^\infty x^{-(1+\alpha)}\cos\lambda x\,dx\right|$$
$$+\left|\alpha\int_1^\infty x^{-(1+\alpha)}\cos\lambda x\,dx - e^{-C_\alpha\lambda^\alpha}\right| := R_1(\lambda)+R_2(\lambda)\,.$$

We estimate each term. First of all,

$$R_1(\lambda) \le \alpha\left(\frac{1}{1-T^{-\alpha}}-1\right)\left|\int_1^T x^{-(1+\alpha)}\cos\lambda x\,dx\right|$$
$$+\alpha\left|\int_T^\infty x^{-(1+\alpha)}\cos\lambda x\,dx\right| \le 2T^{-\alpha}.$$

Further, by the definition of C_α,

$$\big|R_2(\lambda)\big| \le \left|1-e^{-C_\alpha\lambda^\alpha}-C_\alpha\lambda^\alpha\right|$$
$$+\left|C_\alpha\lambda^\alpha-\alpha\int_1^\infty\frac{1-\cos\lambda x}{x^{\alpha+1}}\,dx\right| \le \frac{1}{2}C_\alpha^2\lambda^{2\alpha}+\frac{1}{2(2-\alpha)}\lambda^2.$$

Since $\big|R_2(\lambda)\big| \le 2$, we conclude that

$$\big|R_2(\lambda)\big| \le \min\Big(2, B_\alpha\max\big(\lambda^2,\lambda^{2\alpha}\big)\Big),$$

where $B_\alpha = C_\alpha^2/2+1/(2(2-\alpha))$. Summarizing,

$$\big|\psi(\lambda)-\varphi_T(\lambda)\big| \le \begin{cases} 2T^{-\alpha}+2 & \text{if } |\lambda|>1 \\ 2T^{-\alpha}+B_\alpha\max\big(\lambda^2,\lambda^{2\alpha}\big) & \text{if } |\lambda|\le 1 \end{cases}.$$

Let X be a symmetric Pareto random variable with truncated tails and characteristic function given in (4.6), and Y a symmetric α-stable random variable with the characteristic function given by (4.7). If we select

$$\gamma\in\big(\alpha,\min(2,2\alpha)\big),\quad c\in(0,1)\,,$$

then

$$d_{c,\gamma}(X,Y) \le 2T^{-\alpha}c^{-\gamma}+\max\big(2,B_\alpha\big)\,. \tag{4.8}$$

Substituting (4.8) into (4.4), we obtain

$$B_\gamma(a,\Delta) \le 2C_1(\gamma)\,a^{\gamma+1}\Delta^{-\gamma}nT^{-\alpha}+C_4(\gamma;\alpha)\,a^{\gamma+1}n^{-(\gamma/\alpha-1)}+C_2\Delta+C_3M_ha^{-1},$$

were $C_4(\gamma;\alpha)=C_1(\gamma)\max(2,B_\alpha)$. Consider now the range $n\le T^\alpha$. Selecting

$$\tilde\Delta=\big(nT^{-\alpha}\big)^{1/(\gamma+1)}\tilde a,\quad \tilde a=\left(\frac{C_3M_h}{\big(2C_1(\gamma)+C_2\big)\big(nT^{-\alpha}\big)^{1/(\gamma+1)}}\right)^{1/2},$$

we obtain an upper bound on the distance between the density of the smoothed distribution of S_n, and that of the smoothed distribution of Y, given by

$$K_h(F_{S_n}, F_Y) \leq C_5(\gamma) M_h^{1/2} \big(nT^{-\alpha}\big)^{1/(2(\gamma+1))} + C_6(\gamma;\alpha) M_h^{(\gamma+1)/2} \big(n^{-1}T^{\alpha}\big)^{1/2} n^{-(\gamma/\alpha-1)}\,, \tag{4.9}$$

where $C_5(\gamma) = 2\big(C_3(2C_1(\gamma)+C_2)\big)^{1/2}$ and $C_6(\gamma;\alpha) = C_4(\gamma;\alpha) C_3^{(\gamma+1)/2} (2C_1(\gamma)+C_2)^{-(\gamma+1)/2}$.

What the bound (4.9) shows is that, if the truncation level T is large, and the number of terms n in the sum $X_1 + \ldots + X_n$ is such that

$$T^{\alpha\rho} \ll n \ll T^{\alpha}\,, \tag{4.10}$$

where

$$\rho = \frac{1}{2\gamma/\alpha - 1} \in (0,1),$$

then the distance $K_h(F_{S_n}, F_Y)$ will be small.

Therefore, for the symmetric Pareto random variables with truncated tails, T^{α} provides the natural scale: if the number n is much less than this natural scale, but larger than a certain fractional power of this scale, then the distribution of S_n is well approximated by the distribution of the infinite variance symmetric α-stable random variable with the characteristic function given in (4.7).

It is important to note that, unlike the case of the smoothly truncated Lévy flights, for the symmetric Pareto random variables with truncated tails, the range of n where approximation by an α-stable distribution is good, has a lower bound. This is because the distribution of a single symmetric Pareto random variable is not really close to the corresponding α-stable distribution; see Figure 1 above.

5. Proof of the main result

We start with listing, for ease of reference, several well know properties of convolutions and Fourier transforms. To simplify the notation, we will write L^p for $L^p(\mathbb{R}^d, \mathfrak{B}(\mathbb{R}^d), \lambda_d)$, where $\mathfrak{B}(\mathbb{R}^d)$ is the Borel σ-algebra and λ_d is the Lebesgue measure on $\mathbb{R}^d$. The Fourier transform of a function f is denoted by $\tilde{f}$. As usual, the Fourier transform of a function in L^2 is defined by

$$\tilde{f}(z) = \text{l.i.m.}_{N\to\infty} \int_{|x|\le N} e^{i\langle x,z\rangle} f(x)dx,$$

where l.i.m. is understood to be the limit in L^2.

Theorem 5. (1) *Let $1 \le p,q,r \le \infty$ such that $1+1/r = 1/p+1/q$. If $f \in L^p$ and $g \in L^q$, then $f * g$ exists for almost all x, it is an element of L^r and it satisfies Young's Inequality: $\|f*g\|_r \le \|f\|_p\|g\|_q$. If $r=\infty$ then $f*g$ exists for all x.*

(2) *Let $p \ge 1$ and $g \in L^p$. If F is a finite signed measure, then $F \star g$ is defined for Lebesgue a. e. x and $F \star g \in L^p$.*

(3) *Let $f, g \in L^1$. Then $\widetilde{f*g} = \tilde{f}\tilde{g}$.*

(4) *Let $f, g \in L^2$. Then $f * g(x) = (2\pi)^{-d}\widetilde{(\tilde{f}\tilde{g})}(-x)$ for almost all x.*

(5) *Let $f \in L^2$. If $\tilde{f} \in L^1 \cap L^2$ then $f \in L^\infty \cap L^2$ and $\|f\|_\infty \le (2\pi)^{-d}\|\tilde{f}\|_1$.*

Proof. See Propositions 8.6-8.9 in Folland (1999) for Part (1). Part (2) is in Proposition 3.9.9 in Bogachev (2007). The rest of the statements are in Proposition 6.8.1 and Theorem 6.8.1 in Stade (2005). □

Proof of Theorem 1 Suppose that, for $a > 0$, V_a is a measurable function on $\mathbb{R}^d$ with the following properties. Define $B_a(x) := |x|V_a(x)$, $x \in \mathbb{R}^d$. Assume that $V_a, \tilde{V}_a, B_a \in L^1$, $|\tilde{V}_a| \le M$, $\tilde{V}_a(0) = 1$, and $\tilde{V}_a(x) = 0$ for $x \notin [-2a, 2a]^d$. We will show that, under the assumptions of the theorem, we have a bound

$$\begin{aligned} K_h(F_{S_n}, F_Y) \quad &\le \quad \inf_{a,\Delta>0}\Bigg\{\frac{M d_{\Delta n^{-1/\alpha},\gamma}(X_1,Y)}{n^{\gamma/\alpha-1}}\frac{2^{\gamma+1}(a\sqrt{d})^{\gamma+d}}{\pi^{d/2}\Gamma(d/2)(\gamma+d)} \\ &\qquad + \frac{2M}{\pi^d}[\Delta\wedge(2a)]^d + 2M_h\int_{\mathbb{R}^d}|t||V_a(t)|dt\Bigg\}. \end{aligned} \tag{5.1}$$

The proof of the theorem will then be completed by choosing an appropriate function V_a.

Notice that for every $x \in \mathbb{R}^d$,

$$\begin{aligned} |F_{S_n}\star h(x) - F_Y \star h(x)| &\le |F_{S_n}\star h(x) - (F_{S_n}\star h) * V_a(x)| \\ &+ |F_Y \star h(x) - (F_Y\star h) * V_a(x)| \\ &+ \Big|\Big((F_{S_n}\star h) * V_a(x) - (F_{S_n}\star h) * I * V_a(x)\Big) \end{aligned} \tag{5.2}$$

$$-\Big((F_Y\star h)*V_a(x)-(F_Y\star h)*I*V_a(x)\Big)\Big|$$

$$+\Big|(F_{S_n}\star h)*I*V_a(x)-(F_Y\star h)*I*V_a(x)\Big)\Big| := \sum_{j=1}^{4} T_j(x)\ ,$$

where for $\Delta > 0$,

$$I(x)=\prod_{j=1}^{d}\frac{\sin(\Delta x_j)}{\pi x_j},\ x=(x_1,\dots,x_d)\in\mathbb{R}^d.$$

Note that I is an L^2 function and its Fourier transform is given by $\tilde{I}(z) = 1_{[-\Delta,\Delta]^d}(z)=\prod_{j=1}^{d}1_{[-\Delta,\Delta]}(z_j)$. All the convolutions in (5.2) are well defined by parts (1) and (2) of Theorem 5.

Note that $\int_{\mathbb{R}^d}V_a(x)dx=\tilde{V}_a(0)=1$. If $G\in Lip_M$ then

$$\begin{aligned}|G(x)-G*V_a(x)| &\le \int_{\mathbb{R}^d}|G(x)-G(x-t)|\,|V_a(t)|dt\\ &\le M\int_{\mathbb{R}^d}|t||V_a(t)|dt.\end{aligned}\tag{5.3}$$

Since $h\in Lip_{M_h}$, so are $F_{S_n}\star h$ and $F_Y\star h$. We conclude that

$$T_j(x)\le M_h\int_{\mathbb{R}^d}|t||V_a(t)|dt,\quad j=1,2\,.\tag{5.4}$$

Further, by part (5) of Theorem 5, $V_a\in L^p$ for all $1\le p\le\infty$, and, clearly, so is the function $(F_{S_n}-F_Y)\star h$. By part (1) of Theorem 5, the same is true for the convolution $[(F_{S_n}-F_Y)\star h]*V_a$. Denote by Z is a random vector with density h, independent, where appropriate, of S_n and Y. By parts (3) and (5) of Theorem 5 we obtain

$$\begin{aligned}T_4(x)\le \|[(F_{S_n}-F_Y)\star h]*V_a*I\|_\infty &\le (2\pi)^{-d}\|[\hat{\mu}_{S_n+Z}-\hat{\mu}_{Y+Z}]\tilde{V}_a\tilde{I}\|_1\\ &\le \frac{2M}{\pi^d}[\Delta\wedge(2a)]^d.\end{aligned}\tag{5.5}$$

This leaves only one term to consider in (5.2). By parts (3) and (4) of Theorem 5 we have

$$T_3(x)=(2\pi)^{-d}\left|\int_{\mathbb{R}^d}(1-\tilde{I}(z))\left(\hat{\mu}_{S_n}(z)-\hat{\mu}_Y(z)\right)\tilde{h}(z)\tilde{V}_a(z)e^{-i\langle z,x\rangle}dz\right|.$$

Note that for every $z\in\mathbb{R}^d$ and $\Delta>0$ we have, by the strict stability of Y,

$$\Big|(1-\tilde{I}(z))\hat{\mu}_{S_n}(z)-(1-\tilde{I}(z))\hat{\mu}_Y(z)\Big|\le|z|^\gamma\sup_{|t|>\Delta}\frac{|\hat{\mu}_{S_n}(t)-\hat{\mu}_Y(t)|}{|t|^\gamma}$$

$$= \quad |z|^\gamma d_{\Delta,\gamma}(S_n,Y)$$

$$\begin{aligned}
&\le \quad |z|^\gamma n \sup_{|z|\ge\Delta} \frac{\left|\hat\mu_X(z/n^{1/\alpha}) - \hat\mu_Y(z/n^{1/\alpha})\right|}{|z|^\gamma} \\
&= \quad \frac{|z|^\gamma}{n^{\gamma/\alpha-1}} d_{\Delta n^{-1/\alpha},\gamma}(X,Y).
\end{aligned}$$

Therefore,

$$\begin{aligned}
T_3(x) &\le (2\pi)^{-d}\int_{\mathbb{R}^d}\left|(1-\tilde I(z))[\hat\mu_{S_n}(z)-\hat\mu_Y(z)]\tilde h(z)\tilde V_a(z)\right| dz \\
&\le \frac{(2\pi)^{-d}}{n^{\gamma/\alpha-1}} d_{\Delta n^{-1/\alpha},\gamma}(X,Y)\int_{\mathbb{R}}|z|^\gamma|\tilde V_a(z)|dz \qquad (5.6)\\
&\le \frac{M(2\pi)^{-d}}{n^{\gamma/\alpha-1}} d_{\Delta n^{-1/\alpha},\gamma}(X,Y)\int_{[-2a,2a]^d}|z|^\gamma dz \\
&\le \frac{M(2\pi)^{-d}}{n^{\gamma/\alpha-1}} d_{\Delta n^{-1/\alpha},\gamma}(X,Y)\int_{|z|\le 2a\sqrt{d}}|z|^\gamma dz \\
&= \frac{M d_{\Delta n^{-1/\alpha},\gamma}(X,Y)}{n^{\gamma/\alpha-1}}\frac{2^{1+\gamma}(a\sqrt{d})^{\gamma+d}}{\pi^{d/2}\Gamma(d/2)(\gamma+d)};
\end{aligned}$$

the last line follows by conversion to polar coordinates, see e.g. Section 5.2 in Stroock (1999). Now (5.1) follows from (5.2), (5.4), (5.5) and (5.6).

Let $W(x) = \frac{12\sin^4(x/2)}{\pi x^4}$, $x\in\mathbb{R}$. This is called the Jackson-de la Vallée-Poussin kernel. Its Fourier transform is given by

$$\tilde W(x) = \begin{cases} 1-\frac{3x^2}{2}+\frac{3|x|^3}{4} & |x|\le 1 \\ \frac14(2-|x|)^3 & 1\le|x|\le 2 \\ 0 & |x|\ge 2\end{cases},$$

see page 119 in Achieser (1992). For $a>0$, let $W_a(x) = aW(xa)$, so that $\tilde W_a(x) = \tilde W(x/a)$, and define for $x=(x_1,\dots,x_d)\in\mathbb{R}^d$,

$$V_a(x) = \prod_{j=1}^d W_a(x_j).$$

Then also

$$\tilde V_a(z) = \prod_{j=1}^d \tilde W_a(z_j)$$

for $z=(z_1,\dots,z_d)\in\mathbb{R}^d$.

Let $B_a(x) = |x|V_a(x)$. Note that $\tilde V_a(z)\le 1$, V_a, $\tilde V_a$, $B_a\in L^1$, and $\tilde V(0) = 1$. Therefore, the function V_a satisfies the assumptions imposed in the beginning of the proof. Further, we have

$$\int_{\mathbb{R}^d} |x|\, |V_a(x)|\, dx = \left(\frac{12}{\pi}\right)^d \frac{2^{1-3d}}{a} \int_{\mathbb{R}^d} |x| \prod_{j=1}^{d} \frac{\sin^4(x_i)}{x_i^4} dx \leq \frac{6d}{\pi a}. \tag{5.7}$$

This follows easily from the facts that $|x| \leq \sum_{i=1}^d |x_i|$, $\int_0^\infty \frac{\sin^4 x}{x^4} dx = \pi/3$ (see Gradshteyn and Ryzhik (2000) 3.821), and

$$\int_0^\infty \frac{\sin^4 v}{v^3} dv = \int_0^1 \frac{\sin^4 v}{v^3} dv + \int_1^\infty \frac{\sin^4 v}{v^3} dv \leq \int_0^1 v dv + \int_1^\infty \frac{1}{v^3} dv = 1.$$

This completes the proof of Theorem 1. □

Acknowledgments

M. Grabchak wishes to thank Professor Camil Muscalu for the benefit of several discussions on some of the technical aspects of the paper. A part of G. Samorodnitsky's research was done while visiting the Department of Mathematics at the University of Copenhagen. The hospitality and the support of the department are gratefully acknowledged.

References

N. Achieser (1992): *Theory of Approximation*. Dover Publications, Inc., New York.

V. Akgirav and G. Booth (1988): The stable-law model of stock returns. *J. Bus. Econ. Stat.* 6:51–57.

R. Blattberg and N. Gonedes (1974): A comparison of the stable and Student distributions as statistical models for stock prices. *Journal of Business* 47:244–280.

V. Bogachev (2007): *Measure Theory*, volume I. Springer, Berlin.

P. Carr, H. Geman, D. Madan and M. Yor (2002): The fine structure of asset returns: an empirical investigation. *Journal of Business* 75:305–332.

P. Carr, H. Geman, D. Madan and M. Yor (2003): Stochastic volatility for Lévy processes. *Mathematical Finance* 13:345–382.

R. Cont and P. Tankov (2004): *Financial Modeling With Jump Processes*. Chapman & Hall, Boca Raton.

P. Embrechts, C. Klüppelberg and T. Mikosch (1997): *Modelling Extremal Events for Insurance and Finance*. Springer-Verlag, Berlin.

E. Fama (1965): The behavior of stock market prices. *Journal of Business* 38:34–105.

W. Feller (1971): *An Introduction to Probability Theory and its Applications*, volume 2. Wiley, New York, 2nd edition.

G. Folland (1999): *Real Analysis: Modern Techniques and Their Applications*. Wiley, New York, 2nd edition.

R. Gencay, M. Dacorogna, U. Muller, O. Pictet and R. Olsen (2001): *An Introduction to High-Frequency Finance*. Academic Press, London.

I. Gradshteyn and I. Ryzhik (2000): *Table of Integrals, Series, and Products*. Academic Press, San Diego, sixth edition.

L. Klebanov, S. Rachev and M. Safarian (2000): Local prelimit theorems and their applications to finance. *Applied Mathematics Letters* 13:73–78.

L. Klebanov, S. Rachev and G. Szekely (1999): Pre-limit theorems and their applications. *Acta Applicandae Mathematicae* 58:159–174.

I. Koponen (1995): Analytic approach to the problem of convergence of truncated Lévy flights towards the Gaussian stochastic process. *Physical Review E* 52:1197–1199.

A. Lau, H.-S. Lau and J. Wingender (1990): The Distribution of Stock Returns: New Evidence against the Stable Model. *Journal of Business and Economic Statistics* 8:217–223.

B. Mandelbrot (1963): The variation of certain speculative prices. *Journal of Business* 26:394–419.

T. Mikosch and G. Samorodnitsky (2000): The supremum of a negative drift random walk with dependent heavy–tailed steps. *Annals of Applied Probability* 10:1025–1064.

S. Mittnik and S. Rachev (2000): *Stable Paretian Models in Finance*. Wiley, New York.

J. Rosiński (2007): Tempering stable processes. *Stochastic Processes and their Applications* 117:677–707.

G. Samorodnitsky and M. Taqqu (1994): *Stable Non-Gaussian Random*

Processes. Chapman and Hall, New York.

E. Stade (2005): *Fourier Analysis*. Wiley-Interscience, Hoboken.

D. Stroock (1999): *A Concise Introduction to the Theory of Integration, Third Edition*. Birkhauser, Boston.

N. Taleb (2009): Finiteness of variance is irrelevant in the practice of quantitative finance. *Complexity* 14:66–76.

Department of Statistics, Cornell University, Ithaca, NY 14853
E-mail address: `mg323@cornell.edu`

School of Operations Research and Information Engineering, and Department of Statistical Science, Cornell University, Ithaca, NY 14853
E-mail address: `gennady@orie.cornell.edu`