Financial Economics, Fat-tailed Distributions*

Markus Haas†
Christian Pigorsch‡

First version: April 13, 2007
This version: September 21, 2007

*We are grateful to Bruce Mizrach for his helpful comments. The research of Haas was supported by the German Research Foundation.

†Department of Statistics, University of Munich (LMU), D-80799 München, Akademiestr 1/I. E-mail: haas@stat.uni-muenchen.de. Phone: +49 (0) 89 21 80 - 25 70.

‡Department of Economics, University of Bonn, D-53113 Bonn, Adenauerallee 24-42. E-mail: christian.pigorsch@uni-bonn.de. Phone: +49 (0) 228 73 - 39 20.

Article Outline

1 Glossary

Leptokurtosis
A distribution is leptokurtic if it is more peaked in the center and thicker tailed than the normal distribution with the same mean and variance. Occasionally, leptokurtosis is also identified with a moment–based kurtosis measure larger than three, see Section 3.

Return
Let S_t be the price of a financial asset at time t. Then the *continuous* return, r_t, is $r_t = \log(S_t/S_{t-1})$. The *discrete* return, R_t, is $R_t = S_t/S_{t-1} - 1$. Both are rather similar if $-0.15 < R_t < 0.15$, because $r_t = \log(1 + R_t)$. See Section 3.

Tail
The (upper) tail, denoted by $\overline{F}(x) = P(X > x)$, characterizes the probability that a random variable X exceeds a certain "large" threshold x. For analytical purposes, "large" is often translated with "as $x \to \infty$". For financial returns, a daily change of 5% is already infinitely large. A Gaussian model essentially excludes such an event.

Tail index
The tail index, or *tail exponent*, α, charcaterizes the rate of tail decay if the tail goes to zero, in essence, like a power function, i.e., $\overline{F}(x) = x^{-\alpha}L(x)$, where L is slowly varying. Moments of order lower (higher) than α are (in)finite.

2 Definition of the Subject and Its' Importance

Have a look at Figure 1. Its top plot shows the daily percentage changes, or *returns*, of the S&P500 index ranging from January 2, 1985 to December 29, 2006, a total of 5550 daily observations. We will use this data set throughout the article to illustrate some of the concepts and models to be discussed. Two observations are immediate. The first is that both small and large changes come clustered, i.e., there are periods of low and high volatility. The second is that, from time to time, we observe rather large changes which may be hard to reconcile with *the* standard distributional assumption in statistics and econometrics, that is, normality. The most outstanding return certainly occurred on October 19, 1987, the "Black Monday", where the index lost more than 20% of its value, but the phenomenon is chronic. For example, if we fit a normal distribution to the data, the resulting model predicts that we observe an absolute daily change larger than 5% once in approximately 1860 years, whereas we actually encountered that 13 times during our 22–year sample period. This suggests that, compared to the normal distribution, the distribution of the returns is *fat–tailed*, i.e., the probability of large losses and gains is much higher than would be implied by a time–invariant unconditional Gaussian distribution. The latter is obviously not suitable for describing the booms, busts, bubbles, and bursts of activity which characterize financial markets, and which are apparent in Figure 1.

The two aforementioned phenomena, i.e., volatility clustering and fat tails, have been detected in almost every financial return series that was subject to statistical analysis since the publication of Mandelbrot's [154] seminal study of cotton price changes, and they are of paramount importance for any individual or institution engaging in the financial markets, as well as for financial economists trying to understand their mode of operation. For example, investors holding significant portions of their wealth in risky assets need a realistic assessment of the likelihood of severe losses. Similarly, economists trying to learn about the relation between risk and return, the pricing of financial derivatives, such as options, and the inherent dynamics of financial markets, can only benefit from building their models on adequate assumptions about the stochastic properties of the variables under study, and they have to reconcile the predictions of their models with the actual facts.

This article reviews some of the most important concepts and distributional models that are used in empirical finance to capture the (almost) ubiquitous stochastic properties of returns as indicated above. Section 3 defines in a somewhat more precise manner than above the central variable of interest, the return of a financial asset, and gives a brief account of the early history of the problem. Section 4 discusses various operationalizations of the term "fat–tailedness", and Section 4 summarizes what is or is at least widely believed to be known about the tail characteristics of typical return distributions. Popular parametric distributional models are discussed in Section 6. The alpha stable model as the archetype of a fat–tailed distribution in finance is considered in detail, as is the generalized hyperbolic distribution, which provides a convenient framework for discussing, as special or limiting cases, many of the important distributions employed in the literature. An empirical comparison using the S&P500 returns is also included. In Section 7, the relation between the two "stylized facts" mentioned above, i.e., clusters of volatility and fatness of the tails, is highlighted, where we concentrate on the GARCH approach, which has gained outstanding popularity among financial econometricians. This model has the intriguing property of producing fat–tailed marginal distributions even with light–tailed innovation processes, thus emphasizing the role of the market dynamics. In Section 8, we compare both the unconditional parametric distributional models introduced in Section 6 as well as the GARCH model of Section 7 on an economic basis by evaluating their ability to accurately measure the Value-at-Risk, which is an important tool in risk management. Finally, Section 9 identifies some open issues.

3 Introduction

To fix notation, let S_t be the price of an asset at time t, e.g., a stock, a market index, or an exchange rate. The *continuously compounded* or *log* return from time t to time $t + \Delta t$, $r_{t,t+\Delta t}$, is then defined as

$$r_{t,t+\Delta t} = \log S_{t+\Delta t} - \log S_t. \tag{1}$$

Often the quantity defined in (1) is also multiplied by 100, so that it can be interpreted in terms of *percentage returns*, see Figure 1. Moreover, in applications, Δt is usually set equal to one and represents the horizon over which the returns are calculated, e.g., a day,

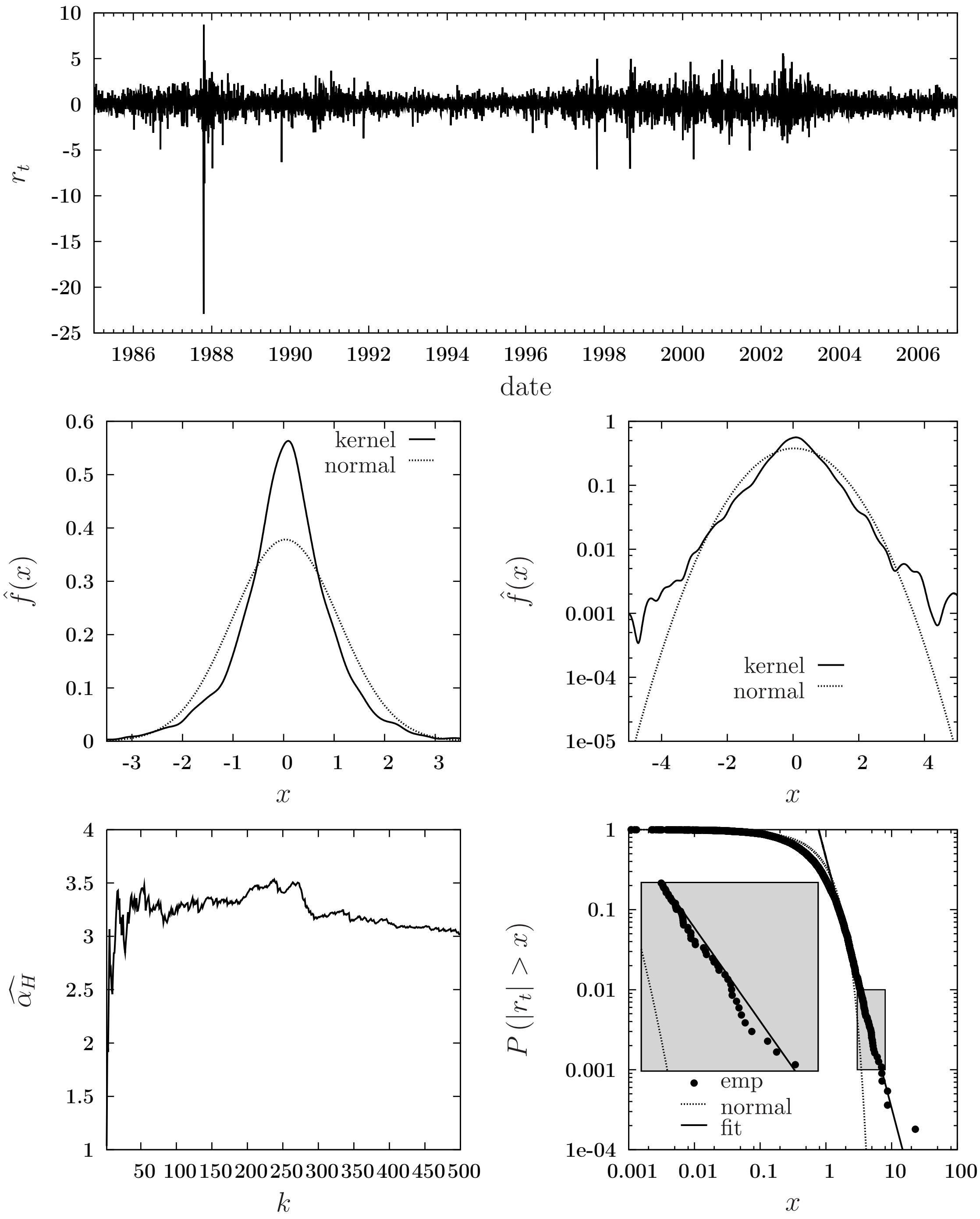

Figure 1: The top plot shows the S&P500 *percentage* returns, r_t, from January 1985 to December 2006, i.e., $r_t = 100 \times \log(S_t/S_{t-1})$, where S_t is the index level at time t. The left plot of the middle panel shows a nonparametric density estimate (solid), along with the fitted normal density (dotted); the right graph is similar but shows the respective log–densities in order to better visualize the tail regions. The bottom left plot represents a *Hill plot* for the S&P500 returns, i.e., it displays $\widehat{\alpha}_{k,n}$ defined in (11) for $k \leq 500$. The bottom right plot shows the complementary cdf, $\overline{F}(x)$, on a log–log scale, see Section 5 for discussion.

week, or month. In this case, we drop the first subscript and define $r_t := \log S_t - \log S_{t-1}$. The log returns (1) can be additively aggregated over time, i.e.,

$$r_{t,t+\tau} = \sum_{i=1}^{\tau} r_{t+i}. \tag{2}$$

Empirical work on the distribution of financial returns is usually based on log returns. In some applications a useful fact is that, over short intervals of time, when returns tend to be small, (1) can also serve as a reasonable approximation to the *discrete* return, $R_{t,t+\Delta t} := S_{t+\Delta t}/S_t - 1 = \exp\left(r_{t,t+\Delta t}\right) - 1$. For further discussion of the relationship between continuous and discrete returns and their respective advantages and disadvantages, see, e.g., [46, 76].

The seminal work of [155], to be discussed in Section 6.1, is often viewed as the beginning of modern empirical finance. As reported in [74], "[p]rior to the work of Mandelbrot the usual assumption ... was that the distribution of price changes in a speculative series is approximately Gaussian or normal". The rationale behind this prevalent view, which was explicitly put forward as early as 1900 by Bachelier [14], was clearly set out in [178]: If the log–price changes (1) from transaction to transaction are independently and identically distributed with finite variance, and if the number of transactions is fairly uniformly distributed in time, then (2) along with the central limit theorem (CLT) implies that the return distribution over longer intervals, such as a day, a week, or a month, approaches a Gaussian shape.

However, it is now generally acknowledged that the distribution of financial returns over horizons shorter than a month is not well described by a normal distribution. In particular, the empirical return distributions, while unimodal and approximately symmetric, are typically found to exhibit considerable *leptokurtosis*, i.e., they are more peaked in the center and have fatter tails than the Gaussian with the same variance. Although this has been occasionally observed in the pre–Mandelbrot literature (e.g., [6]), the first systematic account of this phenomenon appeared in [154] and the follow–up papers by Fama [74, 75] and Mandelbrot [156], and it was consistently confirmed since then. The typical shape of the return distribution, as compared to a fitted Gaussian, is illustrated in the middle panel of Figure 1 for the S&P500 index returns, where a nonparametric kernel density estimator (e.g., [198]) is superimposed on the fitted Gaussian curve (dashed line). Interestingly, this pattern has been detected not only for modern financial markets but also for those of the eighteenth century [103].

The (location and scale–free) standardized fourth moment, or coefficient of *kurtosis*,

$$\mathbb{K}\left[X\right] = \frac{\mathbb{E}\left[(X-\mu)^4\right]}{\sigma^4}, \tag{3}$$

where μ and σ are the mean and the standard deviation of X, respectively, is sometimes used to assess the degree of leptokurtosis of a given distribution. For the normal distribution, $\mathbb{K} = 3$, and $\mathbb{K} > 3$, referred to as *excess kurtosis*, is taken as an indicator of a leptokurtic shape (e.g., [164], p. 429). For example, the sample analogue of (3) for the S&P500 returns shown in Figure 1 is 47.9, indicating very strong excess kurtosis. A formal test could be conducted using the fact that, under normality, the sample kurtosis is asymptotically normal with mean 3 and standard deviation $\sqrt{24/T}$ (T being the sample size), but the result can be anticipated.

As is well–known, however, such moment–based summary measures have to be interpreted with care, because a particular moment need not be very informative about a density's shape. We know from Finucan [82] that if two symmetric densities, f and g, have common mean and variance and finite fourth moment, and if g is more peaked and has thicker tails than f, *then* the fourth moment (and hence $\mathbb{K}$) is greater for g than for f, provided the densities cross exactly twice on both sides of the mean. However, the converse of this statement is, of course, not true, and a couple of (mostly somewhat artificial) counterexamples can be found in [16, 68, 121]. [158] provides some intuition by relating density crossings to moment crossings. For example, (only) if the densities cross more than four times, it may happen that the fourth moment is greater for f, but the sixth and all higher moments are greater for g, reflecting the thicker tails of the latter. Nevertheless, Finucan's result, along with his (in some respects justified) hope that we can view "this pattern as the common explanation of a high observed kurtosis", may serve to argue for a certain degree of usefulness of the kurtosis measure (3), provided the fourth moment is assumed to be finite. However, a nonparametric density estimate will in any case be more informative. Note that the density crossing condition in Finucan's theorem is satisfied for the S&P500 returns in Figure 1.

4 Defining Fat–tailedness

The notion of leptokurtosis as discussed so far is rather vague, and both financial market researchers as well as practitioners, such as risk managers, are interested in a more precise description of the tail behavior of financial variables, i.e., the laws governing the probability of large gains and losses. To this end, we define the *upper tail* of the distribution of a random variable (rv) X as

$$\overline{F}(x) = P(X > x) = 1 - F(x), \tag{4}$$

where F is the cumulative distribution function (cdf) of X. Consideration of the upper tail is the standard convention in the literature, but it is clear that everything could be phrased just as well in terms of the lower tail.

We are interested in the behavior of (4) as x becomes large. For our benchmark, i.e., the normal distribution with (standardized) density (pdf) $\phi(x) = (2\pi)^{-1/2}\exp(-x^2/2)$, we have (cf. [79], p. 131)

$$\overline{F}(x) \cong \frac{1}{\sqrt{2\pi}x}\exp\left(-\frac{x^2}{2}\right) = \frac{\phi(x)}{x} \quad \text{as } x \to \infty, \tag{5}$$

where the notation $f(x) \cong g(x)$ as $x \to \infty$ means that $\lim_{x\to\infty} f(x)/g(x) = 1$. Thus, the tails of the normal tend to zero faster than exponentially, establishing its very light tails.

To appreciate the difference between the general concept of leptokurtosis and the approach that focuses on the tails, consider the class of finite normal mixtures as discussed in Section 6.3. These a leptokurtic in the sense of peakedness and tailedness (compared to the normal), but are light–tailed according to the tail–based perspective.

While it is universally accepted in the literature that the Gaussian is too light–tailed to be an appropriate model for the distribution of financial returns, there is no complete agreement with respect to the actual shape of the tails. This is not surprising because we cannot reasonably expect to find a model that accurately fits all markets at any time and place. However, the current mainstream opinion is that the probability for the occurrence of large (positive and negative) returns can often appropriately be described by Pareto–type tails. Such tail behavior is also frequently adopted as the definition of fat–tailedness per se, but the terminology in the literature is by no means unique.

A distribution has Pareto–type tails if they decay essentially like a power function as x becomes large, i.e., $\overline{F}$ is regularly varying (at infinity) with index $-\alpha$ (written $\overline{F} \in \mathrm{RV}_{-\alpha}$), meaning that

$$\overline{F}(x) = x^{-\alpha}L(x), \quad \alpha > 0, \tag{6}$$

where $L > 0$ is a slowly varying function, which can be interpreted as "slower than any power function" (see [34, 188, 195] for a technical treatment of regular variation). The defining property of a slowly varying function is $\lim_{x\to\infty} L(tx)/L(x) = 1$ for any $t > 0$, and the aforementioned interpretation follows from the fact that, for any $\gamma > 0$, we have (cf. [195], p. 18)

$$\lim_{x\to\infty} x^{\gamma}L(x) = \infty, \quad \lim_{x\to\infty} x^{-\gamma}L(x) = 0. \tag{7}$$

Thus, for large x, the parameter α in (6), called the *tail index* or *tail exponent*, controls the rate of tail decay and provides a measure for the fatness of the tails.

Typical examples of slowly varying functions include $L(x) = c$, a constant, $L(x) = c + o(1)$, or $L(x) = (\log x)^k$, $x > 1$, $k \in \mathbb{R}$. The first case corresponds to strict Pareto tails, while in the second the tails are asymptotically Paretian in the sense that $\overline{F}(x) \cong cx^{-\alpha}$, which includes as important examples in finance the (nonnormal) stable Paretian (see (13) in Section 6.1) and the Student's t distribution considered in Section 6.2.3, where the tail index coincides with the characteristic exponent and the number of degrees of freedom, respectively. As an example for both, the Cauchy distribution with density $f(x) = [\pi(1 + x^2)]^{-1}$ has cdf $F(x) = 0.5 + \pi^{-1}\arctan(x)$. As $\arctan(x) = \sum_0^\infty (-1)^i x^{2i+1}/(2i+1)$ for $|x| < 1$, and $\arctan(x) = \pi/2 - \arctan(1/x)$ for $x > 0$, we have $\overline{F}(x) \cong (\pi x)^{-1}$.

For the distributions mentioned in the previous paragraph, it is known that their moments exist only up to (and excluding) their tail indices, α. This is generally true for rvs with regularly varying tails and follows from (7) along with the well–known connection between moments and tail probabilities, i.e., for a nonnegative rv X, and $r > 0$, $\mathbb{E}[X^r] = r\int_0^\infty x^{r-1}\overline{F}(x)dx$ (cf. [95], p. 75). The only possible minor variation is that, depending on L, $\mathbb{E}[X^\alpha]$ may be finite. For example, a rv X with tail $\overline{F}(x) = cx^{-1}(\log x)^{-2}$ has finite mean. The property that moments greater than α do not exist provides further intuition for α as a measure of tail–fatness.

As indicated above, there is no universal agreement in the literature with respect to the definition of fat–tailedness. For example, some authors (e.g., [72, 196]) emphasize the class of *subexponential* distributions, which are (although not exclusively) characterized by the property that their tails tend to zero slower than any exponential, i.e., for any $\gamma > 0$, $\lim_{x\to\infty} e^{\gamma x}\overline{F}(x) = \infty$, implying that the moment generating function does not exist. Clearly a regularly varying distribution is also subexponential, but further members of this class are, for instance, the lognormal as well as the *stretched exponential*, or heavy tailed Weibull, which has a tail of the form

$$\overline{F}(x) = \exp(-x^b), \quad 0 < b < 1. \tag{8}$$

The stretched exponential has recently been considered by [134, 152, 153] as an alternative to the Pareto–type distribution (6) for modeling the tails of asset returns. Note that, as opposed to (6), both the lognormal as well as the stretched exponential possess power moments of all orders, although no exponential moment.

In addition, [24] coined the term *semi–heavy tails* for the generalized hyperbolic (GH) distribution, but the label may be employed more generally to refer to distributions with slower tails than the normal but existing moment generating function. The GH, which is now very popular in finance and nests many interesting special cases, will be examined in detail in Section 6.2.

As will be discussed in Section 5, results of extreme value theory (EVT) are often employed to identify the tail shape of return distributions. This has the advantage that it allows to concentrate fully on the tail behavior, without the need to model the central part of the distribution. To sketch the idea behind this approach, suppose we attempt to classify distributions according to the limiting behavior of their normalized maxima. To this end, let $\{X_i, i \geq 1\}$ be an iid sequence of rvs with common cdf F, $M_n = \max\{X_1, \ldots, X_n\}$, and assume there exist sequences $a_n > 0$, $b_n \in \mathbb{R}$, $n \geq 1$, such that

$$P\left(\frac{M_n - b_n}{a_n} \leq x\right) = F^n(a_n x + b_n) \overset{n\to\infty}{\longrightarrow} G(x), \tag{9}$$

where G is assumed nondegenerate. To see that normalization is necessary, note that $\lim_{n\to\infty} P(M_n \leq x) = \lim_{n\to\infty} F^n(x) = 0$ for all $x < x_M := \sup\{x : F(x) < 1\} \leq \infty$, so that the limiting distribution is degenerate and of little help. If the above assumptions are satisfied, then, according to the classical Fisher–Tippett theorem of extreme value theory (cf. [188]), the limiting distribution G in (9) must be of the following form:

$$G_\xi(x) = \exp\left(-(1+\xi x)^{-1/\xi}\right), \quad 1 + \xi x > 0, \tag{10}$$

which is known as the *generalized extreme value distribution* (GEV), or *von Mises representation of the extreme value distributions* (EV). The latter term can be explained by the fact that (10) actually nests three different types of EV distributions, namely

(i) the Fréchet distribution, denoted by G_ξ^+, where $\xi > 0$ and $x > -1/\xi$,

(i) the so–called Weibull distribution of EVT, denoted by G_ξ^-, where $\xi < 0$ and $x < -1/\xi$, and

(iii) the Gumbel distribution, denoted by G_0, which corresponds to the limiting case as $\xi \to 0$, i.e., $G_0(x) = \exp(-e^{-x})$, where $x \in \mathbb{R}$.

A cdf F belongs to the *maximum domain of attraction* (MDA) of one of the extreme value distributions nested in (10), written $F \in \text{MDA}(G_\xi)$, if (9) holds, i.e., classifying distributions according to the limiting behavior of their extrema amounts to figuring out the MDAs of the extreme value distributions. It turns out that it is the tail behavior of a distribution F that accounts for the MDA it belongs to. In particular, $F \in \text{MDA}(G_\xi^+)$ if and only if its tail $\overline{F} \in \text{RV}_{-\alpha}$, where $\alpha = 1/\xi$. As an example, for a strict Pareto distribution, i.e., $F(x) = 1 - (u/x)^\alpha$, $x \geq u > 0$, with $a_n = n^{1/\alpha}u/\alpha$ and $b_n = n^{1/\alpha}u$, we have $\lim_{n\to\infty} F^n(a_n x + b_n) = \lim_{x\to\infty}(1 - n^{-1}(1 + x/\alpha)^{-\alpha})^n = G_{1/\alpha}^+(x)$. Distributions in $\text{MDA}(G_\xi^-)$ have a finite right endpoint, while, roughly, most of the remaining distributions, such as the normal, the lognormal and (stretched) exponentials, belong to $\text{MDA}(G_0)$. The latter also accommodates a few distributions with finite right endpoint. See [188] for precise conditions. The important case of non–iid rvs is discussed in [136]. A central result is that, rather generally, *vis–à–vis* an iid sequence with the same marginal cdf, the maxima of stationary sequences converge to the same type of limiting distribution. See [63] and [166] for an application of this theory to ARCH(1) and GARCH(1,1) processes (see Section 7), respectively

One approach to exploit the above results, referred to as the *method of block maxima*, is to divide a given sample of return data into subsamples of equal length, pick the maximum of each subsample, assume that these have been generated by (10) (enriched with location and scale parameters to account for the unknown a_n and b_n), and find the maximum–likelihood estimate for ξ, location, and scale. Standard tests can then be conducted to assess, e.g., whether $\xi > 0$, i.e., the return distribution has Pareto–type tails. An alternative but related approach, which is based on further theoretical developments and often makes more efficient use of the data, is the *peaks over thresholds* (POT) method. See [72] for a critical discussion of these and alternative techniques.

We finally note that $1 - G_{1/\alpha}^+(\alpha(x-1)) \cong x^{-\alpha}$, while $1 - G_0(x) \cong e^{-x}$, i.e., for the extremes, we have asymptotically a Pareto and an exponential tail, respectively. This may provide, on a meta–level, a certain rationale for reserving the notion of genuine fat–tailedness for the distributions with regularly varying tails.

5 Empirical Evidence about the Tails

The first application of power tails in finance appeared in Mandelbrot's [154] study of the log–price changes of cotton. Mandelbrot proposed to model returns with nonnormal *alpha stable*, or *stable Paretian*, distributions, the properties of which will be discussed in some detail in Section 6.1. For the present discussion, it suffices to note that for this model the tail index α in (6), also referred to as *characteristic exponent* in the context of stable distributions, is restricted to the range $0 < \alpha < 2$, and that much of its theoretical appeal derives from the fact that, due to the generalized CLT, "Mandelbrot's hypothesis can actually be viewed as a generalization of the central–limit theorem arguments of Bachelier and Osborne to the case where the underlying distributions of price changes from transaction to transaction ... have infinite variances" [75]. For the cotton price changes, Mandelbrot came up with a tail index of about 1.7, and his work was subsequently complemented by Fama [75] with an analysis of daily returns of the stocks belonging to the Dow Jones Industrial Average. [75] came to the conclusion that Mandelbrot's theory was supported by these data, with an average estimated α close to 1.9.

The findings of Mandelbrot and Fama initiated an extensive discussion about the appropriate distributional model for stock returns, partly because the stable model's implication that the tails are so fat that even the variance is infinite appeared to be too radical to many economists used to work with models built on the assumption of finite second moments. The evidence concerning the stable hypothesis gathered in the course of the debate until the end of the eighties was not ultimately conclusive, but there were many papers reporting mainly negative results [4, 28, 36, 40, 54, 67, 98, 99, 109, 135, 176, 180, 184].

From the beginning of the nineties, a number of researchers have attempted to estimate the tail behavior of asset returns directly, i.e., without making specific assumptions about the entire distributional shape. [86, 115, 142, 143] use the method of block maxima (see Section 4) to identify the maximum domain of attraction of the distribution of stock returns. They conclude that the Fréchet distribution with a tail index $\alpha > 2$ is most likely, implying Pareto–type tails which are thinner than those of the *stable* Paretian.

A second strand of literature assumes a priori the presence of a Pareto–type tail and focuses on the estimation of the tail index α. If, as is often the case, a power tail is deemed adequate, an explicit estimate of α is of great interest both from a practical and an academic viewpoint. For example, investors want to assess the likelihood of large losses of financial assets. This is often done using methods of extreme value theory, which require an accurate estimate of the tail exponent. Such estimates are also important because the properties of statistical tests and other quantities of interest, such as empirical autocorrelation functions, frequently depend on certain moment conditions (e.g., [144, 166]). Clearly the desire to figure out the actual tail shape has also an intrinsic component, as is reflected in the long–standing debate on the stable Paretian hypothesis. People simply wanted to know whether this distribution, with its appealing theoretical properties, is consistent with actual data. Moreover, empirical findings may guide economic theorizing, as they can help both in assessing the validity of certain existing models as well as in suggesting new explanations. Two examples will briefly be discussed at the end of the present section.

Within this second strand of literature, the Hill estimator [106] has become the most popular tool. It is given by

$$\widehat{\alpha}_{k,n} = \left(\frac{1}{k-1} \sum_{j=1}^{k-1} \log X_{j,n} - \log X_{k,n} \right)^{-1}, \tag{11}$$

where $X_{i,n}$ denotes the ith upper order statistic of a sample of length n, i.e., $X_{1,n} \geq X_{2,n} \geq \cdots \geq X_{n,n}$. See [64, 72] for various approaches to deriving (11). If the tail is not regularly varying, the Hill estimator does not estimate anything.

A crucial choice to be made when using (11) is the selection of the threshold value k, i.e., the number of order statistics to be included in the estimation. Ideally, only observations from the tail region should be used, but choosing k to small will reduce the precision of the estimator. There exist various methods for picking k optimally in a mean–squared error sense [62, 61], but much can be learned by looking at the *Hill plot*, which is obtained by plotting $\widehat{\alpha}_{k,n}$ against k. If we can find a range of k–values where the estimate is approximately constant, this can be taken as a hint for where the "true" tail index may be located. As illustrated in [189], however, the Hill plot may not always be so well–behaved, and in this case the semiautomatic methods mentioned above will presumably also be of little help.

The theoretical properties of (11), along with technical conditions, are summarized in [72, 189]. Briefly, for iid data generated from a distribution with tail $\overline{F} \in \mathrm{RV}_{-\alpha}$, the Hill estimator has been shown to be consistent [159] and asymptotically normal with standard deviation $\alpha/\sqrt{k}$ [100]. Financial data, however, are usually not iid but exhibit considerable dependencies in higher–order moments (see Section 7). In this situation, i.e., with ARCH–type dynamics, (11) will still be consistent [190], but little is known about its asymptotic variance. However, simulations conducted in [123] with an IGARCH model, as defined in Section 7, indicate that, under such dependencies, the actual standard errors may be seven to eight times larger than those implied by the asymptotic theory for iid variables.

The Hill estimator was introduced into the econometrics literature in the series of articles [107, 113, 125, 126]. [125, 126], using weekly observations, compare the tails of exchange rate returns in floating and fixed exchange rate systems, such as the Bretton Woods period and the EMS. They find that for the fixed systems, most tail index estimates are below 2, i.e., consistent with the alpha stable hypothesis, while the estimates are significantly larger than 2 (ranging approximately from 2.5 to 4) for the float. [126] interpret these results in the sense that "a float lets exchange rates adjust more smoothly than any other regime that involves some amount of fixity". Subsequent studies of floating exchange rates using data ranging from weekly [107, 110, 111] over daily [58, 89, 144] to very high–frequency [59, 62, 170] have confirmed the finding of these early papers that the tails are not fat enough to be compatible with the stable Paretian hypothesis, with estimated tail indices usually somewhere in the region 2.5–5. [58] is the first to investigate the tail behavior of the euro against the US dollar, and finds that it is similar both to the German mark in the pre–euro era as well as to the yen and the British pound, with estimated exponents hovering around 3–3.5.

Concerning estimation with data at different time scales, a comparison of the results reported in the literature reveals that the impact on the estimated tail indices appears to be moderate. [59] observe an increase in the estimates when moving from 30–minute to daily returns, but they argue that these changes, due to the greater bias at the lower frequencies, are small enough to be consistent with α being invariant under time aggregation.

Note that if returns were independently distributed, their tail behavior would in fact be unaffected by time aggregation. This is a consequence of (2) along with Feller's ([80], p. 278) theorem on the convolution of regularly varying distributions, stating that any finite convolution of a regularly varying cdf $F(x)$ has a regularly varying tail with the same index. Thus, in principle, the tail survives forever, but, as long as the variance is finite, the CLT ensures that in the course of aggregation an increasing probability weight is allocated to the center of the distribution, which becomes closer to a Gaussian shape. The probability of observing a tail event will thus decrease. However, for fat–

tailed distributions, the convergence to normality can be rather slow, as reflected in the observation that pronounced nonnormalities in financial returns are often observed even at a weekly and (occasionally) monthly frequency. See [41] for an informative discussion of these issues. The fact that returns are, in general, not iid makes the interpretation of the approximate stability of the tail index estimates observed across papers employing different frequencies not so clear–cut, but Feller's theorem may nevertheless provide some insight.

There is also an extensive literature reporting tail index estimates of stock returns, mostly based on daily [2, 89, 92, 112, 113, 144, 145, 146, 177] and higher frequencies [2, 91, 92, 147, 181]. The results are comparable to those for floating exchange rates in that the tenor of this literature, which as a whole covers all major stock markets, is that most stock return series are characterized by a tail index somewhere in the region 2.5–5, and most often below 4. That is, the tails are thinner than expected under the stable Paretian hypothesis, but the finiteness of the third and in particular the fourth moments (and hence kurtosis) may already be questionable. Again, the results appear to be relatively insensitive with respect to the frequency of the data, indicating a rather slow convergence to the normal distribution. Moreover, most authors do not find significant differences between the left and the right tail, although, for stock returns, the point estimates tend to be somewhat lower for the left tail (e.g., [115, 145]).

Applications to the bond market appear to be rare, but see [201], who report tail index estimates between 2.5 and 4.5 for 5–minute and 1–hour Bund future returns and somewhat higher values for daily data.

[160] compare the tail behaviors of spot and future prices of various commodities (including cotton) and find that, while future prices resemble stock prices with tail indices in the region 2.5–4, spot prices are somewhat fatter tailed with α hovering around 2.5 and, occasionally, smaller than 2.

Summarizing, it is now a widely held view that the distribution of asset returns can typically be described as fat–tailed in the power law sense but with finite variance. Thus, currently there seems to exist a far reaching consensus that the stable Paretian model is not adequate for financial data, but see [163, 202] for a different viewpoint. A consequence of the prevalent view is that asset return distributions belong to the Gaussian domain of attraction, but that the convergence appears to be very slow.

To illustrate typical findings as reported above, let us consider the S&P500 returns described in Section 2. A first informal check of the appropriateness of a power law can be obtained by means of a log–log plot of the empirical tail, i.e., if $1 - F(x) = \overline{F}(x) \approx cx^{-\alpha}$ for large x, then a plot of the log of the empirical complementary cdf, $\overline{F}(x)$, against $\log x$ should display a linear behavior in its outer part. For the data at hand, such a plot is shown in the bottom right panel of Figure 1. Assuming homogeneity across the tails, we pool negative and positive returns by first removing the sample mean and then taking absolute values. We have also multiplied (1) by 100, so that the returns are interpretable in terms of percentage changes. The plot suggests that a power law regime may be starting from approximately the 90% quantile. Included in Figure 1 is also a regression line ("fit") fitted to the log–tail using the 500 upper (absolute) return observations. This yields, as a rough estimate for the tail index, a slope of $\widehat{\alpha} = 3.13$, with a coefficient of determination $R^2 = 0.99$. A Hill plot for $k \leq 500$ in (11) is shown in

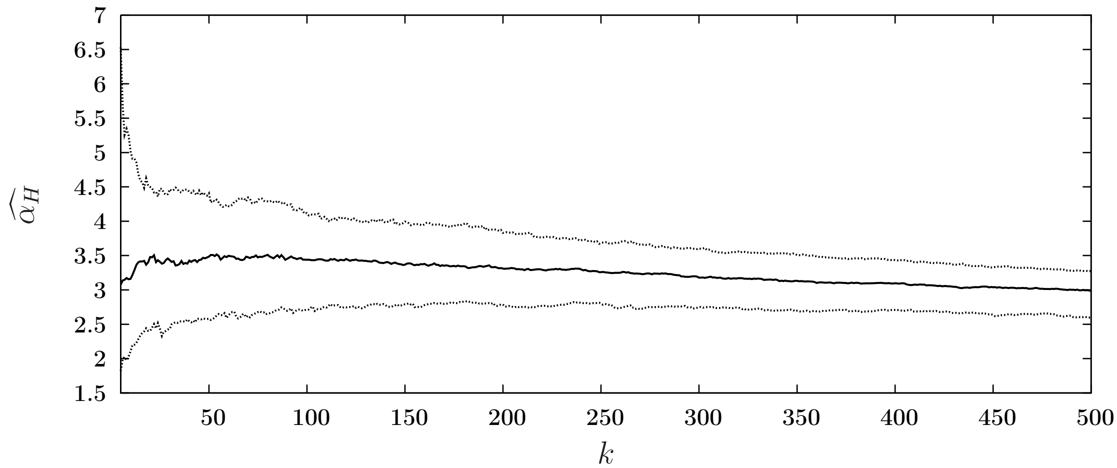

Figure 2: Shown are, for $k \leq 500$, the 95%, 50%, and 5% quantiles of the distribution of the Hill estimator $\widehat{\alpha}_{k,n}$, as defined in (11), over 176 stocks included in the S&P500 stock index.

the bottom left panel of Figure 1. The estimates are rather stable over the entire region and suggest an α somewhere in the interval (3,3.5), which is reconcilable with the results in the literature summarized above. A somewhat broader picture can be obtained by considering individual stocks. Here we consider the 176 stocks that were listed in the S&P500 from January 1985 to December 2006. Figure 2 presents, for each $k \leq 500$, the 5%, 50%, and 95% quantiles of the distribution of (11) over the different stocks. The median is close to 3 throughout, and it appears that an estimate in $(2.5, 4.5)$ would be reasonable for most stocks.

At this point, it may be useful to note that the issue is not whether a power law is true in the strict sense but only if it provides a reasonable approximation in the relevant tail region. For example, it might be argued that asset returns actually have finite support, implying finiteness of all moments and hence inappropriateness of a Pareto–type tail. However, as concisely pointed out in [144], "saying that the support of an empirical distribution is bounded says very little about the nature of outlier activity that may occur in the data".

We clearly cannot expect to identify the "true" distribution of financial variables. For example, [153] have demonstrated that by standard techniques of EVT it is virtually impossible, even in rather large samples, to discriminate between a power law and a stretched exponential (8) with a small value of b, thus questioning, for example, the conclusiveness of studies relying on the block maxima method, as referred to above. A similar point was made in [137], who showed by simulation that a three–factor stochastic volatility model, with a marginal distribution known to have all its moments finite, can generate apparent power laws in practically relevant sample sizes. As put forward in [152], "for most practical applications, the relevant question is not to determine what is the true asymptotic tail, but what is the best effective description of the tails in the domain of useful applications".

As is evident in Figure 1, a power law may (and often does) provide a useful approx-

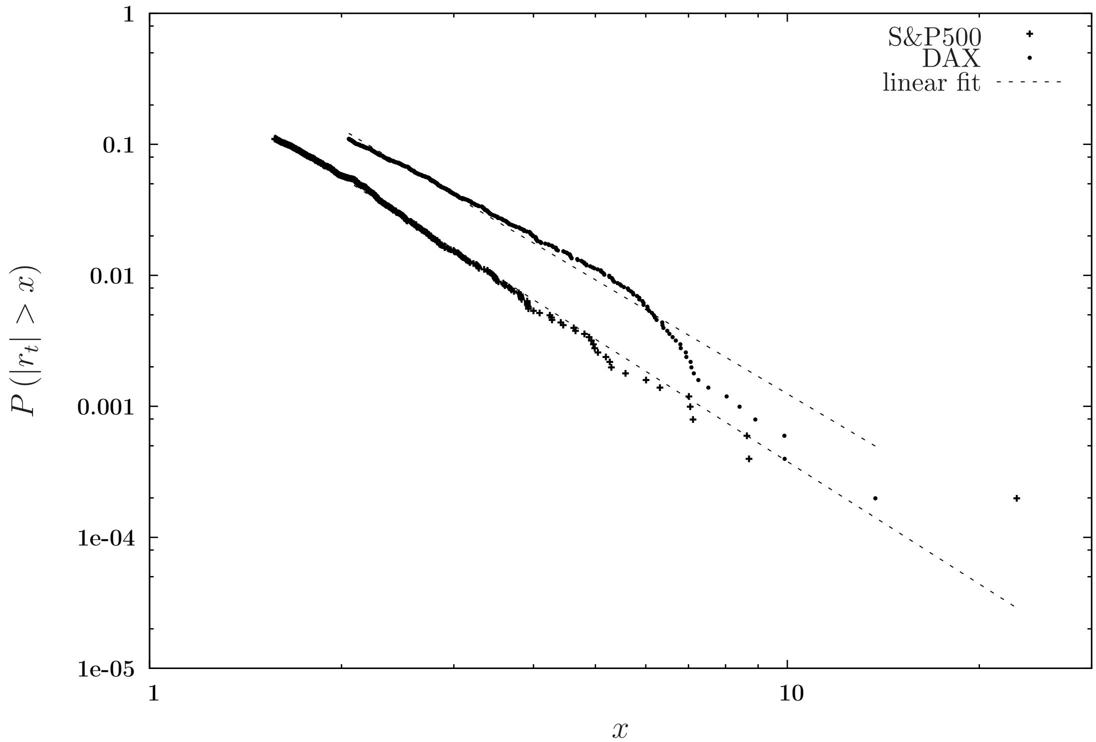

Figure 3: The figure shows, on a log-log scale, the complementary cdf, $\overline{F}(x)$, for the largest 500 absolute return observations both for the daily S&P500 returns from January 1985 to December 2006 and the daily DAX returns from July 1987 to July 2007.

imation to the tail behavior of actual data, but there is no reason to expect that it will appear in every market, and a broad range of heavy and semi–heavy tailed distributions (such as the GH in Section 6.2) may provide adequate fit. For instance, [93] investigate the tail behavior of high–frequency returns of one of the most frequently traded stocks on the Paris Stock Exchange (Alcatel) and conclude that the tails decay at an exponential rate, and [197] and [119] obtain similar results for daily returns of the Nikkei 225 index and various individual US stocks, respectively. As a further illustration, without rigorous statistical testing, Figure 3 shows the log–log tail plot for daily returns of the German stock market index DAX from July 3, 1987 to July 4, 2007 for the largest 500 out of 5218 (absolute) return observations, along with a regression–based linear fit. For purpose of comparison, the corresponding figure for the S&P500 has also been reproduced from Figure 1. While the slopes of the fitted power laws exhibit an astonishing similarity (in fact, the estimated tail index of the DAX is 2.93), it is clear from Figure 3 that an asymptotic power law, although not necessarily inconsistent with the data, is much less self–evident for the DAX than for the S&P500, due to the apparent curvature particularly in the more extreme tail.

It is finally worthwhile to mention that financial theory in general, although some of its models are built on the *assumption* of a specific distribution, has little to say about the distribution of financial variables. For example, according to the efficient markets paradigm, asset prices change in response to the arrival of relevant new information, and, consequently, the distributional properties of returns will essentially reflect those of the news process. As noted by [148], an exception to this rule is the model of rational bubbles of [35]. [148] show that this class of processes gives rise to an (approximate) power law for the return distribution. However, the structure of the model, i.e., the assumption of rational expectations, restricts the tail exponent to be below unity, which is incompatible with observed tail behaviors.

More recently, prompted by the observation that estimated tail indices are often located in a relatively narrow interval around 3, [84, 85, 83] have developed a model to explain a hypothesized "inverse cubic law for the distribution of stock price variations" [91], valid for highly developed economies, i.e., a power law tail with index $\alpha = 3$. This model is based on Zipf's law for the size of large institutional investors and the hypothesis that large price movements are generated by the trades of large market participants via a square–root price impact of volume, V, i.e., $r \cong h\sqrt{V}$, where r is the log return and h is a constant. Putting these together with a model for profit maximizing large funds, which have to balance between trading on a perceived mispricing and the price impact of their actions, leads to a power law distribution of volume with tail index 1.5, which by the square–root price impact function and simple power law accounting then produces the "cubic law". See [78, 182] for a discussion of this model and the validity of its assumptions. In a somewhat similar spirit, [161] find strong evidence for *exponentially* decaying tails of daily Indian stock returns and speculate about a general inverse relationship between the stage of development of an economy and the closeness to Gaussianity of its stock markets, but it is clear that this is really just a speculation.

6 Some Specific Distributions

6.1 Alpha Stable and Related Distributions

As noted in Section 5, the history of heavy tailed distributions in finance has its origin in the *alpha stable* model proposed by Mandelbrot [154, 155]. Being the first alternative to the Gaussian law, the alpha stable distribution has a long history in financial economics and econometrics, resulting in a large number of books and review articles.

Apart from its good empirical fit the stable distribution has also some attractive theoretical properties such as the stability property and domains of attraction. The stability property states that the index of stability (or shape parameter) remains the same under scaling and addition of different stable rv with the same shape parameter. The concept of domains of attraction is related to a generalized CLT. More specifically, dropping the assumption of a finite variance in the classical CLT, the domains of attraction states, loosely speaking, that the alpha stable distribution is the only possible limit distribution. For a more detailed discussion of this concept we refer to [169], who also provide an overview over alternative stable schemes. While the fat–tailedness of the alpha stable distributions makes it already an attractive candidate for modeling financial returns, the concept of the domains of attraction provides a further argument for its use in finance, as under the relaxation of the assumption of a finite variance of the continuously arriving return innovations the resulting return distribution at lower frequencies is generally an alpha stable distribution.

Although the alpha stable distribution is well established in financial economics and econometrics, there still exists some confusion about the naming convention and parameterization. Popular terms for the alpha stable distribution are the *stable Paretian*, *Lévy stable* or simply *stable* laws. The parameterization of the distribution in turn varies mostly with its application. For instance, to numerically integrate the characteristic function, it is preferable to have a continuous parameterization in all parameters.

The numerical integration of the alpha stable distributions is important, since with the exception of a few special cases, its pdf is unavailable in closed form. However, the characteristic function of the standard parameterization is given by

$$\mathbb{E}\left[\exp\left(itX\right)\right] = \begin{cases} \exp\left(-c^{\alpha}\left|t\right|^{\alpha}\left(1-i\beta\operatorname{sign}\left(t\right)\tan\left(\frac{\pi\alpha}{2}\right)\right)+i\tau t\right) & \alpha\neq 1 \\ \exp\left(-c\left|t\right|\left(1+i\beta\frac{2}{\pi}\operatorname{sign}\left(t\right)\ln\left(\left|t\right|\right)\right)+i\tau t\right) & \alpha=1, \end{cases} \tag{12}$$

where $\operatorname{sign}(\cdot)$ denotes the sign function, which is defined as

$$\operatorname{sign}(x) = \begin{cases} -1 & x<0 \\ 0 & x=0 \\ 1 & x>0, \end{cases}$$

$0<\alpha\leq 2$ denotes the *shape parameter*, *characteristic exponent* or *index of stability*, $-1\leq\beta\leq 1$ is the skewness parameter, and $\tau\in\mathbb{R}$ and $c\geq 0$ are the location and scale parameters, respectively.

Figure 4 highlights the impact of the parameters α and β. β controls the skewness of the distribution. The shape parameter α controls the behavior of the tails of the distribution and therefore the degree of leptokurtosis. For $\alpha<2$ moments only up to (and excluding) α exist, and for $\alpha>1$ we have $\mathbb{E}[X]=\tau$. In general, for $\alpha\in(0,1)$ and $\beta=1$ ($\beta=-1$) the support of the distribution is the set (τ,∞) (or $(-\infty,\tau)$) rather than the whole real line. In the following we call this stable distribution with $\alpha\in(0,1)$, $\tau=0$ and $\beta=1$ the positive alpha stable distribution.

Moreover, for $\alpha<2$ the stable law has asymptotically power tails,

$$\begin{aligned} \overline{F}(x)=P(X>x) &\cong c^{\alpha}d(1+\beta)x^{-\alpha} \\ f_S(x,\alpha,\beta,c,\tau) &\cong \alpha c^{\alpha}d(1+\beta)x^{-\alpha+1} \end{aligned}$$

with $d=\sin\left(\frac{\pi\alpha}{2}\right)\Gamma(\alpha)/\pi$.

For $\alpha=2$ the stable law is equivalent to the normal law with variance $2c^2$, for $\alpha=1$ and $\beta=0$ the Cauchy distribution is obtained, and for $\alpha=1/2$, $\beta=1$ and $\tau=0$ the stable law is equivalent to the Lévy distribution, with support over the positive real line.

An additional property of the stable laws is that they are closed under convolution for constant α, i.e., for two independent alpha stable rvs $X_1\sim S(\alpha,\beta_1,c_1,\tau_1)$ and $X_2\sim S(\alpha,\beta_2,c_2,\tau_2)$ with common shape parameter α we have

$$X_1+X_2\sim S\left(\alpha,\frac{\beta_1c_1^{\alpha}+\beta_2c_2^{\alpha}}{c_1^{\alpha}+c_2^{\alpha}},\left(c_1^{\alpha}+c_2^{\alpha}\right)^{1/\alpha},\tau_1+\tau_2\right)$$

and

$$aX_1+b\sim\begin{cases} S(\alpha,\operatorname{sign}(a)\beta,|a|c,a\tau+b) & \alpha\neq 1 \\ S(1,\operatorname{sign}(a)\beta,|a|c,a\tau+b-\frac{2}{\pi}\beta ca\log|a|) & \alpha=1. \end{cases}$$

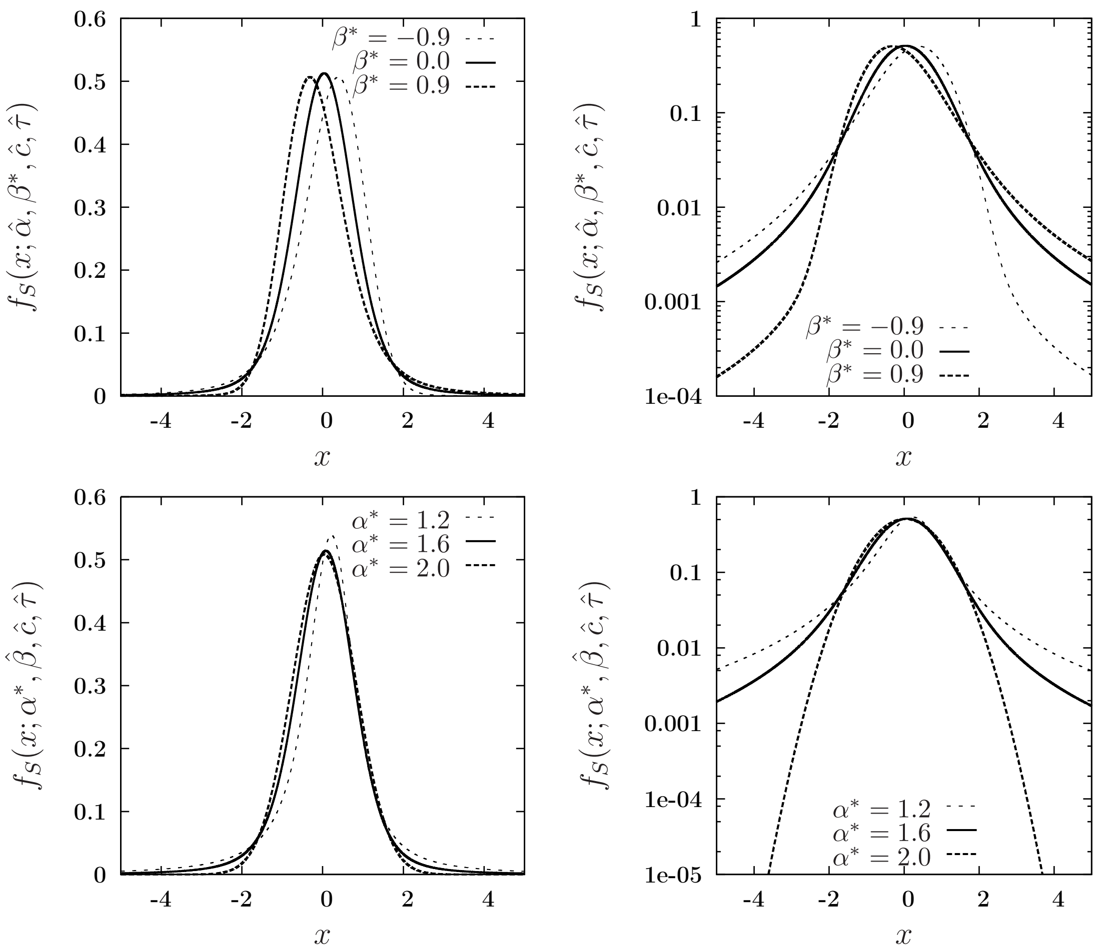

Figure 4: Density function (pdf) of the alpha stable distribution for different parameter vectors. The right panel plots the log–densities to better visualize the tail behavior. The upper (lower) section present the pdf for different values of β (α).

These results can be extended to n stable rvs. The closedness under convolution immediately implies the infinite divisibility of the stable law. As such every stable law corresponds to a Lévy process. A more detailed analysis of alpha stable processes in the context of Lévy processes is given in [192, 193].

The computation and estimation of the alpha stable distribution is complicated by the aforementioned non–existence of a closed form pdf. As a consequence, a number of different approximations for evaluating the density have been proposed, see e.g. [65, 175]. Based on these approximations, parameter estimation is facilitated using for example the maximum–likelihood estimator, see [66], or other estimation methods. As maximum–likelihood estimation relies on computationally demanding numerical integration methods, the early literature focused on alternative estimation methods. The most important methods include the quantile estimation suggested by [77, 162], which is still heavily applied in order to obtain starting values for more sophisticated estimation procedures, as well as the characteristic function approach proposed by [127, 131, 186]. However, based on its nice asymptotic properties and the nowadays available computational power, the maximum–likelihood approach is preferable.

Many financial applications also involve the simulation of a return series. In derivative pricing, for example, the computation of an expectation is oftentimes impossible as the financial instrument is generally a highly nonlinear function of asset returns. A common way to alleviate this problem is to apply Monte Carlo integration, which in turn requires quasi rvs drawn from the respective return distribution, i.e. the alpha stable distribution. A useful simulation algorithm for alpha stable rvs is proposed by [49], which is a generalization of the algorithm of [120] to the non–symmetric case. A random variable X distributed according to the stable law, $S(\alpha, \beta, c, \tau)$, can be generated as follows:

1. draw a rv U, uniformly distributed over the interval $(-\pi/2, \pi/2)$, and an (independent) exponential rv E with unit mean,

2. if $\alpha \neq 1$, compute

$$X = cS\frac{\sin\left(\alpha\left(U+B\right)\right)}{\cos^{1/\alpha}\left(U\right)}\left(\frac{\cos\left(U-\alpha(U+B)\right)}{E}\right)^{(1-\alpha)/\alpha} + \tau$$

where

$$\begin{aligned} B &:= \frac{\arctan\left(\beta\tan\left(\frac{\pi\alpha}{2}\right)\right)}{\alpha} \\ S &:= \left(1+\beta^2\tan^2\left(\frac{\pi\alpha}{2}\right)\right)^{1/(2\alpha)} \end{aligned}$$

for $\alpha = 1$ compute

$$X = c\frac{2}{\pi}\left(\left(\frac{\pi}{2}+\beta U\right)\tan\left(U\right) - \beta\log\left(\frac{(\pi/2)E\cos\left(U\right)}{(\pi/2)+\beta U}\right)\right) + \frac{2}{\pi}\beta c\log\left(c\right) + \tau.$$

Interestingly, for $\alpha = 2$ the algorithm collapses to the Box–Muller method [42] to generate normally distributed rvs.

As further discussed in Section 6.2, the mixing of normal distributions allows to derive interesting distributions having support over the real line and exhibiting heavier tails than the Gaussian. While generally any distribution with support over the positive real line can be used as the mixing distribution for the variance, transformations of the positive alpha stable distribution are often used in financial modeling.

In this context the symmetric alpha stable distributions have a nice representation. In particular, if $X \sim S(\alpha^*, 0, c, 0)$ and A (independent of X) is an α/α^* positive alpha stable rv with scale parameter $\cos^{\alpha^*/\alpha}\left(\frac{\pi\alpha}{2\alpha^*}\right)$ then

$$Z = A^{1/\alpha^*} X \sim S(\alpha, 0, c, 0).$$

For $\alpha^* = 2$ this property implies that every symmetric alpha stable distribution, i.e. an alpha stable distribution with $\beta = 0$, can be viewed as being conditionally normally distributed, i.e., it can be represented as a continuous variance normal mixture distribution.

Generally, the tail behavior of the resulting mixture distribution is completely determined by the (positive) tails of the variance mixing distribution. In the case of the positive alpha stable distribution this implies that the resulting symmetric stable distribution exhibits very heavy tails, in fact the second moment does not exist. As the literature is controversial on the adequacy of such heavy tails (see Section 5), transformations of the positive alpha stable distribution are oftentimes considered to weight down the tails. The method of *exponential tilting* is very useful in this context. In a general setup the exponential tilting of a rv X with respect to a rv Y (defined on the same probability space) defines a new rv $\tilde{X}$, which pdf can be written as

$$f_{\tilde{X}}(x;\theta) = f_X(x)\frac{\mathbb{E}\left[\exp\left(\theta Y\right)|X = x\right]}{\mathbb{E}\left[\exp\left(\theta Y\right)\right]},$$

where the parameter θ determines the "degree of dampening". The exponential tilting of a rv X with respect to itself, known as *Esscher transformation*, is widely used in financial economics and mathematics, see e.g. [88]. In this case the resulting pdf is given by

$$f_{\tilde{X}}(x;\theta) = \frac{\exp\left(\theta x\right)}{\mathbb{E}\left[\exp\left(\theta X\right)\right]} f_X(x) = \exp\left(\theta x - K(\theta)\right) f_X(x), \tag{13}$$

with $K(\cdot)$ denoting the cumulant generating function, $K(\theta) := \log\left(\mathbb{E}\left[\exp\left(\theta X\right)\right]\right)$.

The *tempered stable* (TS) laws are given by an application of the Esscher transform (13) to a positive alpha stable law. Note that the *Laplace transform* $\mathbb{E}\left[\exp\left(-tX\right)\right]$, $t \geq 0$, of a positive alpha stable rv is given by $\exp\left(-\delta(2t)^\alpha\right)$, where $\delta = c^\alpha/(2^\alpha \cos(\frac{\pi\alpha}{2}))$. Thus, with $\theta = -(1/2)\gamma^{1/\alpha} \leq 0$, the pdf of the tempered stable law is given by

$$\begin{aligned} f_{TS}(x;\alpha,\delta,\gamma) &= \frac{\exp\left(-(1/2)\gamma^{1/\alpha}x\right)}{\mathbb{E}\left[\exp\left(-(1/2)\gamma^{1/\alpha}X\right)\right]} f_S(x;\alpha,1,c(\delta,\alpha),0) \\ &= \exp\left(\delta\gamma - \frac{1}{2}\gamma^{1/\alpha}x\right) f_S(x;\alpha,1,c(\delta,\alpha),0) \end{aligned}$$

with $0 < \alpha < 1$, $\delta > 0$, and $\gamma \geq 0$.

A generalization of the TS distribution was proposed by [24]. This class of *modified stable* (MS) laws can be obtained by applying the following transformation

$$f_{MS}(x, \alpha, \lambda, \delta, \gamma) = c(\alpha, \lambda, \delta, \gamma) x^{\lambda+\alpha} f_{TS}(x; \alpha, \delta, \gamma), \tag{14}$$

where $\lambda \in \mathbb{R}$, $\gamma \vee (-\lambda) > 0$ and $c(\alpha, \lambda, \delta, \gamma)$ being a norming constant. For a more detailed analysis, we refer to [24]. Note that the terms "modified stable" or "tempered stable distribution" are not unique in the literature. Very often the so called truncated Lévy flights/processes (see for example [56, 130, 157]) are also called TS processes (or corresponding distributions). These distributions are obtained by applying a smooth downweighting of the large jumps (in terms of the Lévy density).

The MS distribution is a quite flexible distribution defined over the positive real line and nests several important distributions. For instance, for $\alpha = 1/2$, the MS distribution reduces to the *generalized inverse Gaussian* (GIG) distribution, which is of main interest in Section 6.2.

Note that in contrast to the unrestricted MS distribution, the pdf of the GIG distribution is available in closed form and can be straightforwardly obtained by applying the above transformation. In particular, for $\alpha = 1/2$, the positive alpha stable distribution is the Lévy distribution with closed form pdf given by

$$f_S(x; 1/2, 1, c, 0) = \sqrt{\frac{c}{2\pi}} \frac{\exp\left(-\frac{c}{2x}\right)}{x^{3/2}}.$$

Applying the Esscher transformation (13) with $\lambda = -(1/2)\gamma^2$ yields the pdf of the inverse Gaussian (or Wald) distribution

$$f_{IG}(x; \delta, \gamma) = \frac{\delta}{\sqrt{2\pi}} x^{-3/2} \exp\left(\delta\gamma - (\delta^2 x^{-1} + \gamma^2 x)/2\right),$$

where $\delta > 0$ and $\gamma \geq 0$. Applying the transformation (14) delivers the pdf of the GIG distribution,

$$f_{GIG}(x; \lambda, \delta, \gamma) = \frac{(\gamma/\delta)^\lambda}{2K_\lambda(\delta\gamma)} x^{\lambda-1} \exp\left(-\left(\delta^2 x^{-1} + \gamma^2 x\right)/2\right), \tag{15}$$

with $K_\lambda(\cdot)$ being the *modified Bessel function of the third kind* and of order $\lambda \in \mathbb{R}$. Note that this function is oftentimes called the modified Bessel function of the second kind or Macdonald function. Nevertheless, one representation is given by

$$K_\lambda(x) = \frac{1}{2} \int_0^\infty y^{\lambda-1} \exp\left(-\frac{1}{2} z \left(y + y^{-1}\right)\right) dy.$$

The parameters of the GIG distribution are restricted to satisfy the following conditions:

$$\begin{aligned} &\delta \geq 0 \text{ and } \gamma > 0 \quad \text{if } \lambda > 0 \\ &\delta > 0 \text{ and } \gamma > 0 \quad \text{if } \lambda = 0 \\ &\delta > 0 \text{ and } \gamma \geq 0 \quad \text{if } \lambda < 0. \end{aligned} \tag{16}$$

Importantly, in contrast to the positive alpha stable law, all moments exist and are given by

$$\mathbb{E}\left[X^r\right] = \left(\frac{\delta}{\gamma}\right)^r \frac{K_{\lambda+r}(\delta\gamma)}{K_\lambda(\delta\gamma)} \tag{17}$$

for all $r > 0$. For a very detailed analysis of the GIG distribution we refer to [117]. The GIG distribution nests several positive distributions as special cases and as limit distributions. Since all of these distributions belong to a special class of the generalized hyperbolic distribution, we proceed with a discussion of the latter, thus providing a broad framework for the discussion of many important distributions in finance.

6.2 The Generalized Hyperbolic Distribution

The mixing of normal distributions is well suited for financial modeling, as it allows to construct very flexible distributions. For example, the *normal variance–mean mixture*, given by

$$X = \mu + \beta V + \sqrt{V} Z, \tag{18}$$

with Z being normally distributed and V a positive random variable, generally exhibits heavier tails than the Gaussian distribution. Moreover, this mixture possesses interesting properties, for an overview see [22]. First, similarly to other mixtures, the normal variance–mean mixture is a conditional Gaussian distribution with conditioning on the volatility states, which is appealing when modeling financial returns. Second, if the mixing distribution, i.e. the distribution of V, is infinitely divisible, then X is likewise infinitely divisible. This implies that there exists a Lévy process with support over the whole real line, which is distributed at time $t = 1$ according to the law of X. As the theoretical properties of Lévy processes are well established in the literature (see e.g. [26, 194]), this result immediately suggests to formulate financial models in terms of the corresponding Lévy process.

Obviously, different choices for the distribution of V result in different distributions of X. However, based on the above property, an infinitely divisible distribution is most appealing. For the MS distributions discussed in Section 6.1, infinite divisibility is not yet established, although [24] strongly surmise so. However, for some special cases infinite divisibility has been shown to hold. The most popular is the GIG distribution yielding the *generalized hyperbolic* (GH) distribution for X. The latter distribution was introduced by [17] for modeling the distribution of the size of sand particles. The infinite divisibility of the GIG distribution was shown by [21].

To derive the GH distribution as a normal variance–mean mixture, let $V \sim GIG(\lambda, \delta, \gamma)$ as in (15), with $\gamma = \sqrt{\alpha^2 - \beta^2}$, and $Z \sim N(0, 1)$ independent of V. Applying (18) yields the GH distributed rv $X \sim GH(\lambda, \alpha, \beta, \mu, \delta)$ with pdf

$$
\begin{aligned}
f_{GH}(x; \lambda, \alpha, \beta, \mu, \delta) \;\; = \;\; & \frac{(\delta\gamma)^{\lambda} (\delta\alpha)^{1/2-\lambda}}{\sqrt{2\pi}\delta K_{\lambda}(\delta\gamma)} \left(1 + \frac{(x-\mu)^2}{\delta^2}\right)^{\lambda/2-1/4} \\
& K_{\lambda-1/2}\left(\alpha\delta\sqrt{1 + \frac{(x-\mu)^2}{\delta^2}}\right) \\
& \exp\left(\beta(x-\mu)\right)
\end{aligned}
$$

for $\mu \in \mathbb{R}$ and

$$
\begin{array}{ll}
\delta \geq 0 \text{ and } |\beta| < \alpha & \text{if } \lambda > 0 \\
\delta > 0 \text{ and } |\beta| < \alpha & \text{if } \lambda = 0 \\
\delta > 0 \text{ and } |\beta| \leq \alpha & \text{if } \lambda < 0,
\end{array}
$$

which are the induced parameter restrictions of the GIG distribution given in (16).

Note that, based on the mixture representation (18), the existing algorithms for generating GIG distributed rvs can be used to draw rvs from the GH distribution, see [12, 60].

For $|\beta + u| < \alpha$, the moment generating function of the GH distribution is given by

$$
\mathbb{E}\left[\exp\left(uX\right)\right] = \exp\left(\mu u\right) \left(\frac{\alpha^2 - \beta^2}{\alpha^2 - (\beta+u)^2}\right)^{\lambda/2} \frac{K_{\lambda}\left(\delta\sqrt{\alpha^2 - (\beta+u)^2}\right)}{K_{\lambda}\left(\delta\sqrt{\alpha^2 - \beta^2}\right)}. \tag{19}
$$

As the moment generating function is infinitely differentiable in the neighborhood of zero, moments of all orders exist and have been derived in [27]. In particular, the mean and the variance of a GH rv X are given by

$$
\begin{aligned}
\mathbb{E}\left[X\right] \;\; &= \;\; \mu + \frac{\beta\delta}{\sqrt{\alpha^2-\beta^2}} \frac{K_{\lambda+1}(\delta\sqrt{\alpha^2-\beta^2})}{K_{\lambda}(\delta\sqrt{\alpha^2-\beta^2})} \\
&= \;\; \mu + \beta\mathbb{E}\left[X_{GIG}\right] \\
\mathbb{V}\left[X\right] \;\; &= \;\; \frac{\delta}{\sqrt{\alpha^2-\beta^2}} \frac{K_{\lambda+1}(\delta\sqrt{\alpha^2-\beta^2})}{K_{\lambda}(\delta\sqrt{\alpha^2-\beta^2})} \\
& \quad + \frac{\beta^2\delta^2}{\alpha^2-\beta^2} \left(\frac{K_{\lambda+2}(\delta\sqrt{\alpha^2-\beta^2})}{K_{\lambda}(\delta\sqrt{\alpha^2-\beta^2}))} - \frac{K^2_{\lambda+1}(\delta\sqrt{\alpha^2-\beta^2})}{K^2_{\lambda}(\delta\sqrt{\alpha^2-\beta^2})}\right) \\
&= \;\; \mathbb{E}\left[X_{GIG}\right] + \beta^2\mathbb{V}\left[X_{GIG}\right],
\end{aligned}
$$

with $X_{GIG} \sim GIG(\lambda, \delta, \gamma)$ denoting a GIG distributed rv. Skewness and kurtosis can be derived in a similar way using the third and fourth derivative of the moment generating function (19). However, more information on the tail behavior is given by

$$
f_{GH}(x; \lambda, \alpha, \beta, \mu, \delta) \cong |x|^{\lambda-1} \exp\left((\mp\alpha + \beta)x\right),
$$

which shows that the GH distribution exhibits semi–heavy tails, see [24].

The moment generating function (19) also shows that the GH distribution is generally not closed under convolution. However, for $\lambda \in \{-1/2, 1/2\}$, the modified Bessel function of the third kind satisfies

$$K_{-\frac{1}{2}}(x) = K_{\frac{1}{2}}(x) = \sqrt{\frac{\pi}{2x}}\exp(-x),$$

yielding the following form of the moment generating function for $\lambda = -1/2$

$$\mathbb{E}\left[\exp\left(uX\right)|\lambda = -1/2\right] = \exp\left(\mu u\right)\frac{\exp\left(\delta\sqrt{\alpha^2-\beta^2}\right)}{\exp\left(\delta\sqrt{\alpha^2-(\beta+u)^2}\right)},$$

which is obviously closed under convolution. This special class of the GH distribution is called normal inverse Gaussian distribution and is discussed in more detail in Section 6.2.2. The closedness under convolution is an attractive property of this distribution as it facilitates forecasting applications.

Another very popular distribution that is nested in the GH distribution is the hyperbolic distribution given by $\lambda = 1$ (see the discussion in Section 6.2.1). Its popularity is primarily based on its pdf, which can (except for the norming constant) be expressed in terms of elementary functions allowing for a very fast numerical evaluation. However, given the increasing computer power and the slightly better performance in reproducing the unconditional return distribution, the normal inverse Gaussian distribution is now the most often used subclass.

Interestingly, further well–known distributions can be expressed as limiting cases of the GH distribution, when certain of its parameters approach their extreme values. To this end, the following alternative parametrization of the GH distribution turns out to be useful. Setting $\xi = 1/\sqrt{1+\delta\sqrt{\alpha^2-\beta^2}}$ and $\chi = \xi\beta/\alpha$ renders the two parameters invariant under location and scale transformations. This is an immediate result of the following property of the GH distribution. If $X \sim GH(\lambda, \alpha, \beta, \mu, \delta)$, then

$$a + bX \sim GH\left(\lambda, \frac{\alpha}{|b|}, \frac{\beta}{|b|}, a + b\mu, \delta|b|\right).$$

Furthermore, the parameters are restricted by $0 \leq |\chi| < \xi < 1$, implying that they are located in the so–called *shape triangle* introduced by [20]. Figure 5 highlights the impact of the parameters in the GH distribution in terms of χ, ξ and λ. Obviously, χ controls the skewness and ξ the tailedness of the distribution. The impact of λ is not so clear–cut. The lower panels in figure 5 depict the pdfs for different values of λ whereby the first two moments and the values of χ and ξ are kept constant to show the "partial" influence.

As pointed out by [71], the limit distributions can be classified by the values of ξ and χ as well as by the values ϱ and ζ of a second location and scale invariant parametrization of the GH, given by $\varrho = \beta/\alpha$ and $\zeta = \delta\sqrt{\alpha^2-\beta^2}$, as follows:

- $\boldsymbol{\xi = 1}$ **and** $\boldsymbol{-1 \leq \chi \leq 1}$**:** The resulting limit distributions depend here on the values of λ. Note, that in order to reach the boundary either $\delta \to 0$ or $|\beta| \to \alpha$.

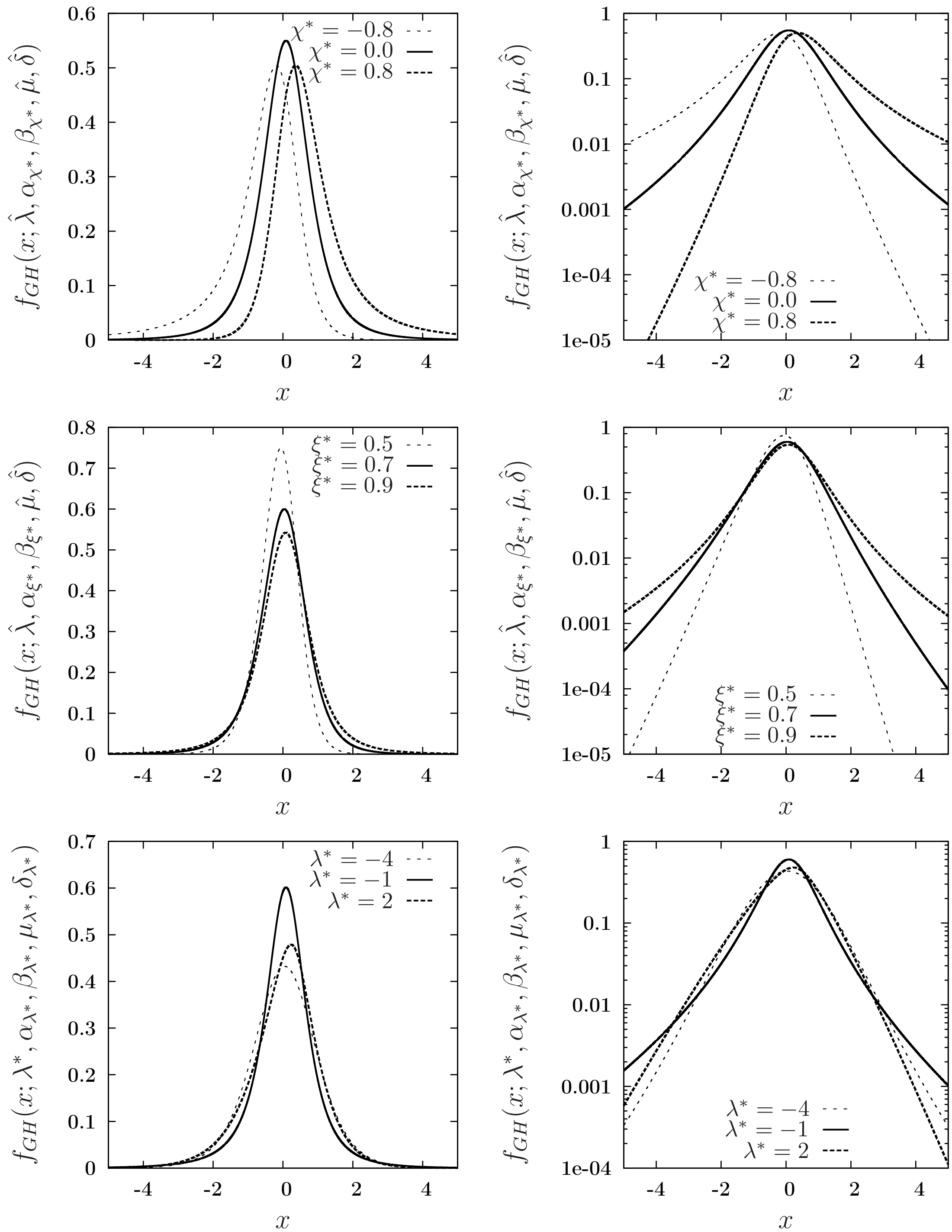

Figure 5: Density function (pdf) of the GH distribution for different parameter vectors. The right panel plots the log–densities to better visualize the tail behavior. The upper and middle section present the pdf for different values of χ and ξ. Note that these correspond to different values of α and β. The lower panel highlights the influence of λ if the the first two moments, as well as χ and ξ, are held fixed. This implies that α, β, μ and δ have to be adjusted accordingly.

- For $\underline{\lambda > 0}$ and $|\beta| \to \alpha$ no limit distribution exists, but for $\delta \to 0$ the GH distribution converges to the distribution of a variance gamma rv (see Section 6.2.4). However, note that $|\beta| < \alpha$ implies $|\chi| < \xi$ and so the limit distribution is not valid in the corners. For these cases, the limit distributions are given by $\xi = |\chi|$ and $0 < \xi \le 1$, i.e. the next case.
- For $\underline{\lambda = 0}$ there exists no proper limit distribution.
- For $\underline{\lambda < 0}$ and $\delta \to 0$ no proper distribution exists but for $\beta \to \pm\alpha$ the limit distribution is given in [185] with pdf

$$\frac{2^{\lambda+1}(\delta^2 + (x-\mu)^2)^{(\lambda-1/2)/2}}{\sqrt{2\pi}\Gamma(-\lambda)\delta^{2\lambda}\alpha^{\lambda-1/2}} K_{\lambda-1/2}\left(\alpha\sqrt{\delta^2 + (x-\mu)^2}\right) \exp\left(\pm\alpha(x-\mu)\right), \tag{20}$$

  which is the limit distribution of the corners, since $\beta = \pm\alpha$ is equivalent to $\chi = \pm\xi$. This distribution was recently called the GH skew t distribution by [1]. Assuming additionally that $\alpha \to 0$ and $\beta = \varrho\alpha \to 0$ with $\varrho \in (-1, 1)$ yields the limit distribution in between

$$\frac{\Gamma(-\lambda + 1/2)}{\sqrt{\pi\delta^2}\Gamma(-\lambda)}\left(1 + \frac{(x-\mu)^2}{\delta^2}\right)^{\lambda-1/2},$$

  which is the scaled and shifted Student's t distribution with -2λ degrees of freedom, expectation μ and variance $4\lambda^2\nu/(\nu-2)$, for more details see Secion 6.2.3.

- $\boldsymbol{\xi = |\chi|}$ **and** $\boldsymbol{0 < \xi \le 1}$**:** Except for the upper corner the limit distribution of the right boundary can be derived for

$$\beta = \alpha - \frac{\phi}{2}; \quad \alpha \to \infty; \quad \delta \to 0; \quad \alpha\delta^2 \to \tau$$

  with $\phi > 0$ and is given by the μ-shifted GIG distribution $GIG(\lambda, \sqrt{\tau}, \sqrt{\phi})$. The distribution for the left boundary is the same distribution but mirrored at $x = 0$. Note that the limit behavior does not depend on λ. However, to obtain the limit distributions for the left and right upper corners we have to distinguish for different values of λ. Recall that for the regime $\xi = 1$ and $-1 \le \chi \le 1$ the derivation was not possible.

  - For $\underline{\lambda > 0}$ the limit distribution is a gamma distribution.
  - For $\underline{\lambda = 0}$ no limit distribution exists.
  - For $\underline{\lambda < 0}$ the limit distribution is the reciprocal gamma distribution.

- $\boldsymbol{\xi = \chi = 0}$**:** This is the case for $\alpha \to \infty$ or $\delta \to \infty$. If only $\alpha \to \infty$ then the limit distribution is the Dirac measure concentrated in μ. If in addition $\delta \to \infty$ and $\delta/\alpha \to \sigma^2$ then the limit distribution is a normal distribution with mean $\mu + \beta\sigma^2$ and variance σ^2.

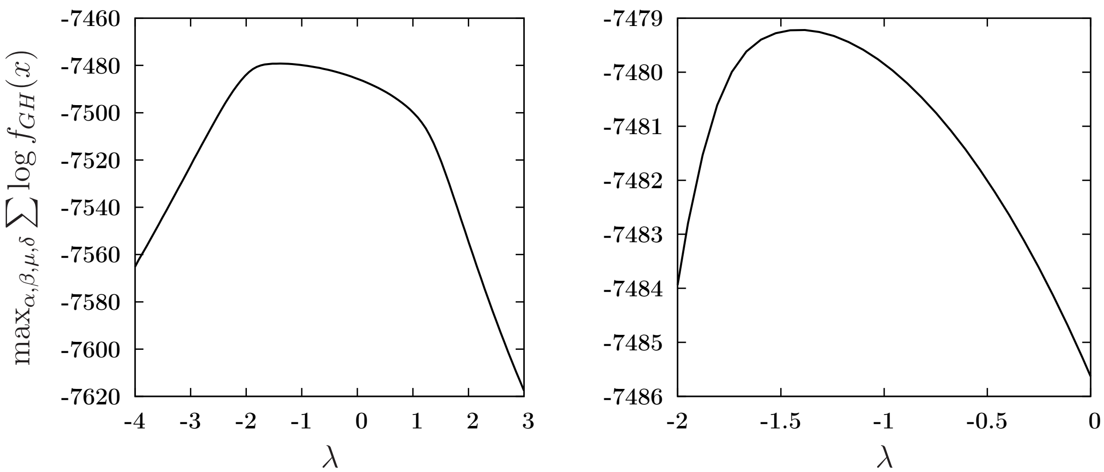

Figure 6: Partially maximized log likelihood, estimated maximum log likelihood values of the GH distribution for different values of λ.

As pointed out by [185] applying the unrestricted GH distribution to financial data results in a very flat likelihood function especially for λ. This characteristic is illustrated in Figure 6, which plots the maximum of the log likelihood for different values of λ using our sample of the S&P500 index returns. This implies that the estimate of λ is generally associated with a high standard deviation. As a consequence, rather than using the GH distribution directly, the finance literature primarily predetermines the value of λ, resulting in specific subclasses of the GH distribution, which are discussed in the sequel. However, it is still interesting to derive the general results in terms of the GH distribution (or the corresponding Lévy process) directly and to restrict only the empirical application to a subclass. For example [191] derived a diffusion process with GH marginal distribution, which is a generalization of the result of [33], who proposed a diffusion process with hyperbolic marginal distribution.

6.2.1 The Hyperbolic Distribution

Recall, that the *hyperbolic* (HYP) distribution can be obtained as a special case of the GH distribution by setting $\lambda = 1$. Thus, all properties of the GH law can be applied to the HYP case. For instance the pdf of the HYP distribution is straightforwardly given by (19) setting $\lambda = 1$

$$\begin{aligned} f_H(x;\alpha,\beta,\mu,\delta) &:= f_{GH}(x;1,\alpha,\beta,\mu,\delta) \\ &= \frac{\sqrt{\alpha^2-\beta^2}}{2\alpha\delta K_1(\delta\sqrt{\alpha^2-\beta^2})}\, e^{-\alpha\sqrt{\delta^2+(x-\mu)^2}+\beta(x-\mu)}, \end{aligned} \tag{21}$$

where $0 \leq |\beta| < \alpha$ are the shape parameter and $\mu \in \mathbb{R}$ and $\delta > 0$ are the location and scale parmeter, respectively.

The distribution was applied to stock return data by [69, 70, 132] while [33] derived a diffusion model with marginal distribution belonging to the class of HYP distributions.

6.2.2 The Normal Inverse Gaussian Distribution

The *normal inverse Gaussian* (NIG) distribution is given by the GH distribution with $\lambda = -1/2$ and has the following pdf

$$\begin{aligned} f_{NIG}(x;\alpha,\beta,\mu,\delta) &:= f_{GH}(x;-\frac{1}{2},\alpha,\beta,\mu,\delta) \\ &= \frac{\alpha\delta K_1\left(\alpha\sqrt{\delta^2+(x-\mu)^2}\right)}{\pi\sqrt{\delta^2+(x-\mu)^2}}\, e^{\delta\gamma+\beta(x-\mu)} \end{aligned} \tag{22}$$

with $0 < |\beta| \le \alpha$, $\delta > 0$ and $\mu \in \mathbb{R}$. The moments of a NIG distributed rv can be obtained from the moment generating function of the GH distribution (19) and are given by

$$\mathbb{E}[X] = \mu + \frac{\delta\beta}{\sqrt{\alpha^2-\beta^2}} \quad \text{and} \quad \mathbb{V}[X] = \frac{\delta\alpha^2}{\sqrt{\alpha^2-\beta^2}^3}$$

$$\mathbb{S}[X] = 3\frac{\beta}{\alpha\sqrt{\delta\sqrt{\alpha^2-\beta^2}}} \quad \text{and} \quad \mathbb{K}[X] = 3\frac{\alpha^2+4\beta^2}{\delta\alpha^2\sqrt{\alpha^2-\beta^2}}.$$

This distribution was heavily applied in financial economics for modeling the unconditional as well as the conditional return distribution, see e.g. [19, 23, 185]; as well as [10, 18, 114], respectively. Recently, [57] used the NIG distribution for modeling realized variance and found improved forecast performance relative to a Gaussian model. A more realistic modeling of the distributional properties is not only important for risk management or forecasting, but also for statistical inference. For example the efficient method of moments, proposed by [87] requires the availability of an highly accurate auxiliary model, which provide the objective function to estimate a more structural model. Recently, [39] provide such an auxiliary model, which uses the NIG distribution and realized variance measures.

Recall that for $\lambda = -1/2$ the mixing distribution is the inverse Gaussian distribution, which facilitates the generation of rvs. Hence, rvs with NIG distribution can be generated in the following way

1. draw a chi-square distributed rv C with one degree of freedom and a uniformly distributed rv over the intervall $(0, 1)$ U

2. compute
$$X_1 = \frac{\delta}{\gamma} + \frac{1}{2\delta\gamma}\left(\frac{\delta C}{\gamma} - \sqrt{4\delta^3 C/\gamma + \delta^2 C^2/\gamma^2}\right)$$

3. if $U < \delta/(\gamma(\delta/\gamma + X_1))$ return X_1 else return $\delta^2/(\gamma^2 X_1)$.

As pointed out by [187] the main difference between the HYP and NIG distribution: “Hyperbolic log densities, being hyperbolas, are strictly concave everywhere. Therefore they cannot form any sharp tips near $x = 0$ without loosing too much mass in the tails ... In contrast, NIG log densities are concave only in an intervall around $x = 0$, and convex in the tails.” Moreover, [18] concludes, “It is, moreover, rather typical that asset returns exhibit tail behaviour that is somewhat heavier than log linear, and this further strengthens the case for the NIG in the financial context”.

6.2.3 The Student t distribution

Next to the alpha stable distribution *Student's t* (t thereafter) distribution has the longest history in financial economics. One reason is that although the nonnormality of asset returns is widely accepted, there still exists some discussion on the exact tail behavior. While the alpha stable distribution implies extremely slowly decreasing tails for $\alpha \neq 2$, the t distribution exhibits power tails and existing moments up to (and excluding) ν. As such, the t distribution might be regarded as the strongest competitor to the alpha stable distribution, shedding also more light on the empirical tail behavior of returns. The pdf for the scaled and shifted t distribution is given by

$$f_t(x;\nu,\mu,\sigma) = \frac{\Gamma((\nu+1)/2)}{\sqrt{\nu\pi}\Gamma(\nu/2)\sigma}\left(1+\frac{1}{\nu}\left(\frac{x-\mu}{\sigma}\right)^2\right)^{-(\nu+1)/2} \tag{23}$$

for $\nu > 0$, $\sigma > 0$ and $\mu \in \mathbb{R}$. For $\mu = 0$ and $\sigma = 1$ the well–known standard t distribution is obtained. The shifted and scaled t distribution can also be interpreted as a mean-variance mixture (18) with a reciprocal gamma distribution as a mixing distribution. The mean, variance, and kurtosis (3) are given by μ, $\sigma^2\nu/(\nu-2)$, and $3(\nu-2)/(\nu-4)$, provided that $\nu > 1$, $\nu > 2$, and $\nu > 4$, respectively. The tail behavior is

$$f_t(x,\nu,\mu,\sigma) \cong cx^{-\nu-1}.$$

The t distribution is one of the standard nonnormal distribution in financial economics, see e.g. [36, 38, 184]. However, as the unconditional return distribution may exhibit skewness, a skewed version of the t distribution might be more adequate in some cases. In fact, several skewed t distributions were proposed in the literature, for a short overview see [1]. The following special form of the pdf was considered in [81, 102]

$$\begin{aligned} f_{t,FS}(x;\nu,\mu,\sigma,\beta) &= \frac{2\beta}{\beta^2+1}\frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\Gamma\left(\nu/2\right)\sqrt{\pi\nu}\sigma} \\ &\quad \left(1+\frac{1}{\nu}\left(\frac{x-\mu}{\sigma}\right)^2\left(\frac{1}{\beta^2}\mathcal{I}(X \geq \mu)+\beta^2\mathcal{I}(x<\mu)\right)\right)^{-\frac{\nu+1}{2}} \end{aligned}$$

with $\beta > 0$. For $\beta = 1$ the pdf reduces to the pdf of the usual symmetric scaled and shifted t distribution. Another skewed t distribution was proposed by [116] with pdf

$$\begin{aligned} f_{t,JF}(x,\nu,\mu,\sigma,\beta) &= \frac{\Gamma(\nu+\beta)2^{1-\nu-\beta}}{\Gamma(\nu/2)\Gamma(\nu/2+\beta)\sqrt{\nu+\beta}\sigma}\left(1+\frac{\frac{x-\mu}{\sigma}}{\sqrt{\nu+\beta+\left(\frac{x-\mu}{\sigma}\right)^2}}\right)^{(\nu+1)/2} \\ &\quad \left(1-\frac{\frac{x-\mu}{\sigma}}{\sqrt{\nu+\beta+\left(\frac{x-\mu}{\sigma}\right)^2}}\right)^{\beta+(\nu+1)/2} \end{aligned}$$

for $\beta > -\nu/2$. Again, the usual t distribution can be obtained as a special case for $\beta = 0$. A skewed t distribution in terms of the pdf and cdf of the standard t distribution

$f_t(x;\nu,0,1)$ and $F_t(x;\nu,0,1)$ is given by [13, 43]

$$f_{t,AC}(x;\nu,\mu,\sigma,\beta) = \frac{2}{\sigma} f_t\left(\frac{x-\mu}{\sigma},\nu,0,1\right) F_t\left(\beta\left(\frac{x-\mu}{\sigma}\right)\sqrt{\frac{\nu+1}{\nu+\left(\frac{x-\mu}{\sigma}\right)^2}},\nu+1,0,1\right)$$

for $\beta \in \mathbb{R}$.

Alternatively, a skewed t distribution can also be obtained as a limit distribution of the GH distribution. Recall that for $\lambda < 0$ and $\beta \to \alpha$ the limit distribution is given by (20) as

$$\begin{aligned} f_{t,GH}(x;\lambda,\mu,\delta,\alpha) &= \frac{2^{\lambda+1}(\delta^2+(x-\mu)^2)^{(\lambda-1/2)/2}}{\sqrt{2\pi}\Gamma(-\lambda)\delta^{2\lambda}\alpha^{\lambda-1/2}} \\ &\quad K_{\lambda-1/2}\left(\alpha\sqrt{\delta^2+(x-\mu)^2}\right)\exp\left(\alpha(x-\mu)\right). \end{aligned}$$

for $\alpha \in \mathbb{R}$. The symmetric t distribution is obtained for $\alpha \to 0$. The distribution was introduced by [185] and a more detailed examination was recently given in [1].

6.2.4 The Variance Gamma Distribution

The *variance gamma* (VG) distribution can be obtained as mean–variance mixture with gamma mixing distribution. Note that the gamma distribution is obtained in the limit from the GIG distribution for $\lambda > 0$ and $\delta \to 0$. The pdf of the VG distribution is given by

$$\begin{aligned} f_{VG}(x;\mu,\alpha,\beta,\lambda) &:= \lim_{\delta\to 0} f_{GH}(x;\lambda,\alpha,\beta,\delta,\mu) \\ &= \frac{\gamma^{2\lambda}|x-\mu|^{\lambda-1/2}K_{\lambda-1/2}\left(\alpha|x-\mu|\right)}{\sqrt{\pi}\Gamma(\lambda)(2\alpha)^{\lambda-1/2}}\, e^{\beta(x-\mu)}. \end{aligned} \tag{24}$$

Note, the usual parameterization of the VG distribution

$$\begin{aligned} f^*_{VG}(x;\sigma^*,\theta^*,\nu^*,\mu^*) &= \frac{2\exp\left(\theta^*(x-\mu^*)/\sigma^{*2}\right)}{\nu^{*1/\nu^*}\sqrt{2\pi\sigma^{*2}}\Gamma(1/\nu^*)}\left(\frac{(x-\mu^*)^2}{2\sigma^{*2}/\nu^*+\theta^{*2}}\right)^{\frac{1}{2\nu^*}-\frac{1}{4}} \\ &\quad K_{\frac{1}{\nu^*}-\frac{1}{2}}\left(\frac{\sqrt{(x-\mu^*)^2(2\sigma^{*2}/\nu^*+\theta^{*2})}}{\sigma^{*2}}\right) \end{aligned}$$

is different from the one used here, however the parameters can be transformed between these repesentation in the following way

$$\sigma^* = \sqrt{\frac{2\lambda}{\alpha^2-\beta^2}}; \quad \theta^* = \frac{2\beta\lambda}{\alpha^2-\beta^2}; \quad \nu^* = \frac{1}{\lambda}; \quad \mu^* = \mu.$$

This distribution was introduced by [149, 150, 151]. For $\lambda = 1$ (the HYP case) we obtain a skewed, shifted and scaled Laplace distribution with pdf

$$\begin{aligned} f_L(x;\alpha,\beta,\mu) &:= \lim_{\delta\to 0} f_{GH}(x;1,\alpha,\beta,\delta,\mu) \\ &= \frac{\alpha^2-\beta^2}{2\alpha}\exp\left(-\alpha|x-\mu|+\beta(x-\mu)\right). \end{aligned}$$

A generalization of the VG distribution to the so called CGMY distribution was proposed by [48].

6.2.5 The Cauchy distribution

Setting $\lambda = -1/2$, $\beta \to 0$ and $\alpha \to 0$ the GH distribution converges to the *Cauchy* distribution with parameters μ and δ. Since the Cauchy distribution belongs to the class of symmetric alpha stable ($\alpha = 1$) and symmetric t distributions ($\nu = 1$) we refer to Section 6.1 and 6.2.3 for a more detailed discussion.

6.2.6 The Normal Distribution

For $\alpha \to \infty$, $\beta = 0$ and $\delta = 2\sigma^2$ the GH distribution converges to the *normal* distribution with mean μ and variance σ^2.

6.3 Finite Mixtures of Normal Distributions

The density of a (finite) mixture of k normal distributions is given by a linear combination of k Gaussian *component densities*, i.e.,

$$f_{NM}(x;\theta) = \sum_{j=1}^{k} \lambda_j \phi(x;\mu_j,\sigma_j^2), \quad \phi(x;\mu,\sigma^2) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right), \tag{25}$$

where $\theta = (\lambda_1, \ldots, \lambda_{k-1}, \mu_1, \ldots, \mu_k, \sigma_1^2, \ldots, \sigma_k^2)$, $\lambda_k = 1 - \sum_{j=1}^{k-1} \lambda_j$, $\lambda_j > 0$, $\mu_j \in \mathbb{R}$, $\sigma_j^2 > 0$, $j = 1, \ldots, k$, and $(\mu_i, \sigma_i^2) \neq (\mu_j, \sigma_j^2)$ for $i \neq j$. In (25), the λ_j, μ_j, and σ_j^2 are called the *mixing weights*, *component means*, and *component variances*, respectively.

Finite mixtures of normal distributions have been applied as early as 1886 in [174] to model leptokurtic phenomena in astrophysics. A voluminous literature exists, see [165] for an overview. In our discussion, we shall focus on a few aspects relevant for applications in finance. In this context, (25) arises naturally when the component densities are interpreted as different *market regimes*. For example, in a two–component mixture ($k = 2$), the first component, with a relatively high mean and small variance, may be interpreted as the bull market regime, occurring with probability λ_1, whereas the second regime, with a lower expected return and a greater variance, represents the bear market. This (typical) pattern emerges for the S&P500 returns, see Table 1. Clearly (25) can be generalized to accommodate nonnormal component densities; e.g., [104] model stock returns using mixtures of generalized error distributions of the form (40). However, it may be argued that in this way much of the original appeal of (25), i.e., within–regime normality along with CLT arguments, is lost.

The moments of (25) can be inferred from those of the normal distribution, with mean and variance given by

$$\mathbb{E}[X] = \sum_{j=1}^{k} \lambda_j \mu_j, \quad \text{and} \quad \mathbb{V}[X] = \sum_{j=1}^{k} \lambda_j (\sigma_j^2 + \mu_j^2) - \left(\sum_{j=1}^{k} \lambda_j \mu_j\right)^2, \tag{26}$$

respectively. The class of finite normal mixtures is very flexible in modeling the leptokurtosis and, if existent, skewness of financial data. To illustrate the first property, consider the *scale normal mixture*, where, in (25), $\mu_1 = \mu_2 = \cdots = \mu_k := \mu$. In fact, when applied to financial return data, it is often found that the market regimes differ mainly in their variances, while the component means are rather close in value, and often their differences are not significant statistically. This reflects the observation that excess kurtosis is a much more pronounced (and ubiquitous) property of asset returns than skewness. In the scale mixture case, the density is symmetric, but with higher peaks and thicker tails than the normal with the same mean and variance. To see this, note that $\sum_j(\lambda_j/\sigma_j) > (\sum_j \lambda_j\sigma_j^2)^{-1/2} \Leftrightarrow (\sum_j \lambda_j\sigma_j^2)^{1/2} > [\sum_j(\lambda_j/\sigma_j)]^{-1}$. But $(\sum_j \lambda_j\sigma_j^2)^{1/2} > \sum_j \lambda_j\sigma_j > [\sum_j(\lambda_j/\sigma_j)]^{-1}$ by Jensen's and the arithmetic–harmonic mean inequality, respectively. This shows $f_{NM}(\mu;\theta) > \phi(\mu;\mu,\sum_j \lambda_j\sigma_j^2)$, i.e., peakedness. Tailedness follows from the observation that the difference between the mixture and the mean–variance equivalent normal density is asymptotically dominated by the component with the greatest variance. Moreover, the densities of the scale mixture and the mean–variance equivalent Gaussian intersect exactly two times on both sides of the mean, so that the scale mixture satisfies the density crossing condition in Finucan's theorem mentioned in Section 3 and observed in the center panel of Figure 1. This follows from the fact that, if $a_1,\ldots,a_n$ and $\gamma_1 < \cdots < \gamma_n$ are real constants, and N is the number of real zeros of the function $\varphi(x) = \sum_i a_i e^{\gamma_i x}$, then $W - N$ is a nonnegative even integer, where W is the number of sign changes in the sequence $a_1,\ldots,a_n$ [183]. Skewness can be incorporated into the model when the component means are allowed to differ. For example, if, in the two–component mixture, the high–variance component has both a smaller mean and mixing weight, then the distribution will be skewed to the left.

Due to their flexibility and the aforementioned economic interpretation, finite normal mixtures have been frequently used to model the unconditional distribution of asset returns [40, 44, 129, 179], and they have become rather popular since the publication of Hamilton's [101] paper on Markov–switching processes, where the mixing weights are assumed to be time–varying according to a k–state Markov chain; see, e.g., [200] for an early contribution in this direction.

However, although a finite mixture of normals is a rather flexible model, its tails decay eventually in a Gaussian manner, and therefore, according to the discussion in Section 5, it may often not be appropriate to model returns at higher frequencies unconditionally. Nevertheless, when incorporated into a GARCH structure (see Section 7), it provides a both useful and intuitively appealing framework for modeling the *conditional* distribution of asset returns, as in [5, 96, 97]. These papers also provide a discussion of alternative interpretations of the mixture model (25), as well as an overview over the extensive literature.

6.4 Empirical Comparison

In the following we empirically illustrate the adequacy of the various distributions discussed in the previous sections for modeling the unconditional return distribution. Table 1 presents the estimation results for the S&P500 index assuming iid returns. The log likelihood values clearly indicate the inadequacy of the normal, Cauchy and stable distributions. This is also highlighted in the upper panel of Figure 7, which clearly shows that the tails of the Cauchy and stable distributions are too heavy, whereas those of the normal distribution are too weak. To distinguish the other distributions in more detail, the lower left panel is an enlarged display of the shadowed box in the upper panel. It illustrates nicely that the two component mixture, VG and HYP distribution exhibit semiheavy tails, which are probably a little bit to weak for an adequate modeling as is indicated by the log likelihood values. Similarly, the two–component finite normal mixture, although much better than the normal, cannot keep up with most of the other models, presumably due to its essentially Gaussian tails. Although the pdf of the NIG distribution lies somewhere in between the pdfs of the HYP and the different t distributions, the log likelihood value clearly indicates that this distribution is in a statistical sense importantly closer to the t distributions. A further distinction between the other distributions including all kinds of t distributions and the GH distribution is nearly impossible, as can be seen from the lower right plot, which is an enlarged display of the lower left panel. The log likelihood values also do not allow for a clear distinction. Note also that the symmetric t distribution performs unexpectedly well. In particular, its log likelihood is almost indistinguishable from those of the skewed versions. Also note that, for all t distributions, the estimated tail index, ν, is close to 3.5, which is in accordance with the results from semiparametric tail estimation in Section 5.

The ranking of the distributions in terms of the log likelihood depends of course heavily on the dataset, and different return series may imply different rankings. However, Table 1 also highlights some less data–dependent results, which are more or less accepted in the literature, e.g., the tails of the Cauchy and stable distributions are too heavy, and those of the HYP and VG are too light for the unconditional distribution. This needs of course no longer be valid in a different modeling setup. Especially in a GARCH framework the conditional distribution don't need to imply such heavy tails because the model itself imposes fatter tails.

In Section 8, the comparison of the models will be continued on the basis of their ability the measure the Value–at–Risk, an important concept in risk management.

7 Volatility Clustering and Fat Tails

It has long been known that the returns of most financial assets, although close to being unpredictable, exhibit significant dependencies in measures of volatility, such as absolute or squared returns. Moreover, the empirical results based on the recent availability of more precise volatility measures, such as the *realized volatility*, which is defined as the sum over the squared intradaily high–frequency returns (see e.g. [7] and [25]), also point towards the same direction. In particular, the realized volatility has been found to exhibit strong persistence in its autocorrelation function, which shows a hyperbolic decay indicating the presence of long memory in the volatility process. In fact, this finding as well as other stylized features of the realized volatility have been observed across different data sets and markets and are therefore by now widely acknowledged and established in the literature. For a more detailed and originating discussion on the stylized facts of the high–frequency based volatility measures for stock returns and exchange returns we refer to [9] and [8], respectively.

The observed dependence of time–varying pattern of the volatility is usually referred to as *volatility clustering*. It is also apparent in the top panel of Figure 1 and was already observed by Mandelbrot [154], who noted that "'large changes tend to be followed by

Table 1: Maximum–likelihood parameter estimates of the iid model.

distribution	parameters					loglik
GH	$\hat{\lambda}$	$\hat{\mu}$	$\hat{\alpha}$	$\hat{\beta}$	$\hat{\delta}$	-7479.2
	-1.422	0.087	0.322	-0.046	1.152	
	(0.351)	(0.018)	(0.222)	(0.022)	(0.139)	
t_{GH}	$\hat{\mu}$	$\hat{\delta}$	$\hat{\lambda}$	$\hat{\alpha}$		-7479.7
	0.084	1.271	3.445	-0.041		
	(0.018)	(0.052)	(0.181)	(0.021)		
t_{JF}	$\hat{\nu}$	$\hat{\mu}$	$\hat{\sigma}$	$\hat{\beta}$		-7480.0
	3.348	0.098	0.684	0.091		
	(0.179)	(0.025)	(0.012)	(0.049)		
t_{AC}	$\hat{\nu}$	$\hat{\mu}$	$\hat{\sigma}$	$\hat{\beta}$		-7480.1
	3.433	0.130	0.687	-0.123		
	(0.180)	(0.042)	(0.013)	(0.068)		
t_{FS}	$\hat{\nu}$	$\hat{\mu}$	$\hat{\sigma}$	$\hat{\beta}$		-7480.3
	3.432	0.085	0.684	0.972		
	(0.180)	(0.020)	(0.012)	(0.017)		
symmetric t	$\hat{\nu}$	$\hat{\mu}$	$\hat{\sigma}$			-7481.7
	3.424	0.056	0.684			
	(0.179)	(0.011)	(0.012)			
NIG	$\hat{\mu}$	$\hat{\alpha}$	$\hat{\beta}$	$\hat{\delta}$		-7482.0
	0.088	0.784	-0.048	0.805		
	(0.018)	(0.043)	(0.022)	(0.028)		
HYP	$\hat{\mu}$	$\hat{\alpha}$	$\hat{\beta}$	$\hat{\delta}$		-7499.5
	0.090	1.466	-0.053	0.176		
	(0.018)	(0.028)	(0.023)	(0.043)		
VG	$\hat{\mu}$	$\hat{\alpha}$	$\hat{\beta}$	$\hat{\lambda}$		-7504.2
	0.092	1.504	-0.054	1.115		
	(0.013)	(0.048)	(0.019)	(0.054)		
alpha stable	$\hat{\alpha}$	$\hat{\beta}$	$\hat{c}$	$\hat{\tau}$		-7522.5
	1.657	-0.094	0.555	0.036		
	(0.024)	(0.049)	(0.008)	(0.015)		
finite mixture ($k = 2$)	$\hat{\lambda}_1$	$\hat{\mu}_1$	$\hat{\mu}_2$	$\hat{\sigma}_1{}^2$	$\hat{\sigma}_2{}^2$	-7580.8
	0.872	0.063	-0.132	0.544	4.978	
	(0.018)	(0.012)	(0.096)	(0.027)	(0.530)	
Cauchy	$\hat{\mu}$	$\hat{\sigma}$				-7956.6
	0.060	0.469				
	(0.010)	(0.008)				
normal	$\hat{\mu}$	$\hat{\sigma}$				-8168.9
	0.039	1.054				
	(0.014)	(0.010)				

Shown are maximum likelihood estimates for iid models with different assumptions about the distribution of the innovations. Standard errors are given in parentheses. "loglik" is the value of the maximized log likelihood function.

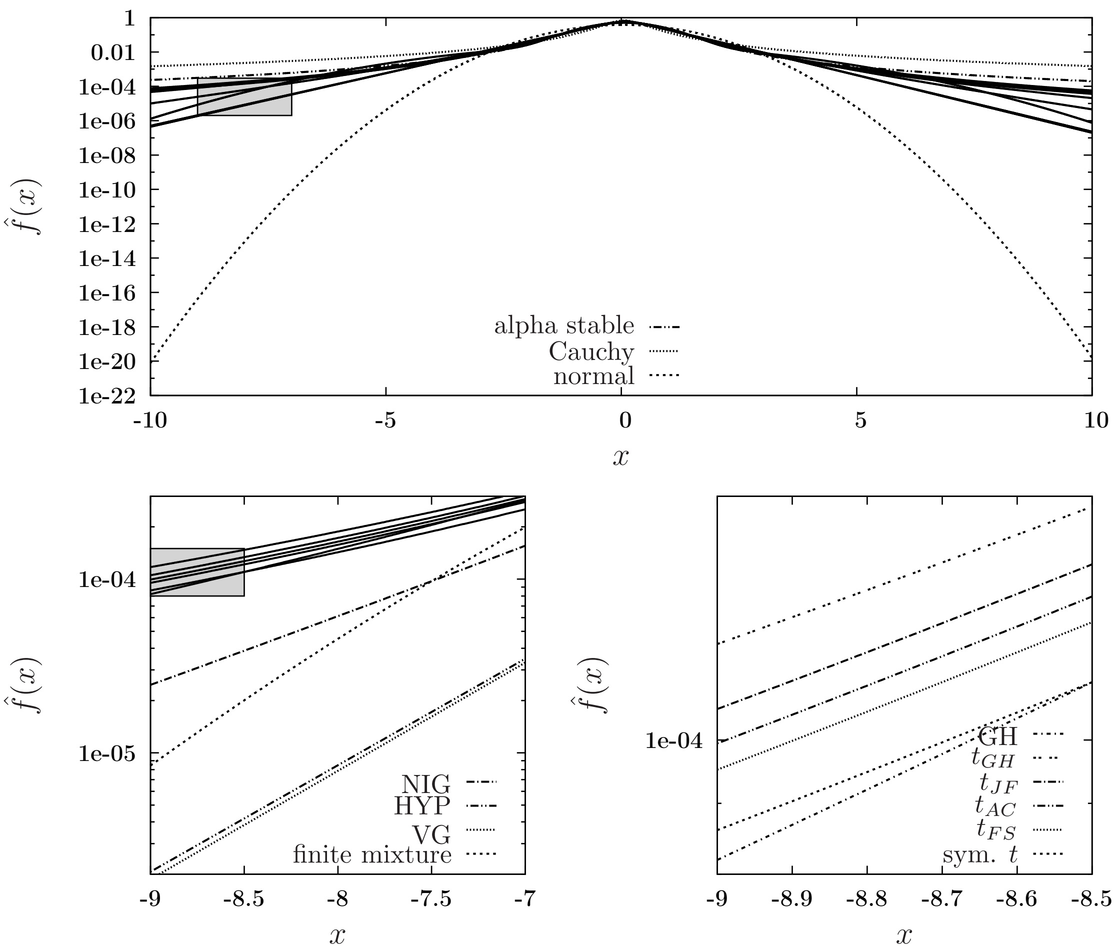

Figure 7: Plot of the estimated pdfs of the different return distributions assuming iid returns.

large changes—of either sign—and small changes tend to be followed by small changes". It is now well understood that volatility clustering can explain at least part of the fat–tailedness of the unconditional return distribution, even if the *conditional* distribution is Gaussian. This is also supported by the recent observation that if the returns are scaled by the realized volatility then the distribution of the resulting series is approximately Gaussian (see [9] and [8]). To illustrate, consider a time series $\{\epsilon_t\}$ of the form

$$\epsilon_t = \eta_t \sigma_t, \tag{27}$$

where $\{\eta_t\}$ is an iid sequence with mean zero and unit variance, with η_t being independent of σ_t, so that σ_t^2 is the conditional variance of ϵ_t. With respect to the kurtosis measure $\mathbb{K}$ in (3), it has been observed by [108], and earlier by [31] in a different context, that, as long as σ_t^2 is not constant, Jensen's inequality implies $\mathbb{E}\left[\epsilon_t^4\right] = \mathbb{E}\left[\eta_t^4\right]\mathbb{E}\left[\sigma_t^4\right] > \mathbb{E}\left[\eta_t^4\right]\mathbb{E}^2[\sigma_t^2]$, so that the kurtosis of the unconditional distribution exceeds that of the innovation process. Clearly $\mathbb{K}$ provides only limited information about the actual shape of the distribution, and more meaningful results can be obtained by specifying the dynamics of the conditional variance, σ_t^2. A general useful result [166] for analyzing the tail behavior of processes such as (27) is that, if ξ_t and σ_t are independent nonnegative random variables with σ_t regularly varying, i.e., $P(\sigma_t > x) = L(x)x^{-\alpha}$ for some slowly varying L, and $\mathbb{E}\left[\xi_t^{\alpha+\delta}\right] < \infty$ for some $\delta > 0$, then $\xi_t\sigma_t$ is likewise regularly varying with tail index α, namely,

$$P(\xi_t\sigma_t > x) \cong \mathbb{E}\left[\xi^\alpha\right] P(\sigma_t > x) \quad \text{as } x \to \infty. \tag{28}$$

Arguably the most popular model for the evolution of σ_t^2 in (27) is the generalized autoregressive conditional heteroskedasticity process of orders p and q, or GARCH(p,q), as introduced by [73, 37], which specifies the conditional variance as

$$\sigma_t^2 = \alpha_0 + \sum_{i=1}^{q} \alpha_i \epsilon_{t-i}^2 + \sum_{i=1}^{p} \beta_i \sigma_{t-i}^2. \tag{29}$$

The case $p = 0$ in (29) is referred to as an ARCH(q) process, which is the specification considered in [73]. To make sure that the conditional variance remains positive for all t, appropriate restrictions have to be imposed on the parameters in (29), i.e., α_i, $i = 0, \ldots, q$, and β_i, $i = 1, \ldots, p$. It is clearly sufficient to assume that α_0 is positive and all the other parameters are nonnegative, as in [37], but these conditions can be relaxed substantially if $p, q > 0$ and $p + q > 2$ [173].

[37] has shown that the process defined by (27) and (29) is covariance stationary iff

$$P(z) = z^m - \sum_{i=1}^{m} (\alpha_i + \beta_i) z^{m-i} = 0 \Rightarrow |z| < 1, \tag{30}$$

where $m = \max\{p, q\}$, and $\alpha_i = 0$ for $i > q$, and $\beta_i = 0$ for $i < p$, which boils down to $\sum_i \alpha_i + \sum_i \beta_i < 1$ in case the nonnegativity restrictions of [37] are imposed. The situation $\sum_i \alpha_i + \sum_i \beta_i = 1$ is referred to as an integrated GARCH (IGARCH) model, and in applications it is often found that the sum is just below unity. This indicates a high degree of volatility persistence, but the interpretation of this phenomenon is not so clear–cut [167]. If (30) holds, the unconditional variance of the process defined by (27) and (29) is given by

$$\mathbb{E}\left[\epsilon_t^2\right] = \frac{\alpha_0}{1 - \sum_{i=1}^{q} \alpha_i - \sum_{i=1}^{p} \beta_i}. \tag{31}$$

In practice, the GARCH(1,1) specification is of particular importance, and it will be the focus of our discussion too, i.e., we shall concentrate on the model (27) with

$$\sigma_t^2 = \alpha_0 + (\alpha_1 \eta_{t-1}^2 + \beta_1)\sigma_{t-1}^2, \quad \alpha_0 > 0, \quad \alpha_1 > 0, \quad 1 > \beta_1 \geq 0. \tag{32}$$

The case $\alpha_1 = 0$ corresponds to a model with constant variance, which is of no interest in the current discussion.

An interesting property of the GARCH process is that its unconditional distribution is fat–tailed even with light–tailed (e.g., Gaussian) innovations, i.e., the distributional properties of the returns will not reflect those of the innovation (news) process. This has been known basically since [73, 37], who showed that, even with normally distributed innovations, (G)ARCH processes do not have all their moments finite. For example, for the GARCH(1,1) model, [37] showed that, with $m \in \mathbb{N}$, the unconditional $(2m)$th moment of ϵ_t in (27) is finite if and only if

$$\mathbb{E}\left[(\alpha_1 \eta_t^2 + \beta_1)^m\right] < 1, \tag{33}$$

which, as long as $\alpha_1 > 0$, will eventually be violated for all practically relevant distributions. The argument in [37] is based on the relation

$$\mathbb{E}\left[\sigma_t^{2m}\right] = \sum_{i=0}^{m} \binom{m}{i} \alpha_0^i \mathbb{E}\left[(\alpha_1 \eta_{t-1}^2 + \beta_1)^{m-i}\right] \mathbb{E}\left[\sigma_{t-1}^{2(m-i)}\right], \tag{34}$$

which follows from (32). The coefficient of $\mathbb{E}\left[\sigma_{t-1}^{2m}\right]$ on the right–hand side of (34) is just the expression appearing in (33), and consequently the $(2m)$th unconditional moment cannot be finite if this exceeds unity. The heavy–tailedness of the GARCH process is sometimes also exemplified by means of its unconditional kurtosis measure (3), which is finite for the GARCH(1,1) model with Gaussian innovations iff $3\alpha_1^2 + 2\alpha_1\beta_1 + \beta_1^2 < 1$. Writing (34) down for $m = 2$, using (31) and substituting into (3) gives

$$\mathbb{K}\left[\epsilon_t\right] = \frac{3[1 - (\alpha_1 + \beta_1)^2]}{1 - (\alpha_1 + \beta_1)^2 - 2\alpha_1^2} > 3,$$

as $\mathbb{E}\left[\epsilon_t^4\right] = 3\mathbb{E}\left[\sigma_t^4\right]$. [73] notes that "[m]any statistical procedures have been designed to be robust to large errors, but ... none of this literature has made use of the fact that temporal clustering of outliers can be used to predict their occurrence and minimize their effects. This is exactly the approach taken by the ARCH model". Conditions for the existence of and expressions for higher–order moments of the GARCH(p, q) model can be found in [50, 105, 122, 139]. The relation between the conditional and unconditional kurtosis of GARCH models was investigated in [15], see also [47] for related results.

A more precise characterization of the tails of GARCH processes has been developed by applying classical results about the tail behavior of solutions of stochastic difference equations as, for example, in [124]. We shall continue to concentrate on the GARCH(1,1) case, which admits relatively explicit results, and which has already been written as a first–order stochastic difference equation in (32). Iterating this,

$$\sigma_t^2 = \sigma_0^2 \prod_{i=1}^{t} (\alpha_1 \eta_{t-i}^2 + \beta_1) + \alpha_0 \left[1 + \sum_{k=1}^{t-1} \prod_{i=1}^{k} (\alpha_1 \eta_{t-i}^2 + \beta_1) \right]. \tag{35}$$

[171] has shown that the GARCH(1,1) process (32) has a strictly stationary solution, given by

$$\sigma_t^2 = \alpha_0 \left[1 + \sum_{k=1}^{\infty} \prod_{i=1}^{k} (\alpha_1 \eta_{t-i}^2 + \beta_1) \right], \tag{36}$$

if and only if

$$\mathbb{E}\left[\log(\alpha_1 \eta_t^2 + \beta_1)\right] < 0. \tag{37}$$

The keynote of the argument in [171] is the application of the strong law of large numbers to the terms of the form $\prod_{i=1}^{k}(\alpha_1 \eta_{t-i}^2 + \beta_1) = \exp\{\sum_1^k \log(\alpha_1 \eta_{t-i}^2 + \beta_1)\}$ in (35), revealing that (35) converges almost surely if (37) holds. Note that $\mathbb{E}\left[\log(\alpha_1 \eta_t^2 + \beta_1)\right] < \log \mathbb{E}\left[\alpha_1 \eta_t^2 + \beta_1\right] = \log(\alpha_1 + \beta_1)$, i.e., stationary GARCH processes need not be covariance stationary. Using (36) and standard moment inequalities, [171] further established that, in case of stationarity, $\mathbb{E}\left[|\epsilon_t|^p\right]$, $p > 0$, is finite if and only if $\mathbb{E}\left[(\alpha_1 \eta_t^2 + \beta_1)^{p/2}\right] < 1$, which generalizes (33) to noninteger moments. It may now be supposed, and, building on the results of [124, 90], has indeed been established by [166], that the tails of the marginal distribution of ϵ_t generated by a GARCH(1,1) process decay asymptotically in a Pareto–type fashion, i.e.,

$$P(|\epsilon_t| > x) \cong c x^{-\alpha} \quad \text{as} \quad x \to \infty, \tag{38}$$

where the tail index α is the unique positive solution of the equation

$$h(\alpha) := \mathbb{E}\left[(\alpha_1 \eta_t^2 + \beta_1)^{\alpha/2}\right] = 1. \tag{39}$$

This follows from (28) along with the result that the tails of σ_t^2 and σ_t are asymptotically Paretian with tail indices $\alpha/2$ and α, respectively. For a discussion of technical conditions, see [166]. [166] also provides an expression for the constant c in (38), which is difficult to calculate explicitly, however. For the ARCH(1) model with Gaussian innovations, (39) becomes $(2\alpha_1)^{\alpha/2}\Gamma[(\alpha+1)/2]/\sqrt{\pi} = 1$, which has already been obtained by [63] and was foreshadowed in the work of [168]. The results reported above have been generalized in various directions, with qualitatively similar conclusions. The GARCH(p, q) case is treated in [29], while [140, 141] consider various extensions of the standard GARCH(1,1) model.

Although the *unconditional* distribution of a GARCH model with Gaussian innovations has genuinely fat tails, it is often found in applications that the tails of empirical return distributions are even fatter than those implied by fitted Gaussian GARCH models, indicating that the *conditional* distribution, i.e., the distribution of η_t in (27), is likewise fat–tailed. Therefore, it has become standard practice to assume that the innovations η_t are also heavy tailed, although it has been questioned whether this is the best modeling strategy [199]. The most popular example of a heavy tailed innovation distribution is certainly the t considered in Section 6.2.3, which was introduced by [38] into the GARCH literature. Some authors have also found it beneficial to let the degrees of freedom parameter ν in (23) be time–varying, thus obtaining time–varying conditional fat–tailedness [45].

In the following, we shall briefly discuss a few GARCH(1,1) estimation results for the S&P500 series in order to compare the tails implied by these models with those from the semiparametric estimation procedures in Section 5. As distributions for the innovation process $\{\eta_t\}$, we shall consider the Gaussian, t, and the generalized error distribution (GED), which was introduced by [172] into the GARCH literature, see [128] for a recent contribution and asymmetric extensions. It has earlier been used in an unconditional context by [94] for the S&P500 returns. The density of the GED with mean zero and unit variance is given by

$$f_{GED}(x;\nu) = \frac{\lambda\nu}{2^{1/\nu+1}\Gamma(1/\nu)} \exp\left(-\frac{|\lambda x|^\nu}{2}\right), \quad \nu > 0, \tag{40}$$

where $\lambda = 2^{1/\nu}\sqrt{\Gamma(3/\nu)/\Gamma(1/\nu)}$. Parameter ν in (40) controls the thickness of the tails. For $\nu = 2$, we get the normal distribution, and a leptokurtic shape is obtained for $\nu < 2$. In the latter case, the tails of (40) are therefore thicker than those of the Gaussian, but they are not fat in the Pareto sense. However, even if one argues for Pareto tails of return distributions, use of (40) may be appropriate as a conditional distribution in GARCH models, because the power law already accompanies the volatility dynamics. To make the estimates of the parameter α_1 in (32) comparable, we also use the unit variance version of the t, which requires multiplying X in (23) by $\sqrt{(\nu-2)/\nu}$. Returns are modeled as $r_t = \mu + \epsilon_t$, where μ is a constant mean and ϵ_t is generated by (27) and (32). Parameter estimates, obtained by maximum–likelihood estimation, are provided in Table 2. In addition to the GARCH parameters in (32) and the shape parameters of the innovation distributions, Table 2 reports the log likelihood values and the implied tail indices, $\widehat{\alpha}$, which are obtained by solving (39) numerically. First note that all the GARCH models have considerably higher likelihood values than the iid models in Table 1, which highlights the importance of accounting for conditional heteroskedasticity. We can also conclude that the Gaussian assumption is still inadequate as a conditional distribution in GARCH models, as both the t and the GED achieve significantly higher likelihood values, and their estimated shape parameters indicate pronounced nonnormalities. However, the degrees of freedom parameter of the t, ν, is somewhat increased in comparison to Table 1, as part of the leptokurtosis is now explained by the GARCH effects.

Compared to the nonparametric tail estimates obtained in Section 5, the tail index implied by the Gaussian GARCH(1,1) model turns out to be somewhat too high, while those of the more flexible models are both between 3 and 4 and therefore more in line with what has been found in Section 5. However, for all three models, the confidence intervals for α, as obtained from 1000 simulations from the respective estimated GARCH

Table 2: GARCH parameter estimates

distribution	$\widehat{\mu}$	$\widehat{\alpha}_0$	$\widehat{\alpha}_1$	$\widehat{\beta}_1$	$\widehat{\nu}$	$\widehat{\alpha}$	loglik
normal	0.059 (0.011)	0.012 (0.002)	0.080 (0.008)	0.911 (0.009)	–	4.70 (3.20,7.22)	–7271.7
GED	0.063 (0.010)	0.007 (0.002)	0.058 (0.007)	0.936 (0.008)	1.291 (0.031)	3.95 (2.52,6.95)	–7088.2
symmetric t	0.063 (0.010)	0.006 (0.002)	0.051 (0.006)	0.943 (0.007)	6.224 (0.507)	3.79 (2.38,5.87)	–7068.1

Shown are maximum–likelihood estimation results for GARCH(1,1) models, as given by (27) and (32), with different assumptions about the distribution of the innovations η_t in (27). Standard errors for the model parameters and 95% confidence intervals for the implied tail indices, $\widehat{\alpha}$, are given in parentheses. "loglik" is the value of the maximized log likelihood function.

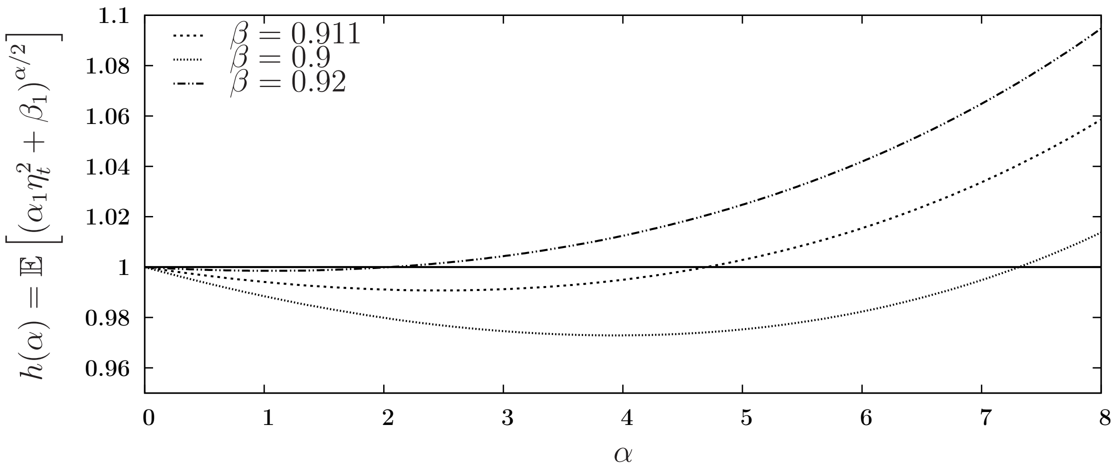

Figure 8: The figure displays the function $h(\alpha)$, as defined in (39), for Gaussian η_t, $\alpha = 0.0799$ and various values of β_1. Note that $\widehat{\alpha}_1 = 0.0799$ and $\widehat{\beta}_1 = 0.911$ are the maximum likelihood estimates for the S&P500 returns, as reported in Table 2.

processes, are rather wide, so that we cannot conclusively rule out the existence of the unconditional fourth (and even fifth) moment. The width of the confidence intervals reflects the fact that the implied tail indices are very sensitive to small variations in the underlying GARCH parameters. For example, if, in the GARCH model with conditional normality, we replace the estimate $\widehat{\beta}_1 = 0.911$ with 0.9, the implied tail index is 7.31, and with $\beta_1 = 0.92$, we get $\alpha = 2.05$, which is close to an infinite variance. The situation is depicted in Figure 8, showing $h(\alpha)$ in (39) for the different values of β_1. The shape of h follows generally from $h(0) = 1$, $h'(0) < 0$ by (37), $h'' > 0$, i.e., h is convex, and $\lim_{\alpha\to\infty} h(\alpha) = \infty$ as long as $P[(\alpha_1\eta_t^2 + \beta_1) > 1] > 0$, so that $h(\alpha) = 1$ has a unique positive solution. Note that both 0.9 and 0.92 are covered by $0.911 \pm 2 \times 0.009$, i.e., a 95% confidence interval for β_1. This shows that the GARCH–implied tail indices are rather noisy.

Alternatively, we may avoid precise assumptions about the distribution of the innovation process $\{\eta_t\}$ and rely on quasi maximum–likelihood results [138]. That is, we estimate the innovations by $\widehat{\eta}_t = \widehat{\epsilon}_t / \widehat{\sigma}_t$, $t = 1, \ldots, 5550$, where $\{\widehat{\sigma}_t\}$ is the sequence of conditional standard deviations implied by the estimated Gaussian GARCH model, and then solve the sample analogue of (39), i.e., $T^{-1} \sum_{t=1}^{T} (\widehat{\alpha}_1 \widehat{\eta}_t^2 + \widehat{\beta}_1)^{\alpha/2} = 1$, a procedure theoretically justified in [32]. Doing so, we obtain $\widehat{\alpha} = 2.97$, so that we recover the "universal cubic law". However, the 95% confidence interval, calculated from 1000 GARCH simulations, where the innovation sequences are obtained by sampling with replacement from the $\widehat{\eta}_t$–series, is $(1.73, 4.80)$, which is still reconcilable with a finite fourth moment, and even with an infinite second moment. These results clearly underline the caveat brought out by [72] (p. 349), that "[t]here is no free lunch when it comes to [tail index] estimation".

8 Application to Value–at–Risk

In this section, we compare the models discussed in Sections 6 and 7 on an economic basis by employing the Value–at–Risk (VaR) concept, which is a widely used measure to describe the downside risk of a financial position both in industry and in academia [118]. Consider a time series of portfolio returns, r_t, and an associated series of ex–ante VaR measures with target probability ξ, $\mathrm{VaR}_t(\xi)$. The $\mathrm{VaR}_t(\xi)$ implied by a model $\mathcal{M}$ is defined by

$$\Pr_{t-1}^{\mathcal{M}}(r_t < -\mathrm{VaR}_t(\xi)) = \xi, \tag{41}$$

where $\Pr_{t-1}^{\mathcal{M}}(\cdot)$ denotes a probability derived from model $\mathcal{M}$ using the information up to time $t-1$, and the negative sign in (41) is due to the convention of reporting VaR as a positive number. For an appropriately specified model, we expect $100 \times \xi\%$ of the observed return values not to exceed the (negative of the) respective VaR forecast. Thus, to assess the performance of the different models, we examine the percentage shortfall frequencies,

$$U_\xi = 100 \times \frac{x}{T} = 100 \times \hat{\xi}, \tag{42}$$

where T denotes the number of forecasts evaluated, x is the observed shortfall frequency, i.e., the number of days for which $r_t < -\mathrm{VaR}_t(\xi)$, and $\hat{\xi} = x/T$ is the empirical shortfall probability. If $\hat{\xi}$ is significantly less (higher) than ξ, then model $\mathcal{M}$ tends to overestimate (underestimate) the risk of the position. In the present application, in order to capture even the more extreme tail region, we focus on the target probabilities $\xi = 0.001$, 0.0025, 0.005, 0.01, 0.025, and 0.05.

To formally test whether a model correctly estimates the risk (according to VaR) inherent in a given financial position, that is, whether the empirical shortfall probability, $\hat{\xi}$, is statistically indistinguishable from the nominal shortfall probability, ξ, we use the likelihood ratio test [133]

$$\mathrm{LRT}_{\mathrm{VaR}} = -2\{x \log(\xi/\hat{\xi}) + (T-x)\log[(1-\xi)/(1-\hat{\xi})]\} \overset{asy}{\sim} \chi^2(1). \tag{43}$$

Based on the first 1000 return observations, we calculate one–day–ahead VaR measures based on parameter estimates obtained from an expanding data window, where the

Table 3: Backtesting Value–at–Risk measures

Unconditional Distributional Models						
distribution	$U_{0.001}$	$U_{0.0025}$	$U_{0.005}$	$U_{0.01}$	$U_{0.025}$	$U_{0.05}$
GH	0.04	0.11^{**}	0.24^{***}	0.73^{*}	2.70	5.89^{***}
t_{GH}	0.07	0.11^{**}	0.22^{***}	0.75^{*}	2.75	5.96^{***}
t_{JF}	0.04	0.11^{**}	0.31^{**}	0.88	2.64	5.32
t_{AC}	0.04	0.11^{**}	0.26^{**}	0.84	2.48	5.16
t_{FS}	0.07	0.13^{*}	0.33^{**}	0.95	2.77	5.38
symmetric t	0.07	0.15	0.31^{**}	0.92	3.08^{**}	6.35^{***}
NIG	0.07	0.15	0.26^{**}	0.70^{**}	2.35	5.34
HYP	0.13	0.24	0.51	0.95	2.50	5.16
VG	0.13	0.24	0.51	0.92	2.46	5.10
alpha stable	0.04	0.11^{**}	0.33^{**}	0.75^{*}	2.44	4.90
finite mixture ($k = 2$)	0.04	0.07^{***}	0.11^{***}	0.37^{***}	2.99^{**}	6.40^{***}
Cauchy	0.00^{***}	0.00^{***}	0.00^{***}	0.00^{***}	0.09^{***}	0.88^{***}
normal	0.48^{***}	0.64^{***}	0.97^{***}	1.36^{**}	2.44	4.02^{***}
GARCH(1,1) Models						
distribution	$U_{0.001}$	$U_{0.0025}$	$U_{0.005}$	$U_{0.01}$	$U_{0.025}$	$U_{0.05}$
normal	0.40^{***}	0.66^{***}	0.92^{***}	1.36^{**}	2.95^{*}	4.57
GED	0.20^{*}	0.33	0.44	0.79	2.48	4.79
symmetric t	0.11	0.26	0.40	0.92	2.86	5.45

The table shows the realized one–day–ahead percentage shortfall frequencies, U_ξ, for given target probabilities, ξ, as defined in (42). Asterisks *, ** and *** indicate significance at the 10%, 5% and 1% levels, respectively, as obtained from the likelihood ratio test (43).

parameters are updated every day. Thus we get, for each model, 4550 one–day–ahead out–of–sample VaR measures.

Table 3 reports the realized one–day–ahead percentage shortfall frequencies for the different target probabilities, ξ, as given above. The upper panel of the table shows the results for the unconditional distributions discussed in Section 6. The results clearly show that the normal distribution strongly *underestimates* ($\hat{\xi} > \xi$) the downside risk for the lower target probabilities, while the Cauchy as well as the alpha stable distributions tend to significantly *overestimate* ($\hat{\xi} < \xi$) the tails. This is in line with what we have observed from the empirical density plots presented in Figure 7, which, in contrast to the out–of–sample VaR calculations, are based on estimates for the entire sample. Interestingly, the finite normal mixture distribution also tends to overestimate the risk at the lower VaR levels, leading to a rejection of correct coverage for almost all target probabilities. In contrast, the HYP distribution, whose empirical tails have been very close to those of the normal mixture in–sample (see Figure 7), nicely reproduces the target probabilities, as does the VG distribution.

Similarly to the log likelihood results presented in Section 6.4 the Value–at–Risk evaluation does not allow for a clear distinction between the different t distributions, the GH and the NIG distribution. Similar to the Cauchy and the stable, they all tend to overestimate the more extreme target probabilities, while they imply too large shortfall probabilities at the five percent quantile.

The fact that most unconditional distributional models tend to overestimate the risk at the lower target probabilities may be due to our use of an expanding data window and the impact of the "Black Monday", where the index decreased by more than 20%, at the beginning of our sample period. In this regard, the advantages of accounting for time–varying volatility via a GARCH(1,1) structure may become apparent, as this model allows the more recent observations to have much more impact on the conditional density forecasts.

In fact, by inspection of the results for the GARCH models, as reported in the lower part of Table 3, it turns out that the GARCH(1,1) model with a normal distribution strongly underestimates the empirical shortfall probabilities at all levels except the largest (5%). However, considering a GED or t distribution for the return innovations within the GARCH model provides accurate estimates of downside risks.

To further discriminate between the GARCH processes and the iid models, tests for *conditional* coverage may be useful, which are discussed in the voluminous VaR literature (e.g., [53]).

Finally, we point out that the current application is necessarily of an illustrative nature. In particular, if the data generating process is not constant but evolves slowly over time and/or is subject to abrupt structural breaks, use of a rolling data window will be preferred to an expanding window.

9 Future Directions

As highlighted in the previous sections, there exists a plethora of different and well–established approaches for modeling the tails of univariate financial time series. However, on the multivariate level the number of models and distributions is still very limited, although the joint modeling of multiple asset returns is crucial for portfolio risk management and allocation decisions. The problem is then to model the dependencies between financial assets. In the literature, this problem has been tackled, for example, by means of multivariate extensions of the mean–variance mixture (18) [18], multivariate GARCH models [30], regime–switching models [11], and copulas [51]. The problem is particularly intricate if the number of assets to be considered is large, and much work remains to understand and properly model their dependence structure.

It is also worth mentioning that the class of GARCH processes, due to its interesting conditional and unconditional distributional properties, has been adopted, for example, in the signal processing literature [52, 55, 3], and it is to be expected that it will be applied in other fields in the future.

A Literature

A.1 Primary Literature

[1] K. Aas and I. H. Haff. The generalized hyperbolic skew Student's t-distribution. *Journal of Financial Econometrics*, 4:275–309, 2006. [25, 28, 29]

[2] A. Abhyankar, L. S. Copeland, and W. Wong. Moment condition failure in high frequency financial data: Evidence from the S&P 500. *Applied Economics Letters*, 2:288–290, 1995. [12]

[3] A. Abramson and I. Cohen. State smoothing in Markov-switching time-frequency GARCH models. *IEEE Signal Processing Letters*, 13:377–380, 2006. [42]

[4] V. Akgiray and G. G. Booth. The stable-law model of stock returns. *Journal of Business & Economic Statistics*, 6:51–57, 1988. [10]

[5] C. Alexander and E. Lazar. Normal mixture GARCH(1,1). Applications to exchange rate modelling. *Journal of Applied Econometrics*, 21:307–336, 2006. [31]

[6] S. S. Alexander. Price movements in speculative markets: Trends or random walks. *Industrial Management Review*, 2:7–25, 1961. [5]

[7] T. G. Andersen and T. Bollerslev. Answering the skeptics: Yes, standard volatility models do provide accurate forecasts. *International Economic Review*, 39:885–905, 1998. [32]

[8] T. G. Andersen, T. Bollerslev, F. X. Diebold, and H. Ebens. The distribution of realized stock return volatility. *Journal of Financial Economics*, 61:43–76, 2001. [32, 35]

[9] T. G. Andersen, T. Bollerslev, F. X. Diebold, and P. Labys. The distribution of realized exchange rate volatility. *Journal of the American Statistical Association*, 96:42–55, 2001. [32, 35]

[10] J. Andersson. On the normal inverse Gaussian stochastic volatility model. *Journal of Business & Economic Statistics*, 19:44–54, 2001. [27]

[11] A. Ang and J. Chen. Asymmetric correlations of equity portfolios. *Journal of Financial Economics*, 63:443–494, 2002. [42]

[12] A. C. Atkinson. The simulation of generalized inverse Gaussian and hyperbolic random variables. *SIAM Journal on Scientific and Statistical Computing*, 3:502–515, 1982. [22]

[13] A. Azzalini and A. Capitanio. Distributions generated by perturbation of symmetry with emphasis on a multivariate skew t-distribution. *Journal of the Royal Statistical Society. Series B*, 65:367–389, 2003. [29]

[14] L. Bachelier. Theory of speculation. In P. H. Cootner, editor, *The Random Character of Stock Market Prices*, pages 17–75. MIT Press, 1964. [5]

[15] X. Bai, J. R. Russell, and G. C. Tiao. Kurtosis of GARCH and stochastic volatility models with non–normal innovations. *Journal of Econometrics*, 114:349–360, 2003. [36]

[16] K. P. Balanda and H. L. MacGillivray. Kurtosis: A critical review. *The American Statistician*, 42:111–119, 1988. [6]

[17] O. E. Barndorff-Nielsen. Exponentially decreasing distributions for the logarithm of particle size. *Proceedings of the Royal Society of London. Series A*, 353:401–419, 1977. [21]

[18] O. E. Barndorff-Nielsen. Normal inverse Gaussian distributions and stochastic volatility modelling. *Scandinavian Journal of Statistics*, 24:1–13, 1997. [27, 42]

[19] O. E. Barndorff-Nielsen. Processes of normal inverse Gaussian type. *Finance and Stochastics*, 2:41–68, 1998. [27]

[20] O. E. Barndorff-Nielsen, P. Blæsild, J. L. Jensen, and M. Sørensen. The fascination of sand. In A. C. Atkinson and S. E. Fienberg, editors, *A Celebration of Statistics*, pages 57–87. Springer, 1985. [23]

[21] O. E. Barndorff-Nielsen and C. Halgreen. Infinite divisibility of the hyperbolic and generalized inverse Gaussian distributions. *Probability Theory and Related Fields*, 38:309–311, 1977. [21]

[22] O. E. Barndorff-Nielsen, J. Kent, and M. Sørensen. Normal variance-mean mixtures and z distributions. *International Statistical Review*, 50:145–159, 1982. [21]

[23] O. E. Barndorff-Nielsen and K. Prause. Apparent scaling. *Finance and Stochastics*, 5:103–113, 2001. [27]

[24] O. E. Barndorff-Nielsen and N. Shephard. Normal modified stable processes. *Theory of Probability and Mathematical Statistics*, 65:1–19, 2001. [8, 20, 21, 22]

[25] O. E. Barndorff-Nielsen and N. Shephard. Estimating quadratic variation using realized variance. *Journal of Applied Econometrics*, 17:457–477, 2002. [32]

[26] O. E. Barndorff-Nielsen and N. Shephard. *Financial Volatility in Continuous Time*. Cambridge University Press, 2007. [21]

[27] O. E. Barndorff-Nielsen and R. Stelzer. Absolute moments of generalized hyperbolic distributions and approximate scaling of normal inverse Gaussian Lévy processes. *Scandinavian Journal of Statistics*, 32:617–637, 2005. [22]

[28] A. Barnea and D. H. Downes. A reexamination of the empirical distribution of stock price changes. *Journal of the American Statistical Association*, 68:348–350, 1973. [10]

[29] B. Basrak, R. A. Davis, and T. Mikosch. Regular variation of GARCH processes. *Stochastic Processes and their Applications*, 99:95–115, 2002. [37]

[30] L. Bauwens, S. Laurent, and J. V. K. Rombouts. Multivariate GARCH models: A survey. *Journal of Applied Econometrics*, 21:79–109, 2006. [42]

[31] E. M. L. Beale and C. L. Mallows. Scale mixing of symmetric distributions with zero mean. *Annals of Mathematical Statistics*, 30:1145–1151, 1959. [35]

[32] I. Berkes, L. Horváth, and P. Kokoszka. Estimation of the maximal moment exponent of a GARCH(1,1) sequence. *Econometric Theory*, 19:565–586, 2003. [40]

[33] B. M. Bibby and M. Sørensen. A hyperbolic diffusion model for stock prices. *Finance and Stochastics*, 1:25–41, 1997. [26]

[34] N. H. Bingham, C. M. Goldie, and J. L. Teugels. *Regular Variation*. Cambridge University Press, 1987. [7]

[35] O. J. Blanchard and M. W. Watson. Bubbles, rational expectations, and financial markets. In P. Wachtel, editor, *Crises in the Economic and Financial Structure*, pages 295–315. Lexington Books, 1982. [15]

[36] R. C. Blattberg and N. J. Gonedes. A comparison of the stable and Student distributions as statistical models for stock prices. *Journal of Business*, 47:244–280, 1974. [10, 28]

[37] T. Bollerslev. Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*, 31:307–327, 1986. [35, 36]

[38] T. Bollerslev. A conditionally heteroskedastic time series model for speculative prices and rates of return. *Review of Economics and Statistics*, 69:542–547, 1987. [28, 38]

[39] T. Bollerslev, U. Kretschmer, C. Pigorsch, and G. Tauchen. A discrete-time model for daily S&P500 returns and realized variations: Jumps and leverage effects. *Journal of Econometrics*, forthcoming, 2007. [27]

[40] P. Boothe and D. Glassman. The statistical distribution of exchange rates. *Journal of International Economics*, 22:297–319, 1987. [10, 31]

[41] J.-P. Bouchaud and M. Potters. *Theory of Financial Risks. From Statistical Physics to Risk Management*. Cambridge University Press, 2000. [12]

[42] G. E. P. Box and M. E. Muller. A note on the generation of random normal deviates. *Annals of Mathematical Statistics*, 29:610–611, 1958. [19]

[43] M. D. Branco and D. K. Dey. A general class of multivariate skew-elliptical distributions. *Journal of Multivariate Analysis*, 79:99–113, 2001. [29]

[44] D. S. Broca. Mixture distribution models of Indian stock returns: An empirical comparison. *Indian Journal of Economics*, 84:525–535, 2004. [31]

[45] C. Brooks, S. P. Burke, S. Heravi, and G. Persand. Autoregressive conditional kurtosis. *Journal of Financial Econometrics*, 3:399–421, 2005. [38]

[46] J. Campbell, A. W. Lo, and A. C. MacKinlay. *The Econometrics of Financial Markets*. Princeton University Press, 1997. [5]

[47] M. A. Carnero, D. Pẽna, and E. Ruiz. Persistence and kurtosis in GARCH and stochastic volatility models. *Journal of Financial Econometrics*, 2:319–342, 2004. [36]

[48] P. Carr, H. Geman, D. B. Madan, and M. Yor. The fine structure of asset returns: An empirical investigation. *Journal of Business*, 75:305–332, 2002. [29]

[49] J. M. Chambers, C. L. Mallows, and B. W. Stuck. A method for simulating stable random variables. *Journal of the American Statistical Association*, 71:340–344, 1976. [18]

[50] M. Chen and H. Z. An. A note on the stationarity and the existence of moments of the GARCH model. *Statistica Sinica*, 8:505–510, 1998. [36]

[51] U. Cherubini, E. Luciano, and W. Vecchiato. *Copula Methods in Finance*. John Wiley & Sons, 2004. [42]

[52] Y. M. Cheung and L. Xu. Dual multivariate auto-regressive modeling in state space for temporal signal separation. *IEEE Transactions on Systems, Man, and Cybernetics—Part B: Cybernetics*, 33:386–398, 2003. [42]

[53] P. F. Christoffersen and D. Pelletier. Backtesting value–at–risk: A duration–based approach. *Journal of Financial Econometrics*, 2:84 – 108, 2004. [42]

[54] P. K. Clark. A subordinated stochastic process model with finite variance for speculative prices. *Econometrica*, 41:135–155, 1973. [10]

[55] I. Cohen. Modeling speech signals in the time-frequency domain using GARCH. *Signal Processing*, 84:2453–2459, 2004. [42]

[56] R. Cont and P. Tankov. *Financial Modelling with Jump Processes*. Chapman & Hall, 2004. [20]

[57] F. Corsi, S. Mittnik, C. Pigorsch, and U. Pigorsch. The volatility of realized volatility. Working Paper, Center for Financial Studies, 2007. [27]

[58] J. Cotter. Tail behaviour of the euro. *Applied Economics*, 37:827–840, 2005. [11]

[59] M. M. Dacorogna, U. A. Müller, O. V. Pictet, and C. G. de Vries. Extremal forex returns in extremely large data sets. *Extremes*, 4:105–127, 2001. [11]

[60] J. S. Dagpunar. An easily implemented generalised inverse Gaussian generator. *Communications in Statistics: Simulations*, 18:703–710, 1989. [22]

[61] J. Daníelsson, L. de Haan, L. Peng, and C. G. de Vries. Using a bootstrap method to choose the sample fraction in tail index estimation. *Journal of Multivariate Analysis*, 76:226–248, 2001. [11]

[62] J. Daníelsson and C. G. de Vries. Tail index and quantile estimation with very high frequency data. *Journal of Empirical Finance*, 4:241–257, 1997. [11]

[63] L. de Haan, S. I. Resnick, H. Rootzén, and C. G. de Vries. Extremal behaviour of solutions to a stochastic difference equation with applications to ARCH processes. *Stochastic Processes and their Applications*, 32:213–234, 1989. [9, 37]

[64] C. G. de Vries. Stylized facts of nominal exchange rate returns. In F. V. D. Ploeg, editor, *The Handbook of International Macroeconomics*, pages 348–389. Blackwell, 1994. [10]

[65] T. Doganoglu and S. Mittnik. An approximation procedure for asymmetric stable Paretian densities. *Computational Statistics*, 13:463–475, 1998. [18]

[66] W. H. DuMouchel. On the asymptotic normality of the maximum-likelihood estimate when sampling from a stable distribution. *Annals of Statistics*, 1:948–957, 1973. [18]

[67] W. H. DuMouchel. Estimating the stable index α in order to measure tail thickness: A critique. *Annals of Statistics*, 11:1019–1031, 1983. [10]

[68] F. J. Dyson. A note on kurtosis. *Journal of the Royal Statistical Society*, 106:360–361, 1943. [6]

[69] E. Eberlein and U. Keller. Hyperbolic distributions in finance. *Bernoulli*, 1:281–299, 1995. [26]

[70] E. Eberlein, U. Keller, and K. Prause. New insights into smile, mispricing, and value at risk: The hyperbolic model. *Journal of Business*, 71:371–405, 1998. [26]

[71] E. Eberlein and E. A. von Hammerstein. Generalized hyperbolic and inverse Gaussian distributions: Limiting cases and approximation of processes. In R. Dalang, M. Dozzi, and F. Russo, editors, *Seminar on Stochastic Analysis, Random Fields and Applications IV*, pages 221–264. Birkhäuser, 2004. [23]

[72] P. Embrechts, C. Klüppelberg, and T. Mikosch. *Modelling Extremal Events for Insurance and Finance*. Springer, 1997. [7, 9, 10, 11, 40]

[73] R. F. Engle. Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. *Econometrica*, 50:987–1006, 1982. [35, 36]

[74] E. F. Fama. Mandelbrot and the stable Paretian hypothesis. *Journal of Business*, 36:420–429, 1963. [5]

[75] E. F. Fama. The behavior of stock market prices. *Journal of Business*, 38:34–105, 1965. [5, 9, 10]

[76] E. F. Fama. *Foundations of Finance*. Basic Books, 1976. [5]

[77] E. F. Fama and R. Roll. Parameter estimates for symmetric stable distributions. *Journal of the American Statistical Association*, 66:331–338, 1971. [18]

[78] J. D. Farmer and F. Lillo. On the origin of power-law tails in price fluctuations. *Quantitative Finance*, 4:C7–C11, 2004. [15]

[79] W. Feller. *An Introduction to Probability Theory and its Applications I.* John Wiley & Sons, 1950. [6]

[80] W. Feller. *An Introduction to Probability Theory and its Applications II.* John Wiley & Sons, 1971. [11]

[81] C. Fernández and M. F. Steel. On Bayesian modeling of fat tails and skewness. *Journal of the American Statistical Association*, 93:359–371, 1998. [28]

[82] H. M. Finucan. A note on kurtosis. *Journal of the Royal Statistical Society. Series B*, 26:111–112, 1964. [6]

[83] X. Gabaix, P. Gopikrishnan, V. Plerou, and E. Stanley. A unified econophysics explanation for the power-law exponents of stock market activity. *Physica A*, 382:81–88, 2007. [15]

[84] X. Gabaix, P. Gopikrishnan, V. Plerou, and H. E. Stanley. A theory of power-law distributions in financial market fluctuations. *Nature*, 423:267–270, 2003. [15]

[85] X. Gabaix, P. Gopikrishnan, V. Plerou, and H. E. Stanley. Institutional investors and stock market volatility. *Quarterly Journal of Economics*, 121:461–504, 2006. [15]

[86] J. W. Galbraith and S. Zernov. Circuit breakers and the tail index of equity returns. *Journal of Financial Econometrics*, 2:109–129, 2004. [10]

[87] A. R. Gallant and G. Tauchen. Which moments to match? *Econometric Theory*, 12:657–681, 1996. [27]

[88] H. U. Gerber and E. S. W. Shiu. Option pricing by Esscher transforms. *Transactions of Society of Actuaries*, 46:99–140, 1994. [19]

[89] D. Ghose and K. F. Kroner. The relationship between GARCH and symmetric stable processes: Finding the source of fat tails in financial data. *Journal of Empirical Finance*, 2:225–251, 1995. [11, 12]

[90] C. M. Goldie. Implicit renewal theory and tails of solutions of random equations. *Annals of Applied Probability*, 1:126–166, 1991. [37]

[91] P. Gopikrishnan, M. Meyer, L. A. N. Amaral, and H. E. Stanley. Inverse cubic law for the distribution of stock price variations. *European Physical Journal B*, 3:139–140, 1998. [12, 15]

[92] P. Gopikrishnan, V. Plerou, L. A. N. Amaral, M. Meyer, and H. E. Stanley. Scaling of the distribution of fluctuations of financial market indices. *Physical Review E*, 60:5305–5316, 1999. [12]

[93] C. Gourieroux and J. Jasiak. Truncated maximum likelihood, goodness of fit tests and tail analysis. Working Paper, CREST, 1998. [14]

[94] J. B. Gray and D. W. French. Empirical comparisons of distributional models for stock index returns. *Journal of Business Finance and Accounting*, 17:451–459, 1990. [38]

[95] A. Gut. *Probability: A Graduate Course*. Springer, 2005. [7]

[96] M. Haas, S. Mittnik, and M. S. Paolella. Mixed normal conditional heteroskedasticity. *Journal of Financial Econometrics*, 2:211–250, 2004. [31]

[97] M. Haas, S. Mittnik, and M. S. Paolella. A new approach to Markov–switching GARCH models. *Journal of Financial Econometrics*, 2:493–530, 2004. [31]

[98] R. L. Hagerman. More evidence on the distribution of security returns. *Journal of Finance*, 33:1213–1221, 1978. [10]

[99] J. A. Hall, W. Brorsen, and S. H. Irwin. The distribution of futures prices: A test of the stable Paretian and mixture of normals hypotheses. *Journal of Financial and Quantitative Analysis*, 24:105–116, 1989. [10]

[100] P. Hall. On some simple estimates of an exponent of regular variation. *Journal of the Royal Statistical Society. Series B*, 44:37–42, 1982. [11]

[101] J. D. Hamilton. A new approach to the economic analysis of nonstationary time series and the business cycle. *Econometrica*, 57:357–384, 1989. [31]

[102] B. E. Hansen. Autoregressive conditional density estimation. *International Economic Review*, 35:705–730, 1994. [28]

[103] P. Harrison. Similarities in the distribution of stock market price changes between the eighteenth and twentieth centuries. *Journal of Business*, 71:55–79, 1996. [5]

[104] A. Hazan, Z. Landsman, and U. E. Makov. Robustness via a mixture of exponential power distributions. *Computational Statistics and Data Analysis*, 42:111–121, 2003. [30]

[105] C. He and T. Teräsvirta. Fourth moment structure of the GARCH(p, q) process. *Econometric Theory*, 15:824–846, 1999. [36]

[106] B. M. Hill. A simple general approach to inference about the tail of a distribution. *Annals of Statistics*, 3:1163–1174, 1975. [10]

[107] M. C. A. B. Hols and C. G. de Vries. The limiting distribution of extremal exchange rate returns. *Journal of Applied Econometrics*, 6:287–302, 1991. [11]

[108] D. A. Hsieh. Modeling heteroskedasticity in daily foreign-exchange rates. *Journal of Business & Economic Statistics*, 7:307–317, 1989. [35]

[109] D.-A. Hsu, R. B. Miller, and D. W. Wichern. On the stable Paretian behavior of stock-market prices. *Journal of the American Statistical Association*, 69:108–113, 1974. [10]

[110] R. Huisman, K. G. Koedijk, C. J. M. Kool, and F. Palm. Tail-index estimates in small samples. *Journal of Business & Economic Statistics*, 19:208–216, 2001. [11]

[111] R. Huisman, K. G. Koedijk, C. J. M. Kool, and F. Palm. The tail-fatness of FX returns reconsidered. *De Economist*, 150:299–312, 2002. [11]

[112] N. Hyung and C. G. de Vries. Portfolio diversification effects of downside risk. *Journal of Financial Econometrics*, 3:107–125, 2005. [12]

[113] D. W. Jansen and C. G. de Vries. On the frequency of large stock returns: Putting booms and busts into perspective. *Review of Economics and Statistics*, 73:18–24, 1991. [11, 12]

[114] M. B. Jensen and A. Lunde. The NIG-S&ARCH model: A fat-tailed, stochastic, and autoregressive conditional heteroskedastic volatility model. *Econometrics Journal*, 4:319–342, 2001. [27]

[115] E. Jondeau and M. Rockinger. Testing for differences in the tails of stock-market returns. *Journal of Empirical Finance*, 10:559–581, 2003. [10, 12]

[116] M. C. Jones and M. J. Faddy. A skew extension of the t-distribution, with applications. *Journal of the Royal Statistical Society. Series B*, 65:159–174, 2003. [28]

[117] B. Jørgensen. *Statistical Properties of the Generalized Inverse Gaussian Distribution*. Springer, 1982. [21]

[118] P. Jorion. *Value at Risk: The New Benchmark for Controlling Derivatives Risk*. McGraw-Hill, 2006. [40]

[119] T. Kaizoji and M. Kaizoji. Empirical laws of a stock price index and a stochastic model. *Advances in Complex Systems*, 6:303–312, 2003. [14]

[120] M. Kanter. Stable densities under change of scale and total variation inequalities. *Journal of the American Statistical Association*, 3:697–707, 1975. [18]

[121] I. Kaplansky. A common error concerning kurtosis. *Journal of the American Statistical Association*, 40:259, 1945. [6]

[122] M. Karanasos. The second moment and the autocovariance function of the squared errors of the GARCH model. *Journal of Econometrics*, 90:63–76, 1999. [36]

[123] P. Kearns and A. Pagan. Estimating the density tail index for financial time series. *Review of Economics and Statistics*, 79:171–175, 1997. [11]

[124] H. Kesten. Random difference equations and renewal theory for products of random matrices. *Acta Mathematica*, 131:207–248, 1973. [37]

[125] K. G. Koedijk, M. M. A. Schafgans, and C. G. de Vries. The tail index of exchange rate returns. *Journal of International Economics*, 29:93–108, 1990. [11]

[126] K. G. Koedijk, P. A. Stork, and C. G. de Vries. Differences between foreign exchange rate regimes: The view from the tails. *Journal of International Money and Finance*, 11:462–473, 1992. [11]

[127] S. M. Kogon and D. B. Williams. Characteristic function based estimation of stable distribution parameters. In R. J. Adler, R. E. Feldman, and M. S. Taqqu, editors, *A Practical Guide to Heavy Tails: Statistical Techniques and Applications*, pages 311–335. Birkhäuser, 1998. [18]

[128] I. Komunjer. Asymmetric power distribution: Theory and applications to risk management. *Journal of Applied Econometrics*, 22:891–921, 2007. [38]

[129] S. J. Kon. Models of stock returns: A comparison. *Journal of Finance*, 39:147–165, 1984. [31]

[130] I. Koponen. Analytic approach to the problem of convergence of truncated Lévy flights towards the Gaussian stochastic process. *Physical Review E*, 52:1197–1199, 1995. [20]

[131] I. A. Koutrouvelis. Regression-type estimation of the parameters of stable laws. *Journal of the American Statistical Association*, 75:918–928, 1980. [18]

[132] U. Küchler, K. Neumann, M. Sørensen, and A. Streller. Stock returns and hyperbolic distributions. *Mathematical and Computer Modelling*, 29:1–15, 1999. [26]

[133] P. H. Kupiec. Techniques for verifying the accuracy of risk management models. *Journal of Derivatives*, 3:73–84, 1995. [40]

[134] J. Laherrère and D. Sornette. Stretched exponential distributions in nature and economy: "fat tails" with characteristic scales. *European Physical Journal B*, 2:525–539, 1998. [8]

[135] A. H.-L. Lau, H.-S. Lau, and J. R. Wingender. The distribution of stock returns: New evidence against the stable model. *Journal of Business & Economic Statistics*, 8:217–223, 1990. [10]

[136] M. R. Leadbetter, G. Lindgren, and H. Rootzén. *Extremes and Related Properties of Random Sequences and Processes*. Springer, 1983. [9]

[137] B. LeBaron. Stochastic volatility as a simple generator of apparent financial power laws and long memory. *Quantitative Finance*, 1:621–631, 2001. [13]

[138] S.-W. Lee and B. E. Hansen. Asymptotic theory for the GARCH(1,1) quasi-maximum likelihood estimator. *Econometric Theory*, 10:29–52, 1994. [40]

[139] S. Ling and M. McAleer. Necessary and sufficient moment conditions for the GARCH(r, s) and asymmetric power GARCH(r, s) models. *Econometric Theory*, 18:722–729, 2002. [36]

[140] J.-C. Liu. On the tail behaviors of a family of GARCH processes. *Econometric Theory*, 22:852–862, 2006. [37]

[141] J.-C. Liu. On the tail behaviors of Box-Cox transformed threshold GARCH(1,1) processes. *Statistics and Probability Letters*, 76:1323–1330, 2006. [37]

[142] F. M. Longin. The asymptotic distribution of extreme stock market returns. *Journal of Business*, 69:383–408, 1996. [10]

[143] F. M. Longin. The choice of the distribution of asset returns: How extreme value theory can help? *Journal of Banking and Finance*, 29:1017–1035, 2005. [10]

[144] M. Loretan and P. C. B. Phillips. Testing the covariance stationarity of heavy-tailed time series. *Journal of Empirical Finance*, 1:211–248, 1994. [10, 11, 12, 13]

[145] T. Lux. The stable Paretian hypothesis and the frequency of large returns: An examination of major German stocks. *Applied Financial Economics*, 6:463–475, 1996. [12]

[146] T. Lux. On moment condition failure in German stock returns: An application of recent advances in extreme value statistics. *Empirical Economics*, 25:641–652, 2000. [12]

[147] T. Lux. The limiting extremal behaviour of speculative returns: An analysis of intra-daily data from the Frankfurt stock exchange. *Applied Financial Economics*, 11:299–315, 2001. [12]

[148] T. Lux and D. Sornette. On rational bubbles and fat tails. *Journal of Money, Credit and Banking*, 34:589–610, 2002. [15]

[149] D. B. Madan, P. P. Carr, and E. C. Chang. The variance gamma process and option pricing. *European Finance Review*, 2:79–105, 1998. [29]

[150] D. B. Madan and F. Milne. Option pricing with v.g. martingale components. *Mathematical Finance*, 1:39–55, 1991. [29]

[151] D. B. Madan and E. Seneta. The variance gamma (v.g.) model for share market returns. *Journal of Business*, 63(4):511–524, 1990. [29]

[152] Y. Malevergne, V. Pisarenko, and D. Sornette. Empirical distributions of stock returns: Between the stretched exponential and the power law? *Quantitative Finance*, 5:379–401, 2005. [8, 13]

[153] Y. Malevergne, V. Pisarenko, and D. Sornette. On the power of generalized extreme value (GEV) and generalized Pareto distribution (GDP) estimators for empirical distributions of stock returns. *Applied Financial Economics*, 16:271–289, 2006. [8, 13]

[154] B. Mandelbrot. New methods in statistical economics. *Journal of Political Economy*, 71:421–440, 1963. [3, 5, 9, 15, 32]

[155] B. Mandelbrot. The variation of certain speculative prices. *Journal of Business*, 36:394–419, 1963. [5, 15]

[156] B. Mandelbrot. The variation of some other speculative prices. *Journal of Business*, 40:393–413, 1967. [5]

[157] R. N. Mantegna and H. E. Stanley. Stochastic process with ultraslow convergence to a Gaussian: The truncated Lévy flight. *Physical Review Letters*, 73:2946–2949, 1994. [20]

[158] G. Marsaglia, A. W. Marshall, and F. Proschan. Moment crossings as related to density crossings. *Journal of the Royal Statistical Society. Series B*, 27:91–93, 1965. [6]

[159] D. M. Mason. Laws of large numbers for sums of extreme values. *Annals of Probability*, 10:754–764, 1982. [11]

[160] K. Matia, L. A. N. Amaral, S. P. Goodwin, and H. E. Stanley. Different scaling behaviors of commodity spot and future prices. *Physical Review E*, 66:045103, 2002. [12]

[161] K. Matia, M. Pal, H. Salunkay, and H. E. Stanley. Scale-dependent price fluctuations for the Indian stock market. *Europhysics Letters*, 66:909–914, 2004. [15]

[162] J. H. McCulloch. Simple consistent estimators of stable distribution parameters. *Communications in Statistics - Simulations*, 15:1109–1136, 1986. [18]

[163] J. H. McCulloch. Measuring tail thickness to estimate the stable index α: A critique. *Journal of Business & Economic Statistics*, 15:74–81, 1997. [12]

[164] J. B. McDonald. Probability distributions for financial models. In G. S. Maddala and C. R. Rao, editors, *Handbook of Statistics 14: Statistical Methods in Finance*, pages 427–461. Elsevier, 1996. [5]

[165] G. J. McLachlan and D. Peel. *Finite Mixture Models*. John Wiley & Sons, 2000. [30]

[166] T. Mikosch and C. Stărică. Limit theory for the sample autocorrelations and extremes of a GARCH(1,1) process. *Annals of Statistics*, 28:1427–1451, 2000. [9, 10, 35, 37]

[167] T. Mikosch and C. Stărică. Nonstationarities in financial time series, the long-range dependence, and the IGARCH effects. *Review of Economics and Statistics*, 86:378 – 390, 2004. [36]

[168] A. Milhøj. The moment structure of ARCH processes. *Scandinavian Journal of Statistics*, 12:281–292, 1985. [37]

[169] S. Mittnik and S. T. Rachev. Modeling asset retunrs with alternative stable distributions. *Econometric Reviews*, 12:261–330, 1993. [15]

[170] T. Mizuno, S. Kurihara, M. Takayasu, and H. Takayasu. Analysis of high-resolution foreign exchange data of USD-JPY for 13 years. *Physica A*, 324:296–302, 2003. [11]

[171] D. B. Nelson. Stationarity and persistence in the GARCH(1,1) model. *Econometric Theory*, 6:318–334, 1990. [37]

[172] D. B. Nelson. Conditional heteroskedasticity in asset returns: A new approach. *Econometrica*, 59:347–370, 1991. [38]

[173] D. B. Nelson and C. Q. Cao. Inequality constraints in the univariate GARCH model. *Journal of Business & Economic Statistics*, 10:229–235, 1992. [35]

[174] S. Newcomb. A generalized theory of the combination of observations so as to obtain the best result. In S. M. Stigler, editor, *American Contributions to Mathematical Statistics in the Nineteenth Century*, volume 2, pages 343–366. Arno Press, 1980. [30]

[175] J. P. Nolan. Numerical calculation of stable densities and distribution functions. *Communications in Statistics - Stochastic Models*, 13:759–774, 1997. [18]

[176] R. R. Officer. The distribution of stock returns. *Journal of the American Statistical Association*, 67:807–812, 1972. [10]

[177] M. F. Omran. Moment condition failure in stock returns: UK evidence. *Applied Mathematical Finance*, 4:201–206, 1997. [12]

[178] M. F. M. Osborne. Brownian motion in the stock market. *Operations Research*, 7:145–173, 1959. [5]

[179] A. Peiró. The distribution of stock returns: International evidence. *Applied Financial Economics*, 4:431–439, 1994. [31]

[180] P. R. Perry. More evidence on the nature of the distribution of security returns. *Journal of Financial and Quantitative Analysis*, 18:211–221, 1983. [10]

[181] V. Plerou, P. Gopikrishnan, L. A. N. Amaral, M. Meyer, and H. E. Stanley. Scaling of the distribution of price fluctuations of individual companies. *Physical Review E*, 60:6519–6529, 1998. [12]

[182] V. Plerou, P. Gopikrishnan, X. Gabaix, and E. Stanley. On the origin of power-law fluctuations in stock prices. *Quantitative Finance*, 4:C11–C15, 2004. [15]

[183] G. Pólya and G. Szegö. *Problems and Theorems in Analysis II*. Springer, 1976. [31]

[184] P. D. Praetz. The distribution of share price changes. *Journal of Business*, 45:49–55, 1972. [10, 28]

[185] K. Prause. *The Generalized Hyperbolic Model: Estimation, Financial Derivatives, and Risk Measures*. PhD thesis, Albert-Ludwigs-Universität Freiburg i. Br., 1999. [25, 26, 27, 29]

[186] S. J. Press. Estimation in univariate and multivariate stable distributions. *Journal of the American Statistical Association*, 67:842–846, 1972. [18]

[187] S. Raible. *Lévy Processes in Finance: Theory, Numerics, and Empirical Facts*. PhD thesis, Albert-Ludwigs-Universität Freiburg i. Br., 2000. [27]

[188] S. I. Resnick. *Extreme Values, Regular Variation, and Point Processes*. Springer, 1987. [7, 8, 9]

[189] S. I. Resnick. Heavy tail modeling and teletraffic data. *Annals of Statistics*, 25:1805–1849, 1997. [11]

[190] S. I. Resnick and C. Stărică. Tail index estimation for dependent data. *Annals of Applied Probability*, 8:1156–1183, 1998. [11]

[191] T. H. Rydberg. Generalized hyperbolic diffusion processes with applications in finance. *Mathematical Finance*, 9:183–201, 1999. [26]

[192] G. Samorodnitsky and M. S. Taqqu. *Stable Non-Gaussian Random Processes: Stochastic Models with Infinite Variance*. Chapman & Hall, 1994. [18]

[193] K.-I. Sato. *Lévy Processes and Infinitely Divisible Distributions*. Cambridge University Press, 1999. [18]

[194] W. Schoutens. *Lévy Processes in Finance: Pricing Financial Derivatives*. John Wiley & Sons, 2003. [21]

[195] E. Seneta. *Regularly Varying Functions*. Springer, 1976. [7]

[196] K. Sigman. Appendix: A primer on heavy-tailed distributions. *Queueing Systems*, 33:261–275, 1999. [7]

[197] A. C. Silva, R. E. Prange, and V. M. Yakovenko. Exponential distribution of financial returns at mesoscopic time lags: A new stylized fact. *Physica A*, 344:227–235, 2004. [14]

[198] B. W. Silverman. *Density Estimation for Statistics and Data Analysis*. Chapman & Hall, 1986. [5]

[199] I. Tsiakas. Periodic stochastic volatility and fat tails. *Journal of Financial Econometrics*, 4:90–135, 2006. [38]

[200] C. M. Turner, R. Startz, and C. R. Nelson. A Markov model of heteroskedasticity, risk, and learning in the stock market. *Journal of Financial Economics*, 25:3–22, 1989. [31]

[201] T. Werner and C. Upper. Time variation in the tail behavior of bund future returns. *Journal of Futures Markets*, 24:387–398, 2004. [12]

[202] R. Weron. Levy-stable distributions revisited: Tail index > 2 does not exclude the Levy-stable regime. *International Journal of Modern Physics C*, 12:209–223, 2001. [12]

A.2 Books and Reviews

[203] T. Bollerslev, R. F. Engle, and D. B. Nelson. ARCH models. In R. F. Engle and D. L. McFadden, editors, *Handbook of Econometrics IV*, pages 2959–3038. Elsevier, 1994.

[204] S. Borak, W. Härdle, and R. Weron. Stable distributions. In P. Cizek, W. Härdle, and R. Weron, editors, *Statistical Tools for Finance and Insurance*, pages 21–44. Springer, 2005.

[205] A. Janicki and A. Weron. *Simulation and Chaotic Behavior of α-Stable Stochastic Processes*. Marcel Dekker, 1993.

[206] J. H. McCulloch. Financial applications of stable distributions. In G. S. Maddala and C. R. Rao, editors, *Handbook of Statistics 14: Statistical Methods in Finance*, pages 393–425. Elsevier, 1996.

[207] T. Mikosch. Modeling dependence and tails of financial time series. In B. Finkenstädt and H. Rootzén, editors, *Extreme Values in Finance, Telecommunications, and the Environment*, pages 185–286. Chapman & Hall, 2004.

[208] S. Mittnik, S. T. Rachev, and M. S. Paolella. Stable Paretian modeling in finance: Some empirical and theoretical aspects. In R. J. Adler, R. E. Feldman, and M. S. Taqqu, editors, *A Practical Guide to Heavy Tails: Statistical Techniques and Applications*, pages 79–110. Birkhäuser, 1998.

[209] R. B. Nelsen. *An Introduction to Copulas*. Springer, 2nd edition, 2006.

[210] A. Pagan. The econometrics of financial markets. *Journal of Empirical Finance*, 3:15–102, 1996.

[211] F. C. Palm. GARCH models of volatility. In G. S. Maddala and C. R. Rao, editors, *Handbook of Statistics 14*, pages 209–240. Elsevier, 1996.

[212] S. T. Rachev and S. Mittnik. *Stable Paretian Models in Finance*. John Wiley & Sons, 2000.

[213] V. M. Zolotarev. *One-Dimensional Stable Distributions*. American Mathematical Society, 1986.