Received: 16 March 2016 | Revised: 22 November 2017
DOI: 10.1002/jae.2628

RESEARCH ARTICLE

WILEY Journal of APPLIED ECONOMETRICS

Exploiting tail shape biases to discriminate between stable and student *t* alternatives

Pengfei Sun | Casper G. de Vries

Department of Economics, Erasmus School of Economics, Erasmus University Rotterdam, Rotterdam, The Netherlands

Correspondence
Casper G. de Vries, Department of Economics, Erasmus School of Economics, Erasmus University Rotterdam, PO Box 1738, 3000 DR Rotterdam, The Netherlands.
Email: cdevries@ese.eur.nl

Funding information
European Community's Seventh Framework Programme, Grant/Award Number: FP7-PEOPLE-ITN-2008 and PITN-GA-2009-23798

Summary

The nonnormal stable laws and Student *t* distributions are used to model the unconditional distribution of financial asset returns, as both models display heavy tails. The relevance of the two models is subject to debate because empirical estimates of the tail shape conditional on either model give conflicting signals. This stems from opposing bias terms. We exploit the biases to discriminate between the two distributions. A sign estimator for the second-order scale parameter strengthens our results. Tail estimates based on asset return data match the bias induced by finite-variance unconditional Student *t* data and the generalized autoregressive conditional heteroscedasticity process.

1 | INTRODUCTION

There is a long-standing debate as to whether the nonnormal stable laws or Student *t* distributions are a better description of the unconditional distribution of financial asset returns. Initially, (Mandelbrot, 1963) advanced the stable distributions as these display a heavy tail and scaling properties that are so characteristic for asset return data. Later it was realized that other distributions, like the Student *t* class, also exhibit the heavy-tail property in the sense of a power law, but do not require an infinite variance. Student *t* distributions do not possess the self-scaling property globally; however, the tails of these distributions are nevertheless self-scaling. This follows from Feller's convolution theorem (1971, Section VIII, 8). The convolution theorem holds that for any distribution with a tail that is regularly varying—that is, exhibits a power law—the tail probability of the sum of independent variables is equal to the sum of the tail probabilities. The application of these two types of distribution in finance comprises a wide range of literature, (e.g., Blattberg & Gonedes, 1974; Fama & Miller, 1972; Hagerman, 1978; Jorion, 1996; McCulloch, 1997; Nolan, 2003; Stoyanov, Rachev, Racheva-Iotova, & Fabozzi, 2011; Tsionas, 2002). Campbell, Lo, and MacKinlay (1997) argue in their textbook that the Student *t* class is preferred over the stable class because data seem to point to finite second moments. But whether moments are bounded or not is difficult to check as empirical results are all bounded in any finite dataset.

Empirical comparisons of the two types of distribution are also found in the literature. Blattberg and Gonedes (1974), for example, use the likelihood ratio and conclude that the Student *t* distribution provides a better empirical description of asset returns than the stable laws. The likelihood approach has two drawbacks. First, there only exists an approximation of the likelihood function for the stable distribution since there is no explicit functional form for the probability density function. Second, the likelihood assigns the same weight to each of the observations, regardless of whether these are from the tail or in the center of the distribution. Lau and Lau (1993) argue that this type of test is unreliable for testing the stable hypothesis, as the tail data are underweighted. Furthermore, McCulloch (1997) notes that tail index estimates in excess of two are expected for stable distributed data with a true tail index of around 1.65. The tail index reflects the power in the power law expansion of the tail of the distribution; it is also one-to-one with the number of bounded moments (degrees of freedom in the case of the Student distribution). Thus, according to McCulloch, the stable data have the appearance of finite-variance Student *t*-like data (with degrees of freedom larger than two[1]). Unlike Blattberg and Gonedes (1974), Shao, Yu, and Yu (2001) produce a new test statistic by assigning more weight to the observations in the tail than in the center. Their test fails to reject the distributions with finite variance in smaller samples, but in the larger sample the finite variance models are rejected. A disadvantage of their approach is that the alternatives are not nested, whereas in our tail approach the alternatives are nested. There also exists a considerable literature that combines the unconditional heavy tail feature of the data with the conditional volatility clustering. In generalized autoregressive conditional heteroscedasticity (GARCH) models, the distribution of the innovations can have either normal-like tails or heavy tails, but the distribution of the stationary solution is "necessarily" heavy tailed. Paolella (2016) reports evidence against GARCH-type models with stable innovations, but Student *t* distributed innovations are commonly used in the literature. Thus the literature finds that while (the range of) values of the tail shape parameters of the two laws are theoretically distinct, the empirical estimates may cross.

This paper explains how the confusion about the tail shape parameters in the existing literature arises from opposing biases of the tail parameter estimates. We examine this issue theoretically by means of the higher-order tail expansions of stable and Student *t* distributions. We then propose to exploit the opposite signs of the bias in the Hill estimator[2] to discriminate empirically between the two models. In addition, we also investigate the unconditional tail behavior of the GARCH models. The implied tail shape and bias that the GARCH data induce is compared with the tail features implied by the two unconditional laws.

By far the more popular tail index estimator is the Hill estimator. We show that the two classes of distribution induce a bias in the Hill estimator with opposing signs. The tails of both types of distribution are characterized by a power shape, as holds for the Pareto distribution. The tail shape is governed by a parameter that is generally known as the tail index. This tail index equals the characteristic exponent of the infinite-variance stable class and the degrees of freedom of the finite-variance Student *t* class. The former class has a tail index below two, and the latter one has a tail index above two. The two intervals do not overlap, but the estimates do cross in practice. The crossing is responsible for the confusion in the literature.

In Figure 1 we have applied the Hill estimator to simulated data from both laws. The *x*-axis indicates the number of upper-order statistics that are used to calculate the Hill estimator. This gives the so-called Hill plot. From the figure, one sees that the tail index estimate from the Student *t* model quickly dives below its true value once more extreme-order statistics are included in the estimation. Eventually the average of the estimates even falls below two. The tail index estimates from the stable model are instead hump shaped and the top of the hump is larger than two. Below, we formally derive the shape of the bias in the Hill plot under the two laws. The biases are such that if one goes too deep into the center of distribution by using relatively many higher-order statistics, the Hill estimator applied to sum stable data may exceed two, while it can fall below two if the data in fact come from a Student *t* distribution. The latter conclusion also applies to data generated by GARCH-like processes with bounded unconditional second moment. This "comedy of biases" has confused the empirical literature (see the above citations), but can be exploited to become a blessing in disguise.

The ways in which these biases behave are used to discriminate between the two classes of distribution by means of the shape of the Hill plot. Rather than relying on a particular point estimate, we propose that one first uses the shape of the Hill plot to identify the type of distribution. For a stable distribution with tail index between one and two, the Hill plot first displays an upward bias, but eventually the plot turns downward, producing a hump-shaped graph. As one goes too deep into the center of the Student *t* distribution, the Hill plot declines more or less monotonically, yielding a one-way downward bias. Thus a monotonic decline points to Student *t* data, while the hump shape is revealing for data generated by a stable distribution. Once one knows the distribution class, one can correct the bias. For Student *t* data this will be an upward correction, whereas stable data require a downward correction (to at least a value below two).

The identification procedure through the Hill plot is further supported by the application of a hitherto unused sign estimator for the bias term. Ferreira and Vries (2004) introduced an estimator for the sign of the second-order scale term in the expansion of the tail of the distribution. The opposing behavior of the Hill plots directly relates to the opposite signs of the second-order tail terms of the two types of distribution. The sign estimates together with the shape of the Hill plots suggest that the unconditional distribution of financial asset returns resemble Student *t* distributed data rather than data generated by a stable distribution.

[1]In the remainder of this paper, we only focus on the Student *t* distribution with degrees of freedom larger than two—that is, with finite variance.
[2]The Hill estimator is a nonparametric extreme-value based estimator that uses only the more extreme-order statistics to estimate the tail index.

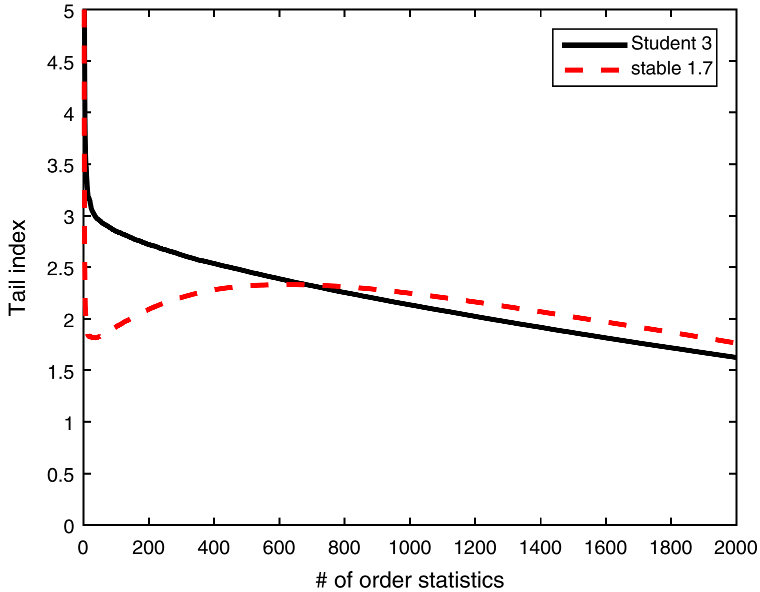

FIGURE 1 Hill plot: stable 1.7 and Student t 3. This figure depicts the Hill estimates for the stable distribution with tail index 1.7 and the Student t distribution with tail index 3 against the number of upper-order statistics used. The sample size is $n = 10^4$. Using pseudo random numbers, we draw 500 different samples. This figure gives the pointwise average estimate of tail index over these samples [Colour figure can be viewed at wileyonlinelibrary.com]

Financial asset returns are also well known to display clusters of volatility. GARCH models are one of the most popular econometric models to capture this feature. A surprising consequence is that the stationary solution implied by GARCH models displays heavy tails (see Davis & Mikosch, 2009; de Haan, Resnick, Rootzén, & de Vries, 1989; Goldie, 1991), even if the innovations are normally distributed. We show that the GARCH(1, 1) process with normal innovations exhibits a downward bias in the Hill plot, analogous to the case of the Student t distribution.

The rest of the paper is organized as follows. In Section 2 we introduce the second-order tail approximation for the competing classes of distributions. This enables us to deduce the asymptotic bias and mean squared error of the Hill estimator and a competing tail index estimator. Simulations in Section 3 illustrate the opposing bias patterns of the tail index estimators as more and more of the tail data are used. The different shapes of Hill plot for these two types of distribution are explained analytically in Section 4. To investigate the robustness of our results, we determine the direction of the bias in an alternative way by means of a sign estimator in Section 5. In Section 6 we discuss the Hill bias of the GARCH(1, 1) process with normal innovations. Section 7, concludes.

2 | SECOND-ORDER THEORY

Suppose the distribution of tails satisfies the following approximation:

$$F(x) = 1 - ax^{-\alpha}\left[1 + bx^{-\beta} + o\left(x^{-\beta}\right)\right], \quad \beta > 0, \quad \text{as } x \to \infty. \tag{1}$$

The first term is analogous to the power of a Pareto distribution. Both the symmetric stable distribution with $1 < \alpha < 2$ and the Student t distribution with degrees of freedom $\alpha > 2$ satisfy this approximation and an analogous expansion for the left tail. The four parameters for stable and Student t classes are displayed in Table 1.

Note that $b > 0$ for the stable class if $\alpha \in (1, 2)$, whereas $b < 0$ for the Student t class. This sign difference for the second-order scale parameter is responsible for the opposite bias behavior of the two laws when estimating the first-order tail index α. We propose to exploit this property in order to discriminate between the two rival distribution classes.

2.1 | Hill estimator

Consider the Hill estimator. It attains the best possible rate for the asymptotic mean squared error (AMSE) within the classes of distributions of Equation 1 (see Goldie & Smith, 1987; Hall & Welsh, 1984). It is defined as follows.

TABLE 1 Parameters under the second-order tail approximation

	Stable	Student
α	$(1,2)$	$(2,+\infty)$
β	α	2
a	$\frac{1}{\pi}\Gamma(\alpha)\sin\left(\frac{\alpha\pi}{2}\right)$	$\frac{1}{\sqrt{\alpha\pi}}\frac{\Gamma\left(\frac{\alpha+1}{2}\right)}{\Gamma\left(\frac{\alpha}{2}\right)}\alpha^{(\alpha-1)/2}$
b	$-\frac{1}{2}\frac{\Gamma(2\alpha)\sin(\alpha\pi)}{\Gamma(\alpha)\sin\left(\frac{\alpha\pi}{2}\right)}$	$-\frac{\alpha^2}{2}\frac{\alpha+1}{\alpha+2}$

Let $\{X_1,\cdots,X_n\}$ be a sample of size n independent[3] and identically distributed (i.i.d.) random variables with common cumulative distribution function (c.d.f.) $F(x)$ which satisfies Equation 1. Consider the descending order statistics from this sample around a given threshold s:

$$X_{(1)} \geq X_{(2)} \geq \cdots \geq X_{(m)} > s \geq X_{(m+1)} \geq \cdots \geq X_{(n)}.$$

Denote by A the set of indices for those order statistics which strictly exceed threshold s:

$$A = \{i|X_i > s,\ i = 1,\cdots,n\}. \tag{2}$$

Note that the number of elements in A is random if the threshold s is fixed before sampling. The number of elements is then denoted by $m(s)$ to reveal its dependence on the threshold choice. The indicator function $1_{\{i\in A\}}$ takes the value of 1 when X_i is in A and 0 otherwise.

Define the empirical conditional first logarithmic moment from the sample $X_{(1)}, \ldots, X_{(m)}$:

$$u(s) = \frac{1}{m}\sum_{i=1}^{m}\log\frac{X_{(i)}}{s}, \tag{3}$$

where $u = 0$ if A is empty. The $u(s)$ statistic is the so-called Hill estimator (see Hill, 1975).

The first two asymptotic moments of the estimator are given in the following two lemmas.

Lemma 1. *For the class of random variables with distributions that satisfy Equation 1, and letting $s_n^{\alpha}/n \to 0,\ s_n \to \infty$ as $n \to \infty$, the asymptotic bias of the statistic u is*

$$E_A\left[u(s_n) - \frac{1}{\alpha}\right] = -b\frac{\beta}{\alpha(\alpha+\beta)}s_n^{-\beta} + o\left(s_n^{-\beta}\right). \tag{4}$$

Proof. See Goldie and Smith (1987). □

From this lemma, it is clear that the sign of the second-order scale parameter b determines the sign of the bias. In particular, we demonstrate that the estimate of the inverse tail index $1/\alpha$ is upward biased for Student t class and downward biased for stable class. Often the tail index α is plotted rather than its inverse; in that case the Hill estimate is downward biased for Student t case and hump-shaped upward biased for stable distribution, as shown in Figure 1.

Furthermore, the second asymptotic moment is in Lemma 2.

Lemma 2. *Consider the class of random variables with distributions that satisfy Equation 1. Let $s_n^{\alpha}/n \to 0,\ s_n \to \infty$ as $n \to \infty$; then the asymptotic variance of the statistic u is*

$$\mathrm{var}_A\left[u(s_n) - \frac{1}{\alpha}\right] = \frac{s_n^{\alpha}}{an\alpha^2} + o\left(\frac{s_n^{\alpha}}{n}\right). \tag{5}$$

Proof. See Goldie and Smith (1987). □

[3]The independence is not crucial (see Resnick & Stărică, 1998).

2.2 | Optimal threshold levels

For either estimator, one needs to select the optimal threshold level s. We first investigate the theoretically optimal level. Upon combining the variance from Equation 5 with the bias squared from Equation 4, we obtain the AMSE:

$$\text{AMSE}\left(u\left(s_n\right)\right)=\frac{1}{a\alpha^2}\frac{s_n^{\alpha}}{n}+\frac{b^2\beta^2}{\alpha^2(\alpha+\beta)^2}s_n^{-2\beta}.$$

It is optimal to let the threshold s depend on the sample size n in such a way that both parts vanish at the same rate. As $n\to\infty, s_n\to\infty$, and the two terms in the AMSE are balanced and the AMSE vanishes at the best possible rate.

Proposition 1. *For n sufficiently large, the unique AMSE-minimizing asymptotic threshold $\bar{s}_n$ is*

$$\bar{s}_n=\left[\frac{2ab^2\beta^3}{\alpha(\alpha+\beta)^2}\right]^{\frac{1}{\alpha+2\beta}}n^{\frac{1}{\alpha+2\beta}}, \tag{6}$$

and the associated minimal AMSE is

$$\begin{aligned}&\overline{AMSE}\left[u\left(\bar{s}_n\right)\right]\\&=\frac{1}{a\alpha}\left[\frac{1}{\alpha}+\frac{1}{2\beta}\right]\left[\frac{2ab^2\beta^3}{\alpha(\alpha+\beta)^2}\right]^{\frac{\alpha}{2\beta+\alpha}}n^{-\frac{2\beta}{2\beta+\alpha}}+o\left(n^{-\frac{2\beta}{2\beta+\alpha}}\right).\end{aligned} \tag{7}$$

Proof. The optimal rate $\bar{s}_n$ is determined from the first-order condition $\partial\text{AMSE}/\partial s_n=0$, and the associated minimal $\overline{\text{AMSE}}$ can be derived by substituting the value of the optimal rate s_n in the AMSE. □

The overbars in Equations 6 and 7 reflect that the choice of s_n sequence that minimizes the AMSE. The result in Hall and Welsh (1984) shows that this rate cannot be improved upon by other estimators for distributions in the class of Equation 1.

In the practical implementation of the estimators, it is more convenient to replace the threshold s by a higher-order statistic $X_{(m+1)}$. Since $1-F(s)=as^{-\alpha}\left[1+O\left(s^{-\beta}\right)\right]$, the following result for the number of upper-order statistics is immediate:

$$n^{-\frac{2\beta}{2\beta+\alpha}}m\left(\bar{s}_n\right)\to a\left[\frac{2ab^2\beta^3}{\alpha(\alpha+\beta)^2}\right]^{-\frac{\alpha}{\alpha+2\beta}}, \tag{8}$$

with probability 1 as $n\to\infty$.

If the data come from a symmetric stable distribution, the optimal number of order statistics (Equation 8) in the case of the Hill estimator is

$$\begin{aligned}m&\simeq a\left[\frac{2ab^2\beta^3}{\alpha(\alpha+\beta)^2}\right]^{-\frac{\alpha}{\alpha+2\beta}}n^{\frac{2\beta}{2\beta+\alpha}}=\\&=\frac{2}{\pi}\Gamma(\alpha)\sin\left(\frac{\alpha\pi}{2}\right)\left[\frac{1}{\pi}\frac{(\Gamma(2\alpha))^2(\sin(\alpha\pi))^2}{\Gamma(\alpha)\sin\left(\frac{\alpha\pi}{2}\right)}\right]^{-1/3}n^{2/3}.\end{aligned}$$

Alternatively, if the data come from a Student t distribution, the optimal number of order statistics for the Hill estimator is

$$m\simeq\frac{1}{\sqrt{\alpha\pi}}\frac{\Gamma\left(\frac{\alpha+1}{2}\right)}{\Gamma\left(\frac{\alpha}{2}\right)}\alpha^{(\alpha-1)/2}\left[4\frac{\Gamma\left(\frac{\alpha+1}{2}\right)\alpha^{(5+\alpha)/2}(\alpha+1)^2}{\Gamma\left(\frac{\alpha}{2}\right)\sqrt{\alpha\pi}(\alpha+2)^4}\right]^{-\frac{\alpha}{\alpha+4}}n^{\frac{4}{4+\alpha}}.$$

Table 2 lists the optimal number of upper-order statistics m supposing that $n=5,000$, 10,000 and 20,000, to reflect the case of absolute values, that is, when both tails are combined in the case of symmetric distributions. For the stable class,

TABLE 2 Optimal m for Hill estimator

	Stable				Student			
α	**1.5**	**1.7**	**1.8**	**1.9**	**2.5**	**3**	**4**	**6**
$n = 5,000$	92	47	31	17	121	89	55	30
$n = 10,000$	146	74	49	27	186	132	78	40
$n = 20,000$	232	117	78	43	285	196	110	53

the number m decreases to zero as the tails become thinner (one approaches the normal distribution). At the other end of the spectrum (the case of Cauchy distribution), m approaches n.

Early literature on the estimation of the stable distribution parameters, (see, e.g., DuMouchel, 1983), suggests that the choice of sample fraction to estimate the tail index should be about 10%. The 10% rule is often followed in the literature when applied to the Hill estimator (see, e.g., Quintos, Fan, & Phillips, 2001; Ibragimov, Ibragimov, & Kattuman, 2013). Sometimes an even larger number m is chosen such that the tail region comprises 30–40% of the entire sample size (see, e.g., Dufour & Kim, 2010; Mittnik, Paolella, & Rachev, 1998). However, our results suggest that one should use a relatively small number rather than a large number of upper-order statistics.

For the Student t class, the closer one gets to the normal distribution (higher degrees of freedom), the fewer observations one should use. So for both models, it holds that the closer the law is to the normal distribution, the fewer order statistics should be used in estimating α. The $n = 5,000$ case is comparable to the 20-year series of daily financial returns that is used in the later sections.

3 | THE SHAPES OF HILL PLOTS

In this section, we first perform simulations for both estimators and compare the Hill plots for the simulated data of the competing distributions.

3.1 | Simulation experiments: Stable distribution

We use a simulated sample size $n = 10^4$ and replicate this 500 times to obtain an average estimate of the tail index. For any given number of order statistics used in the Hill estimator, the average of the estimates across 500 samples proxies the mean of the estimator. Thus, by plotting the averaged estimate against the number of order statistics used, we capture the bias of the Hill estimator.

Recall that the Hill estimator for α of the stable model is upward biased. The thick continuous curves in Figure 2 are the averaged Hill plots for two particular α values: 1.7 and 1.9. For estimates of each α, the Hill plot first displays an upward bias, but eventually turns down, producing a hump-shaped graph. Notably, the hump shape is clearly discernible for the tail index value of 1.9.

Given the expression for the bias in Equation 4, one can correct the bias of the Hill estimator as follows:

$$\hat{u}\left(s_n\right) = u\left(s_n\right) + \frac{b\beta}{\alpha(\alpha+\beta)} s_n^{-\beta}. \tag{9}$$

The dashed black curves in Figure 2 are the second-order bias-corrected Hill plots. The Hill bias is not important at low values of α and becomes more pronounced as α tends to 2 (normal distribution). The second-order bias corrections improve the Hill estimates. The bias-corrected plots, however, still exhibit a hump shape, particularly for the case $\alpha = 1.9$.

Using the optimal number of upper-order statistics presented from Table 2, we report the bias and RMSE of α (for a sample size of 10,000 and 500 rounds of simulations) before and after second-order bias adjustment in Table 3. One observes that for both stable distributions, although the second-order bias adjustment reduces the bias, the bias-adjusted Hill estimates are still upward biased. The higher-order terms of the tail expansion are responsible for this phenomenon, as we argue in Section 4.

3.2 | Simulation experiments: Student *t* distribution

Similarly, for the case of Student t distribution, we also perform simulations with sample size $n = 10^4$. There are 500 replications to obtain an average estimate of tail index α for different number of order statistics. Recall that the Hill

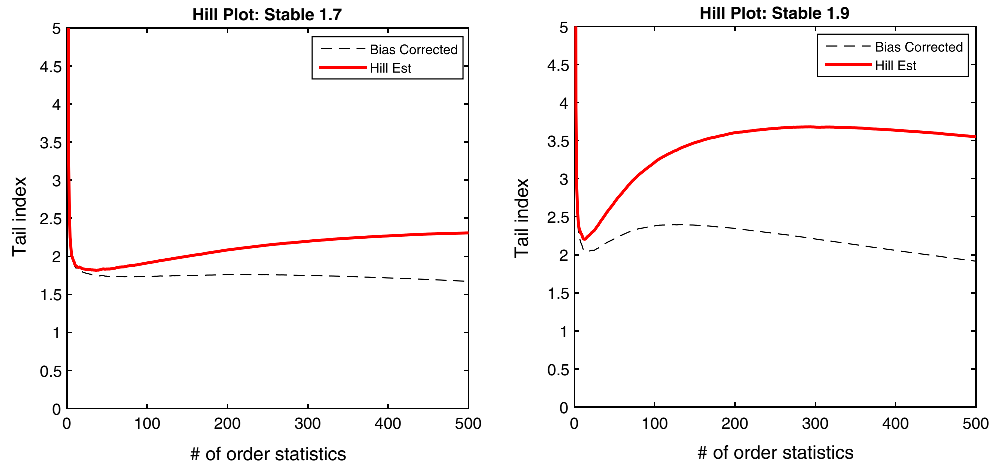

FIGURE 2 Hill plot and second-order bias correction: stable [Colour figure can be viewed at wileyonlinelibrary.com]

TABLE 3 Bias and RMSE comparison before and after second-order bias adjustment

	Bias	RMSE	Bias adjusted	RMSE adjusted
Stable 1.7	0.167	0.274	0.033	0.189
Stable 1.9	0.452	0.689	0.173	0.431
Student *t* 3	−0.202	0.311	0.044	0.283
Student *t* 4	−0.420	0.575	0.087	0.522

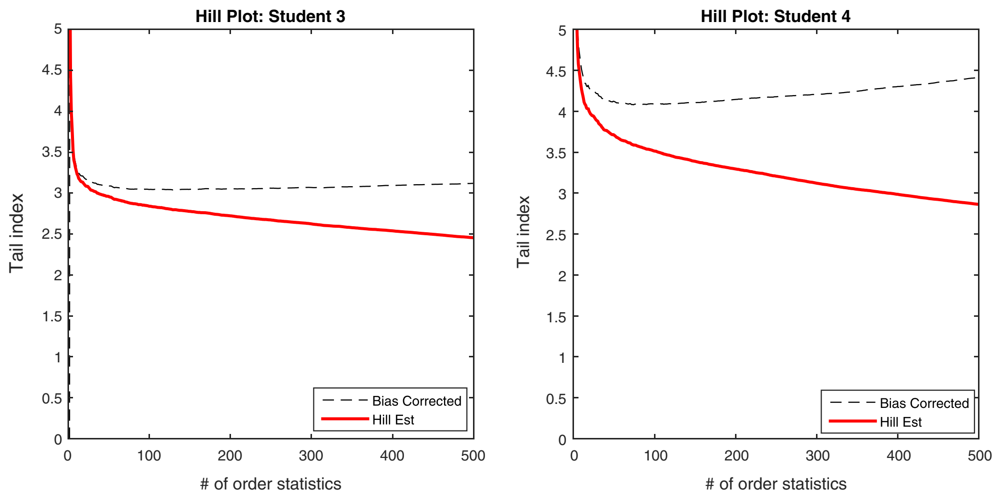

FIGURE 3 Hill plot and second-order bias correction: Student t [Colour figure can be viewed at wileyonlinelibrary.com]

estimator for α of the Student t model is downward biased (upward for $1/\alpha$). The thick continuous curves in Figure 3 depict this bias behavior with Hill plots for the cases $\alpha = 3$ and 4.

Again, we can apply the bias correction from Equation 4; see the dashed black curves in Figure 3. We observe that the Hill estimates are improved by the second-order bias correction in the upper tail region. However, the bias correction leads to an overestimation of the tail index if one uses too many order statistics.

We use the optimal number of upper-order statistics presented in Table 2 to report the bias and RMSE of α (for sample size of 10,000 and 500 rounds simulations) before and after second-order bias adjustment (see Table 3). One observes that the bias adjustment improves both bias and RMSE. Note that the bias sign flipped after the bias adjustment for both distributions. This is related to the higher-order terms in the tail expansion in Section 4.

4 | HIGHER-ORDER THEORY

From the simulation experiments in Subsections 3.1 and 3.2, we observed that the Hill plots of the stable distribution and Student distribution exhibit different shapes. The second-order bias correction improves the Hill estimates. The bias-corrected Hill plots for the stable distribution ($\alpha = 1.7, 1.9$) nevertheless retain the hump shape, whereas it leads to overestimation for the Student t distribution ($\alpha = 3, 4$), as a larger number of upper statistics are used. We provide a theoretical explanation for these phenomena using a higher-order tail approximation.

Consider the third-order tail approximation

$$F(x) = 1 - ax^{-\alpha}\left[1 + bx^{-\beta} + cx^{-\gamma} + o\left(x^{-\gamma}\right)\right], \text{ as } x \to \infty, \tag{10}$$

where γ is the third-order index and c is the third-order scale parameter. The values for the two classes of distribution are given in Table 4. The detailed derivations can be found in the Appendices A1 and A2. In what follows, the corresponding asymptotic bias becomes

$$E_A\left[u\left(s_n\right) - \frac{1}{\alpha}\right] = -\frac{b\beta}{\alpha(\alpha+\beta)}s_n^{-\beta} - \frac{c\gamma}{\alpha(\alpha+\gamma)}s_n^{-\gamma} + o\left(s_n^{-\gamma}\right). \tag{11}$$

Under the second-order tail approximation, we showed that the sign of the second-order scale parameter b determines the sign of Hill bias. Recall that $b > 0$ for stable distribution, whereas $b < 0$ for Student t distribution. Under the third-order tail approximation, both the second-order scale parameter b and third-order scale parameter c determine the bias sign. Note from Table 4 it follows that $\gamma > \beta$ for both classes of distribution.

To compare the importance of the bias terms with respect to the threshold s_n or number of upper-order statistics, we consider the ratio between the third- and second-order bias terms

$$\tau = \frac{b}{c}\frac{\beta(\alpha+\gamma)}{\gamma(\alpha+\beta)}s_n^{\beta-\gamma}.$$

Figure 4 displays the ratio τ against the number of order statistics m; that is, increasing m corresponds to a decrease in s. For both types of distribution, the second-order bias term is the dominant part in the tail, while it becomes less important

TABLE 4 Parameters under the third-order tail approximation

	Stable	Student
γ	2α	4
c	$\frac{1}{6}\frac{\Gamma(3\alpha)}{\Gamma(\alpha)}\frac{\sin(\frac{3\alpha\pi}{2})}{\sin(\frac{\alpha\pi}{2})}$	$\frac{\alpha^3}{8}\frac{(\alpha+1)(\alpha+3)}{(\alpha+4)}$

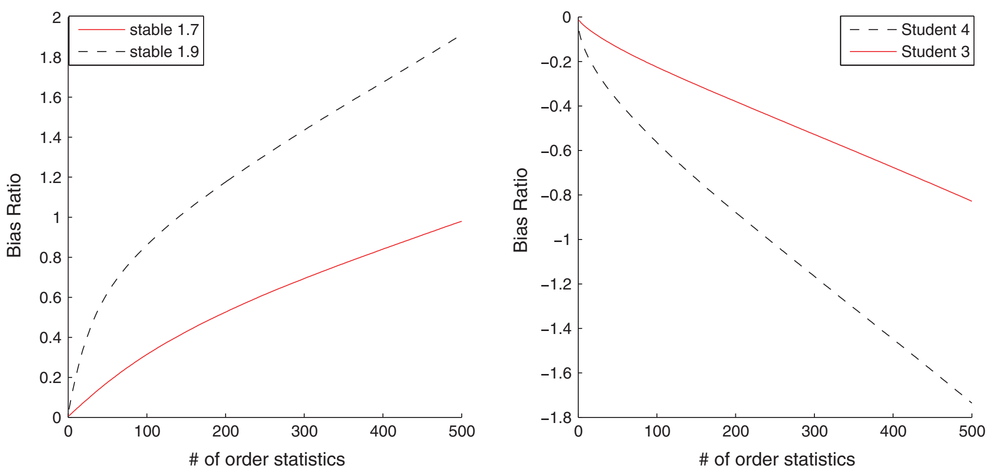

FIGURE 4 Comparison between the third-order and second-order bias terms [Colour figure can be viewed at wileyonlinelibrary.com]

as more and more upper-order statistics are used. This phenomenon is more pronounced when the tail index is closer to 2 for the stable case and at higher degrees of freedom for the Student t case.

For the stable distribution, note that $c > 0$ if $\alpha > \frac{4}{3}$, $c < 0$ as $\alpha < \frac{4}{3}$, and $c = 0$ for $\alpha = \frac{4}{3}$. Thus the second-order and third-order scale parameters share a positive sign in the "neighborhood" of the normal distribution, that is, if the characteristic exponent is close to 2. Therefore, the upward bias in the tail is enhanced by the third-order bias term, which induces the hump shape in the Hill plot. This explains why the second-order bias-corrected Hill plots for the cases $\alpha = 1.7$, 1.9 still exhibit a hump-shaped graph.

For the Student t distribution, the scale parameters $b > 0$ and $c < 0$. Although the bias is to some extent balanced in the tail area, the third-order bias term becomes more important as a larger number of upper-order statistics are used. The upward bias from the third-order term explains the overestimation of the tail index for the second-order bias-corrected plot close to the center (high m, low s).

Similarly, the higher-order terms can still affect the bias correction in the third-order tail approximation. We derive the expansion for both types of distribution. The tail of the two types of distribution is as follows:

$$F(x) = 1 - ax^{-\alpha}\left[1 + bx^{-\beta} + cx^{-\gamma} + \sum_{i=4}^{I} C_i x^{-\alpha_i} + o\left(x^{-\alpha_I}\right)\right], \quad \alpha_i > 0, \quad \text{as } x \to \infty, \tag{12}$$

where the higher-order index α_i and scale parameter C_i $(i > 2)$ values are given in Table 5.

The detailed derivations can be found in the Appendix. In what follows, the corresponding asymptotic bias is

$$E_A\left[u\left(s_n\right) - \frac{1}{\alpha}\right] = -\frac{b\beta}{\alpha(\alpha+\beta)} s_n^{-\beta} - \frac{c\gamma}{\alpha(\alpha+\gamma)} s_n^{-\gamma} - \sum_{i=4}^{I} \frac{C_i \alpha_i}{\alpha(\alpha+\alpha_i)} s_n^{-\alpha_i} + o\left(s_n^{-\alpha_I}\right). \tag{13}$$

From Equation 13, the signs of all the scale parameters jointly determine the sign of Hill bias. Particularly for stable distribution, the term $(-1)^{i-1}\sin(\frac{i\alpha\pi}{2})$ determines the sign of ith-order scale parameter C_i:

- If i is odd, $\sin(\frac{i\alpha\pi}{2}) > 0$ as $\alpha \to 2$;
- if i is even, $-\sin(\frac{i\alpha\pi}{2}) > 0$ as $\alpha \to 2$.

This implies that for α close to 2 the first Ith-order scale parameters share a positive sign. Hence, for distributions with a tail index α close to 2, and if s_n is sufficiently large, the first I bias terms combine and dominate over the higher-order bias terms. The large upward bias in the tail induces the hump shape in the Hill plot; moreover, the hump is more pronounced the closer α is to 2.

For the Student t distribution, the scale parameters flip signs: if i is odd $C_i > 0$, and when i is even $C_i < 0$. The bias is balanced to some extent when including more high-order bias terms, but the second-order scale parameter dominates. The Hill plot therefore declines more or less monotonically, yielding a downward bias.

We formalize the above results in the following proposition. The proof also explains the asymptotic behavior in the left-hand corner of the Hill plots.

Proposition 2. *Consider the class of random variables with distribution that satisfy Equation 10. As one uses more and more of the higher-order statistics in the calculation of the Hill estimator:*

- *For a symmetric stable distributed random variable $\alpha \in (1, 2)$, the Hill plot of $\hat{\alpha}$ is hump shaped and upward biased.*
- *For a Student t distributed random variable with $\alpha > 2$, the Hill plot of $\hat{\alpha}$ is declining and downward biased.*

Proof. see Appendix A3. □

TABLE 5 Parameters in the tail expansion

	Stable	Student
α_i	$(i-1)\alpha$	$2(i-1)$
C_i	$\frac{(-1)^{i-1}}{i!}\frac{\Gamma(i\alpha)}{\Gamma(\alpha)}\frac{\sin(\frac{i\alpha\pi}{2})}{\sin(\frac{\alpha\pi}{2})}$	$\frac{(-1)^{i-1}\alpha^i}{(i-1)!}\frac{(\alpha+1)\cdots(\alpha+2i-3)}{2^{i-1}(\alpha+2i-2)}$
sign(C_i)	"+" if $\alpha \in (\frac{2(i-1)}{i}, 2)$	$(-1)^{i-1}$

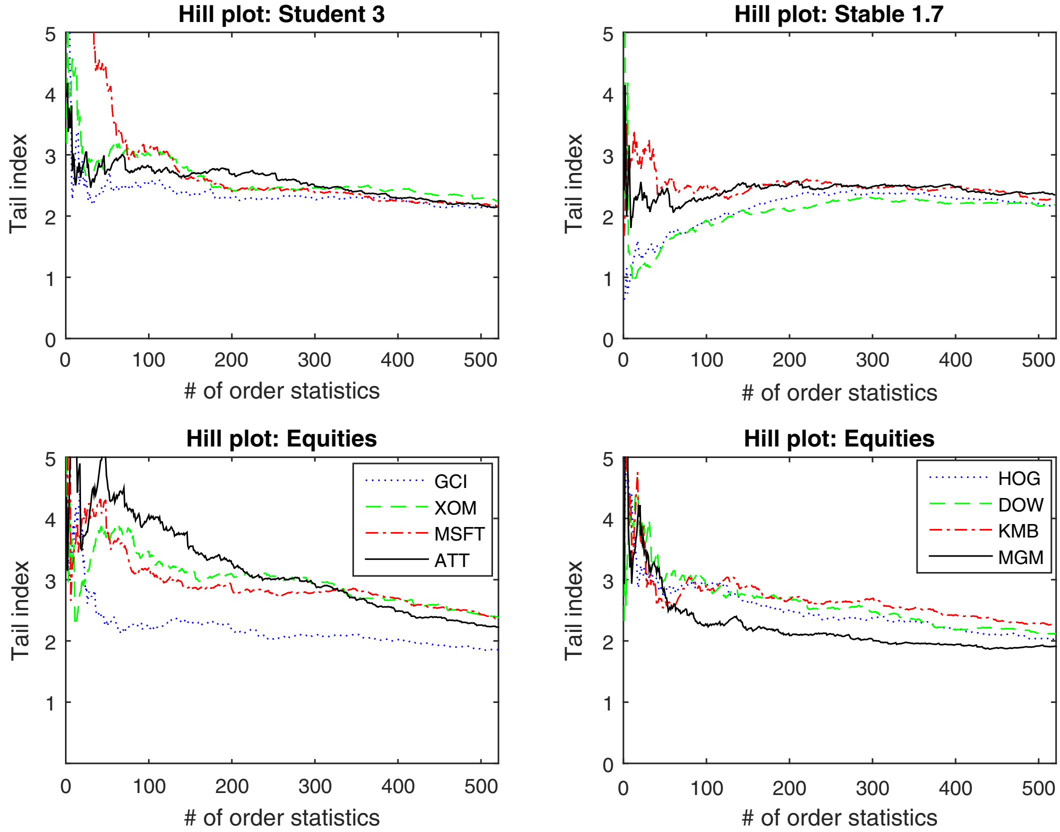

FIGURE 5 Hill plot: equities; January 1991 to December 2010 [Colour figure can be viewed at wileyonlinelibrary.com]

5 | AN EMPIRICAL STUDY

One can be agnostic about which model (distribution) to use. Which model is appropriate depends, in part, on the questions that one asks. In the empirical finance literature, both sum-stable and Student *t* distributions have been used to model the unconditional distribution of the returns. With the above apparatus, one can at least try to discriminate between the two models by using a small number of higher-order statistics in computation of the Hill estimator, as this keeps the bias of both models small. Moreover, one can try to partially correct for the bias by means of a second-order correction. We apply these ideas to some stock market data. Furthermore, we also apply the sign estimator by Ferreira and Vries (2004) to determine which way the bias goes. As we argued above, the two models differ regarding the sign of the second-order scale parameter. The sign estimator directly tries to capture this feature of the data.

5.1 | Equity

We randomly choose eight equities[4] from different sectors in the US market and four equity indices—S&P, DAX , Nikkei, and Hang Seng—from January 1991 to December 2010, giving 5,217 daily returns in total. The data are collected from DATASTREAM. We also simulate four independent series of observations that follow a stable distribution with $\alpha = 1.7$ and Student *t* distribution with $\alpha = 3$. The simulated sample size is 5,000. The corresponding Hill plots are displayed in Figures 5 and 6. For the individual equities, the general tendency appears to be a downward movement as more of the higher number of order statistics are included, but one could also detect some hump-shape behavior in the case of XOM and ATT. The Hill plots for the equity indices show a clear downward tendency as the number of order statistics increases, as in the Student *t* case.

[4]The equities are: Gannett CO. (GCI), Exxon Mobile (XOM), Microsoft (MSFT), AT&T (ATT), Harley-Davidson (HOG), DOW, Kimberly-Clark(KMB), and MGM Resorts International (MGM).

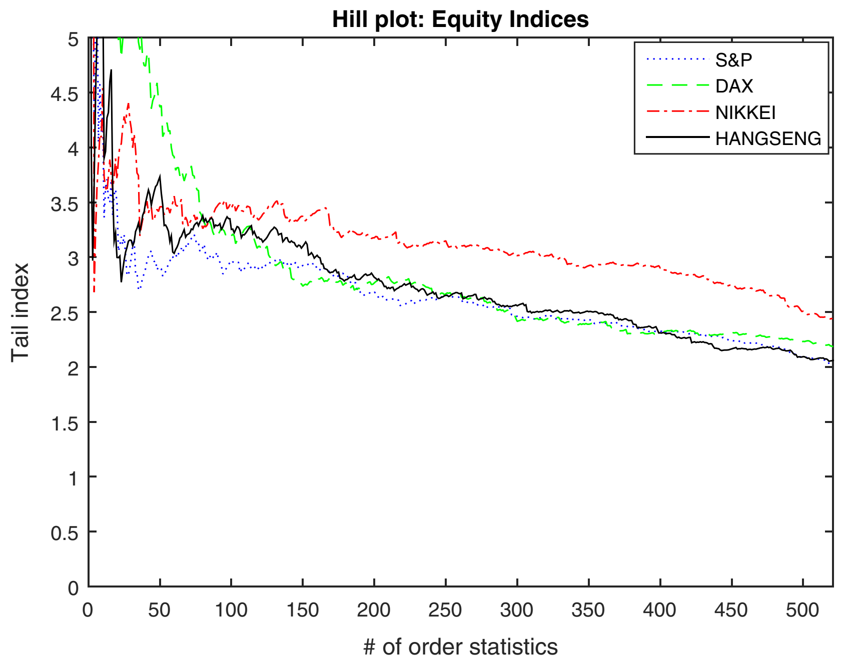

FIGURE 6 Hill plot: indices; January 1991 to December 2010 [Colour figure can be viewed at wileyonlinelibrary.com]

TABLE 6 Financial asset returns sign(b)

	CGI	XOM	MSFT	ATT	HOG	DOW	KMB	MGM	S&P	DAX	NIK	HNG
Whole period	−	−	−	−	−	−	−	−	−	−	−	−
Subperiod 1	−	−	−	−	+	−	−	−	−	−	−	−
Subperiod 2	−	+	−	−	−	+	−	−	−	−	−	−
Subperiod 3	−	−	−	−	−	−	+	−	−	−	−	−
Subperiod 4	−	−	−	−	+	−	−	+	−	−	−	−
Subperiod 5	−	−	−	−	−	−	+	−	−	−	−	−

Note. This table presents the sign estimates of b for eight equities and four equity indices.

5.2 | A sign estimator

As discussed in the previous section, the key difference between the two distributions is the sign of the second-order scale parameter b, which determines the sign of the bias of the Hill estimator. Ferreira and Vries (2004) develop an estimator for the sign of b. The estimator is defined as follows:

$$\widehat{\text{sign}} = \text{sign}\left(\frac{1}{b_n - a_n + 1}\sum_{i=a_n}^{b_n} \hat{u}_i - \hat{u}_{c_n}\right), \tag{14}$$

where a_n, b_n, and c_n are the intermediate sequences such that

$$a_n < b_n \leq c_n, \ \forall n, \ \frac{a_n}{b_n} \to \theta \in [0, 1).$$

The sign estimator is consistent if b_n is such that $a(\frac{n}{b_n})\sqrt{b_n} \to \infty$, as $n \to \infty$.

Empirically, we follow the suggestion by Ferreira and Vries (2004) to choose $a_n = \log(n)$. Moreover, as introduced in Section 2, the hump shape of the Hill plot for the stable distribution behaves differently with respect to different tail index values. To detect the hump shape we choose $b_n = c_n = 0.4n/\log(n)$.

We apply the sign estimator to these eight equities and four stock indices. The top row in Table 6 shows that the second-order scale parameter b exhibits a negative sign for all asset returns if the entire sample period is used. This yields the following interpretation. Since the sign estimator is a consistent estimator and a plus or a minus are the only possible outcomes, the evidence suggests that the asymptotic expansion (Equation 1) for equity returns has $b < 0$.

For the application of the binomial test for the sign estimates, it is important to have independent samples. Since equity returns are driven by common factors, these are not independent. Therefore we split the sample for each stock into five subsamples and apply the binomial test to these subsamples, rather than applying the test across the different equities. Since the question is whether the stable or Student law applies, we take p in the binomial distribution equal to 0.5. Suppose the null hypothesis is that the sign is negative; that is, the probability that the sign of b negative is greater than 0.5. With five trials, the approximate 95% critical region for the one-tailed test implies rejection in case four or or five pluses are recorded. Given that the maximum number of recorded pluses is two, the null of a negative sign is not rejected. As a consequence, these equity returns more likely follow a Student t distribution than a stable distribution.

6 | GARCH

Another stylized fact of the financial asset returns is volatility clustering. The ARCH and GARCH models are designed to capture these volatility clusters. Meanwhile, the stationary solution of these stochastic processes are distributions that also exhibit heavy-tailedness even if the innovations do not possess this feature. In this section, we briefly explore the tail behavior of the stationary solution of the GARCH(1, 1) model and compare the tail shape of that solution with those of the Student and stable laws. Consider the GARCH(1, 1) stochastic process

$$\begin{aligned} Y_t &= N_t H_t, \\ H_t^2 &= \omega + \delta Y_{t-1}^2 + \tau H_{t-1}^2, \end{aligned}$$

where N_t are i.i.d., $\omega > 0, \delta > 0, \tau > 0$ and $\delta + \tau < 1$. Combine the two equations to obtain a stochastic difference equation:

$$\begin{aligned} H_t^2 &= \omega + \delta N_{t-1}^2 H_{t-1}^2 + \tau H_{t-1}^2 \\ &= \omega + Q_{t-1} H_{t-1}^2 \end{aligned} \tag{15}$$

say, and where

$$Q_{t-1} = \tau + \delta N_{t-1}^2$$

is a strictly positive random coefficient.

According to the theorem of Kesten (1973), the stationary solution of the process Y_t^2 is heavy-tail distributed (see, e.g., Davis & Mikosch, 2009; de Haan et al., 1989; Mikosch & Starica, 2000). The more surprising part of Kesten (1973) is that the result may even hold if the innovations N come from a distribution that does not exhibit heavy tails; in fact, the distribution may even have a bounded support. Conjecture that the tail approximation of the stationary distribution accords with the expansion in the Equation 1. In the case of ARCH(1), the first-order tail index is obtained by de Haan et al. (1989), while Goldie (1991) obtains the boundaries of the first-order scale parameter. Furthermore, for the GARCH(1, 1) model, Sun and Zhou (2014) show that if one tries to recover the innovations from a GARCH(1, 1) process one likely finds that these are heavy tailed even if they are not, due to an inherent bias in the recovery process.

Assume that the tail approximation of the stationary process is as indicated in Equation 1, then we have the following lemma.

Lemma 3. *For the GARCH(1, 1) process with i.i.d. normally distributed innovations, suppose the tail approximation of Y_t^2 satisfies Equation 1, where κ is the solution to $E[Q^\kappa] = 1$. Then $\beta = 1$ and*

$$b = \frac{\kappa\omega}{1 - E[Q^{\kappa+1}]} < 0. \tag{16}$$

Proof. see Appendix A4. □

The asymptotic bias of the Hill estimator is again determined by the sign of second-order scale parameter b. From Lemma 3, the second-order scale b has a negative sign, implying that the Hill estimate of the GARCH(1, 1) process is downward biased. This holds regardless of whether the innovations are heavy tailed or not. We perform simulations for the GARCH(1, 1) by taking the sample size $n = 10^4$ replicated 500 times to obtain an average estimate for tail index $\alpha = 2\kappa$

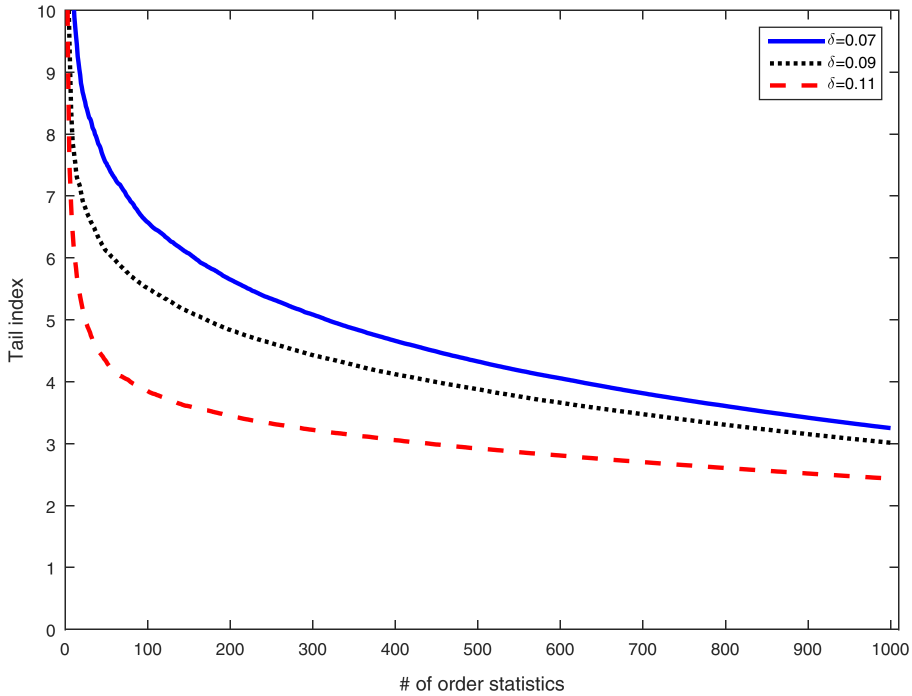

FIGURE 7 Hill Plot: GARCH(1,1). Note: We perform simulations for the GARCH(1,1) by taking the sample size $n = 10^4$ replicated 500 times to obtain an average estimate for tail index $\alpha = 2\kappa$ at three special cases $\delta = 0.07, 0.09, 0.11$ given $\tau = 0.88$ [Colour figure can be viewed at wileyonlinelibrary.com]

for three special cases $\delta = 0.07, 0.09, 0.11$, given $\tau = 0.88$ (see Figure 7). Clearly, the estimates show a downward bias in comparison with their theoretical values.[5]

7 | CONCLUSION

To summarize, under the null of a symmetric stable distribution or a Student *t* distribution, the optimal number of order statistics for the Hill estimator is quite low. This contrasts with the practice in several empirical studies. For the stable model, the Hill plot, after first being upward biased, eventually turns down again, delivering a hump-shaped graph. The hump shape is more pronounced as the stable class approaches the normal distribution. This occurs because more high-order scale parameters in the tail expansion share the positive sign as α approaches 2. Due to the hump shape there are two regions that cover the true value of the tail index. The stable law often gives tail estimates that are larger than 2, conflicting with the stable hypothesis but, due to the downward bias in the case that the Student *t* model is the correct law, just the opposite occurs.

This paper proposes to exploit the shape of the bias stemming from the Hill plot to discriminate between these two classes of distribution. We further propose to apply the sign estimator by Ferreira and Vries (2004) as a test of the two alternative models. These ideas are subsequently applied to the case of eight equities and four equity indices. The results indicate that these financial asset returns are more likely to resemble a Student *t* distribution than a stable distribution. Lastly, we show that the stationary solution of a GARCH(1, 1) process implies a bias similar to the Student *t* model.

ACKNOWLEDGMENTS

Pengfei Sun was a Marie Curie Fellow at Erasmus University Rotterdam. The research leading to these results has received funding from the European Community's Seventh Framework Programme FP7-PEOPLE-ITN-2008 under grant agreement number PITN-GA-2009-237984 (project name: RISK). The funding is gratefully acknowledged. Casper de Vries is grateful to the ESI at Chapman University for its hospitality during a sabbatical leave. We would like to thank Chen Zhou for helpful discussion and suggestions. We benefited from reports by two anonymous referees.

[5]The corresponding tail index values are available from Sun and Zhou (2014). For δequal to 0.07, 0.09, and 0.11 given $\tau = 0.88$, the corresponding αvalues are respectively 13.98, 7.90, and 3.70.

OPEN RESEARCH BADGES

This article has earned an Open Data Badge for making publicly available the digitally-shareable data necessary to reproduce the reported results. The data is available at [http://qed.econ.queensu.ca/jae/2018-v33.5/sun-de_vries/].

REFERENCES

Baek, C., Pipiras, V., Wendt, H., & Abry, P. (2009). Second order properties of distribution tails and estimation of tail exponents in random difference equations. *Extremes*, *12*(4), 361–400.

Blattberg, R., & Gonedes, N. (1974). A comparison of the stable and student distributions as statistical models for stock prices. *Journal of Business*, *47*(2), 244–280.

Campbell, J., Lo, A., & MacKinlay, A. (1997). *The econometrics of financial markets*. Princeton, NJ: Princeton University Press.

Davis, R., & Mikosch, T. (2009). Extreme value theory for GARCH processes. In Mikosch, T, Kreiß, J. P, Davis, R, & Andersen, T (Eds.), *Handbook of Financial Time Series* (pp. 187–200).Berlin, Germany: Springer.

de Haan, L., Resnick, S., Rootzén, H., & de Vries, C. (1989). Extremal behaviour of solutions to a stochastic difference equation with applications to ARCH processes. *Stochastic Processes and their Applications*, *32*(2), 213–224.

DuMouchel, W. (1983). Estimating the stable index α in order to measure tail thickness: a critique. *Annals of Statistics*, *11*(4), 1019–1031.

Dufour, J., & Kim, J. (2010). Exact inference and optimal invariant estimation for the stability parameter of symmetric alpha-stable distributions. *Journal of Empirical Finance*, *17*(2), 180–194.

Fama, E., & Miller, M. (1972). *The theory of finance*. New York, NY: Holt, Rinehart and Winston.

Feller, W. (1971). An introduction to probability theory and its applications. (Vol. 2, 2nd ed.). New York:John Wiley and Sons Inc.

Ferreira, A., & Vries, C. (2004). Optimal confidence intervals for the tail index and high quantiles. (*Technical report 04-090/2*). Amsterdam, Netherlands: Tinbergen Institute.

Goldie, C. (1991). Implicit renewal theory and tails of solutions of random equations. *Annals of Applied Probability*, *1*(1), 126–166.

Goldie, C., & Smith, R. (1987). Slow variation with remainder: Theory and applications. *Quarterly Journal of Mathematics*, *38*(1), 45–71.

Hagerman, R. (1978). More evidence on the distribution of security returns. *Journal of Finance*, *33*(4), 1213–1221.

Hall, P., & Welsh, A. (1984). Best attainable rates of convergence for estimates of parameters of regular variation. *Annals of Statistics*, *12*(3), 1079–1084.

Hill, B. (1975). A simple general approach to inference about the tail of a distribution. *Annals of Statistics*, *3*(5), 1163–1174.

Ibragimov, M., Ibragimov, R., & Kattuman, P. (2013). Emerging markets and heavy tails. *Journal of Banking and Finance*, *37*(7), 2546–2559.

Ibragimov, I., & Linnik, Y. (1971). *Independent and stationary sequences of random variables*. Groningen, Netherlands: Wolters-Noordhoff.

Jorion, P. (1996). Risk2: Measuring the risk in value at risk. *Financial Analysts Journal*, *52*(6), 47–56.

Kesten, H. (1973). Random difference equations and renewal theory for products of random matrices. *Acta Mathematica*, *131*(1), 207–248.

Lau, H., & Lau, A. (1993). The reliability of the stability-under-addition test for the stable-paretian hypothesis. *Journal of Statistical Computation and Simulation*, *48*(1–2), 67–80.

Mandelbrot, B. (1963). The variation of certain speculative prices. *Journal of Business*, *36*(4), 394–419.

McCulloch, J. (1997). Measuring tail thickness to estimate the stable index α: A critique. *Journal of Business and Economic Statistics*, *15*(1), 74–81.

Mikosch, T., & Starica, C. (2000). Limit theory for the sample autocorrelations and extremes of a GARCH(1, 1) process. *Annals of Statistics*, *28*(5), 1427–1451.

Mittnik, S., Paolella, M., & Rachev, S. (1998). A tail estimator for the index of the stable paretian distribution? *Communications in Statistics: Theory and Methods*, *27*(5), 1239–1262.

Nolan, J. (2003). *Stable distributions: Models for heavy-tailed data*. Birkhauser: Basel, Switzerland.

Paolella, M. (2016). Stable-GARCH models for financial returns: Fast estimation and tests for stability. *Econometrics*, *4*(2), 25.

Quintos, C., Fan, Z., & Phillips, P. (2001). Structural change tests in tail behaviour and the asian crisis. *Review of Economic Studies*, *68*(3), 633–663.

Resnick, S., & Stărică, C. (1998). Tail index estimation for dependent data. *Annals of Applied Probability*, *8*(4), 1156–1183.

Shao, Q., Yu, H., & Yu, J. (2001). Do stock returns follow a finite variance distribution? *Annals of Economics and Finance*, *2*, 467–486.

Stoyanov, S., Rachev, S., Racheva-Iotova, B., & Fabozzi, F. (2011). Fat-tailed models for risk estimation. (Working Paper Series in Economics 30), Karlsruhe Institute of Technology, Karlsruhe, Germany.

Sun, P., & Zhou, C. (2014). Diagnosing the distribution of GARCH innovations. *Journal of Empirical Finance*, *29*, 287–303.

Tsionas, E. (2002). Likelihood-based comparison of stable paretian and competing models: Evidence from daily exchange rates. *Journal of Statistical Computation and Simulation*, *72*(4), 341–353.

How to cite this article: Sun P, de Vries CG. Exploiting tail shape biases to discriminate between stable and student *t* alternatives. *J Appl Econ*. 2018;33:708–726. https://doi.org/10.1002/jae.2628

APPENDIX A

A.1 Tail approximation: Stable distribution

Ibragimov and Linnik (1971) provide the following tail approximation of the probability density function (p.d.f.) of a symmetric stable distribution:

$$f(x) = \frac{1}{\pi}\sum_{i=1}^{I}(-1)^{i+1}\frac{\Gamma(i\alpha+1)}{i!x^{i\alpha+1}}\sin\left(\frac{i\alpha\pi}{2}\right) + o(x^{-I\alpha-1}),\ I \geqslant 2.$$

By integration, one obtains

$$\begin{aligned}1-F(x) &= \int_x^\infty f(t)\mathrm{d}t = \frac{1}{\pi}\sum_{i=1}^{I}(-1)^{i}\frac{\Gamma(i\alpha)}{i!x^{i\alpha}}\sin\left(\frac{i\alpha\pi}{2}\right) + o(x^{-I\alpha}) \\ &= \frac{1}{\pi}\Gamma(\alpha)\sin\left(\frac{\alpha\pi}{2}\right)x^{-\alpha}\left[1+\sum_{i=2}^{I}\frac{(-1)^{i-1}}{i!}\frac{\Gamma(i\alpha)}{\Gamma(\alpha)}\frac{\sin(\frac{i\alpha\pi}{2})}{\sin(\frac{\alpha\pi}{2})}x^{-(i-1)\alpha}\right] + o(x^{-I\alpha}).\end{aligned}$$

Thus the ith-order tail index $\alpha_i = (i-1)\alpha$ and the ith-order scale parameter has the following form:

$$C_i = \frac{(-1)^{i-1}}{i!}\frac{\Gamma(i\alpha)}{\Gamma(\alpha)}\frac{\sin(\frac{i\alpha\pi}{2})}{\sin(\frac{\alpha\pi}{2})},\quad i=2,\cdots,I.$$

A.2 Tail approximation: Student *t* distribution

The p.d.f. of the Student t distribution is as follows:

$$f(x) = \frac{\Gamma(\frac{\alpha+1}{2})}{\sqrt{\alpha\pi}\Gamma(\frac{\alpha}{2})}\left[1+\frac{x^2}{\alpha}\right]^{-\frac{\alpha+1}{2}}.$$

Let $y=\frac{\alpha}{x^2}, z=\frac{\alpha+1}{2}$, as $x\to\infty (y\to 0)$. By a Taylor expansion around zero, we have that

$$\begin{aligned}g(y) &= \frac{\Gamma(\frac{\alpha+1}{2})}{\sqrt{\alpha\pi}\Gamma(\frac{\alpha}{2})}\left[1+\frac{1}{y}\right]^{-z} = \frac{\Gamma(\frac{\alpha+1}{2})}{\sqrt{\alpha\pi}\Gamma(\frac{\alpha}{2})}y^z[1+y]^{-z} \\ &= \frac{\Gamma(\frac{\alpha+1}{2})}{\sqrt{\alpha\pi}\Gamma(\frac{\alpha}{2})}y^z\left[1+(-z)y+\frac{(-z)(-z-1)}{2!}y^2+\cdots\right. \\ &\quad \left.+\frac{(-z)\cdots(-z-n+2)}{(n-1)!}y^{n-1}+o(y^{n-1})\right].\end{aligned}$$

This implies the following tail approximation for the p.d.f.:

$$\begin{aligned}f(x) = &\frac{\Gamma(\frac{\alpha+1}{2})}{\sqrt{\alpha\pi}\Gamma(\frac{\alpha}{2})}\alpha^{\frac{\alpha+1}{2}}x^{-(\alpha+1)}[1-\alpha\frac{\alpha+1}{2}x^{-2}+\frac{\alpha^2}{2!}\frac{(\alpha+1)(\alpha+3)}{4}x^{-4}+\cdots \\ &+\frac{\alpha^{n-1}}{(n-1)!}\frac{(\alpha+1)\cdots(\alpha+2n-3)}{2^{n-1}}x^{-2(n-1)}+o(x^{-2(n-1)})].\end{aligned}$$

Upon integrating, we obtain the tail expansion of the c.d.f.:

$$\begin{aligned}1-F(x) = \int_x^\infty f(t)\mathrm{d}t = &\frac{\Gamma(\frac{\alpha+1}{2})}{\sqrt{\alpha\pi}\Gamma(\frac{\alpha}{2})}\alpha^{\frac{\alpha-1}{2}}x^{-\alpha}[1-\frac{\alpha^2(\alpha+1)}{2(\alpha+2)}x^{-2}+\frac{\alpha^3(\alpha+1)(\alpha+3)}{8(\alpha+4)}x^{-4}+\cdots \\ &+\frac{(-1)^{n-1}\alpha^n}{(n-1)!}\frac{(\alpha+1)\cdots(\alpha+2n-3)}{2^{n-1}(\alpha+2n-2)}x^{-2(n-1)}+o(x^{-2(n-1)})].\end{aligned}$$

Hence the ith-order tail index $\alpha_i = 2(i-1)$ and the ith-order scale parameter is

$$C_i = \frac{(-1)^{n-1}\alpha^i}{(i-1)!}\frac{(\alpha+1)\cdots(\alpha+2i-3)}{2^{i-1}(\alpha+2i-2)}, \quad i = 2, \cdots, I.$$

A.3 Proof of Proposition 2

Let $F(x)$ be the distribution function and $f(x)$ be the probability density function of a random variable X. In this paper we consider the symmetric stable and Student t distributions with $F(0) = \frac{1}{2}$. Therefore we can focus on the positive tail only. We obtain the behavior of the Hill plot at the origin, that is, for s close to zero in the center of the distribution, and deep into the tail, that is, for s large. Consider the expected value of the Hill estimator, $u(s) = \widehat{1/\alpha}$, conditional on the threshold s:

$$E\left[u(s)\right] = \frac{1}{1-F(s)}\int_s^\infty \ln\left(\frac{x}{s}\right) f(x)\mathrm{d}x. \tag{A1}$$

We first obtain the shape of the Hill plot as $s\downarrow 0$. This corresponds to the far right-hand side in the plots of Figures 1, 2, and 3. For $s < 1$ the integral can be split as follows:

$$\begin{aligned}\frac{1}{1-F(s)}\int_s^\infty \ln\left(\frac{x}{s}\right) f(x)\mathrm{d}x &= \frac{-\ln s}{1-F(s)}\int_s^\infty f(x)\mathrm{d}x + \frac{1}{1-F(s)}\int_s^\infty \ln(x) f(x)\mathrm{d}x \\ &= -\ln s + \frac{1}{1-F(s)}\left\{\int_s^1 \ln(x) f(x)\mathrm{d}x + \int_1^\infty \ln(x) f(x)\mathrm{d}x\right\}.\end{aligned}$$

By the unimodality of both distribution functions, the density attains its maximum at zero, so that

$$\begin{aligned}\frac{1}{1-F(s)}\int_s^1 \ln(x) f(x)\mathrm{d}x &< \frac{f(0)}{1-F(s)}\int_s^1 \ln(x)\,\mathrm{d}x \\ &= \frac{f(0)}{1-F(s)}\left(x\ln x - x\right)\big|_s^1 \\ &= \frac{f(0)}{1-F(s)}\left(-1 - s\ln s + s\right).\end{aligned}$$

Since, by l'Hôpital's rule,

$$\lim_{s\downarrow 0} s\ln s = \lim_{s\downarrow 0}\frac{\ln s}{1/s} = \lim_{s\downarrow 0}\frac{1/s}{1/s^2} = 0$$

we get

$$\lim_{s\downarrow 0}\frac{f(0)}{1-F(s)}\left(-1 - s\ln s + s\right) = -\frac{f(0)}{1/2} < 0.$$

For the second part

$$\frac{1}{1-F(s)}\int_1^\infty \ln(x) f(x)\mathrm{d}x < \frac{1}{1-F(s)}\int_1^\infty x f(x)\mathrm{d}x < +\infty$$

since $\alpha > 1$ by assumption, so that $E\left[x\right]$ exists and is finite. We find that

$$\lim_{s\downarrow 0}\frac{1}{1-F(s)}\int_s^\infty \ln\left(\frac{x}{s}\right) f(x)\mathrm{d}x = \lim_{s\downarrow 0}\left(-\ln s\right) = +\infty.$$

Thus, for s close to zero, $E\left[\widehat{\alpha}\right] \simeq E\left[1/\widehat{1/\alpha}\right] = 1/E\left[u(s)\right] \to 0$. Furthermore, it is not hard to see that, for any $s > 0$, $\int_s^\infty \ln(x/s)\, f(x)dx > 0$ as $x/s \geq 1$. Hence, if s moves to the middle from the far right-hand side in Figures 1, 2, and 3, the Hill plot is increasing for both classes of distributions. Or, to put it differently, the Hill plot for both the stable and the Student t distribution declines to zero, as $s\downarrow 0$. That is, as more and more data from the center are used—that is, moving to the right in the right-hand side corner of the Hill plots—$\widehat{\alpha}$ is decreasing for both types of distribution.

Next, we consider the asymptotic nature of the Hill plot on the very left-hand side of the Hill plots in Figures 1, 2, and 3. At high s-levels only the very few highest-order statistics are included in the calculation of the Hill statistic. Ultimately, only the highest-order statistic $X_{(1)}$ is in the set A as defined in Equation 2. In this case the Hill estimator (Equation 3) collapses to

$$\begin{aligned} u(s) &= \frac{1}{m}\sum_{i=1}^{n} 1_{\{i\in A\}}\left(\log\frac{X_i}{s}\right)\Bigg| X_1 \geq s > X_2 \geq \cdots \geq X_n \\ &= \log\frac{X_1}{s}. \end{aligned}$$

From this it is immediate that if $s = X_1,\ u(s) = 0$. Thus, at this threshold level, the Hill plot explodes, as it plots the inverse $1/u(s)$. This explains the asymptotic nature of the estimator close to the origin for both types of distribution.

The picture is quite different more towards the middle. The stable distributions induce a hump shape, while the Student plot is monotonic, as can be seen most clearly in Figure 1. To explain this behavior, we study the first- and second-order derivatives at intermediate s -levels.

At intermediate s-levels the second-order tail expansion from Equation 1,

$$E[u(s)] \simeq \frac{1}{\alpha}\left[1 - \frac{\beta}{\alpha+\beta} b s^{-\beta}\right],$$

implies

$$E\left[\frac{1}{u(s)}\right] \simeq \frac{1}{E[u(s)]} \simeq \frac{\alpha}{1 - \frac{\beta}{\alpha+\beta} b s^{-\beta}}.$$

Recall that, for larger s levels,

$$\frac{1}{E[u(s)]} \simeq \alpha.$$

Taking the first-order derivative, we obtain

$$\frac{\partial(1/E[u(s)])}{\partial s} = -\frac{\alpha}{\left[1 - \frac{\beta}{\alpha+\beta} b s^{-\beta}\right]^2} \frac{\beta^2}{\alpha+\beta} b s^{-\beta-1},$$

while the second-order derivative is

$$\begin{aligned} \frac{\partial^2(1/E[u(s)])}{\partial s^2} &= \frac{\alpha(\beta+1)}{\left[1 - \frac{\beta}{\alpha+\beta} b s^{-\beta}\right]^2} \frac{\beta^2}{\alpha+\beta} b s^{-\beta-2} + \frac{2\alpha}{\left[1 - \frac{\beta}{\alpha+\beta} b s^{\beta}\right]^3} \left(\frac{\beta^2}{\alpha+\beta} b s^{-\beta-1}\right)^2 \\ &= \frac{\alpha(\beta+1)}{\left[1 - \frac{\beta}{\alpha+\beta} b s^{-\beta}\right]^2} \frac{\beta^2}{\alpha+\beta} b s^{-\beta-2} + \frac{2\alpha}{\left[1 - \frac{\beta}{\alpha+\beta} b^2 s^{\beta}\right]^3} \frac{\beta^4}{(\alpha+\beta)^2} b^2 s^{-2\beta-2}. \end{aligned}$$

For the Student t $\partial(1/E[u(s)])/\partial s > 0$, since $b < 0$. This explains the increasing nature of the Hill plot as one moves to the left, that is as s increases, in Figure 3. Furthermore, as $\partial^2(1/E[u(s)])/\partial s^2 < 0$ for the Student t data, the increase in the middle occurs at a decreasing rate, though this is not so easy to see from the figures. This effect is overtaken, however, by the asymptotic behavior at very high s-levels as very few order statistics are included.

The picture is different for the stable distribution as $b > 0$, so that $\partial(1/E[u(s)])/\partial s < 0$. We explained before that that at very low threshold levels the Hill plot for the stable distribution declines similarly to the plot for the Student distribution, but more towards the middle part the limiting behavior loses its force and the Hill plot starts to decline again as one moves further to the left (as s increases). This decline occurs at an increasing rate, since the second derivative $\partial^2(1/E[u(s)])/\partial s^2$ is positive. But then, at very high s-levels, the asymptotic effect of having just very few order statistics that determine the Hill statistic overtakes, and thus close to the origin the Hill plots as in Figure 2 increases without bound. Taken together, the behavior at the very endpoints and in the middle explains the characteristic hump shape for the stable data.

A.4 Proof of Lemma 3

Assume that the upper tail of the stationary distribution $F(x)$ of the stochastic process $\{H_t^2\}$ from Equation 15 satisfies expansion in Equation 1, that is,

$$\bar{F}(x) = \Pr\{H^2 > x\} = ax^{-\kappa}\left[1 + bx^{-\beta} + o(x^{-\beta})\right],\ x \to \infty. \tag{A2}$$

Let $f(x)$ be the density function of $F(x)$. Following Equation 15, we obtain

$$\begin{aligned}\Pr\{H_t^2 > x\} &= \Pr\{\omega + Q_{t-1}H_{t-1}^2 > x\}\\ &= E_Q[\Pr\{H^2 > \frac{x-\omega}{Q}|Q\}]\\ &= \int_0^\infty \bar{F}(\frac{x-\omega}{q})f(q)\mathrm{d}q.\end{aligned}$$

By assumption, the term $\bar{F}(\frac{x-\omega}{q})$ has the following expansion:

$$\begin{aligned}\bar{F}(\frac{x-\omega}{q}) &= a\left(\frac{x-\omega}{q}\right)^{-\kappa}\left[1 + b\left(\frac{x-\omega}{q}\right)^{-\beta} + o\left(\left(\frac{x-\omega}{q}\right)^{-\beta}\right)\right]\\ &= aq^{\kappa}x^{-\kappa}\left(1-\frac{\omega}{x}\right)^{-\kappa}\left[1 + bq^{\beta}x^{-\beta}\left(1-\frac{\omega}{x}\right)^{-\beta} + o\left(\left(\frac{x-\omega}{q}\right)^{-\beta}\right)\right].\end{aligned}$$

A Taylor expansion around $\omega/x = 0$ gives

$$\left(1-\frac{\omega}{x}\right)^{-\kappa} = 1 + \kappa\frac{\omega}{x} + \frac{\kappa(1+\kappa)}{2}\left(\frac{\omega}{x}\right)^2 + o\left(x^{-2}\right).$$

Thus we can simplify the term $\bar{F}$:

$$\bar{F}\left(\frac{x-\omega}{q}\right) = aq^{\kappa}x^{-\kappa}\left[1 + bq^{\beta}x^{-\beta} + \kappa\omega x^{-1} + \max\{o(x^{-\beta}), o(x^{-1})\}\right].$$

Finally, we obtain the expression for the integral as

$$\int_0^\infty \bar{F}(\frac{x-\omega}{q})f(q)\mathrm{d}q = ax^{-\kappa}E[Q^{\kappa}] + abx^{-\kappa-\beta}E[Q^{\kappa+\beta}] + a\kappa\omega x^{-\kappa-1}E[Q^{\kappa}]\ + \max\{o(x^{-\kappa-\beta}), o(x^{-\kappa-1})\}]. \tag{A3}$$

The left-hand-side tail expansion in Equation A2 has to agree with the right-hand-side expansion in Equation A3:

$$ax^{-\kappa}\left[1 + bx^{-\beta} + o(x^{-\beta})\right] = ax^{-\kappa}\left[\left(1+\kappa\omega x^{-1}\right)E[Q^{\kappa}] + bx^{-\beta}E[Q^{\kappa+\beta}] + \max\{o(x^{-\beta}), o(x^{-1})\}\right]$$

or

$$1 + bx^{-\beta} + o(x^{-\beta}) = E[Q^{\kappa}] + \kappa\omega x^{-1}E[Q^{\kappa}] + bx^{-\beta}E[Q^{\kappa+\beta}] + \max\{o(x^{-\beta}), o(x^{-1})\}. \tag{A4}$$

For both sides to agree in Equation A4, a first requirement is that we need $E[Q^{\kappa}] = 1$. This is the condition from Kesten (1973) that determines the tail index of the stationary distribution, indicating that the distribution displays heavy tails. Since

$$Q = \tau + \delta N^2,$$

and by assumption $\delta > 0, \tau > 0$, and $\delta + \tau < 1$, a nontrivial solution $\kappa > 0$ exists depending on the distribution function of the innovations N.

If the innovations are normally distributed, so that N^2 is chi square distributed with one degree of freedom, it follows that if $\kappa = 1, E[Q] = \delta + \tau < 1$. By Jensen's inequality, this applies for all $\kappa < 1$. So $\kappa > 1$ is required. It is readily seen from

$$\Gamma(\lambda+1/2)=\sqrt{\pi}(2\delta)^{-\lambda}$$

that for any $\delta\varepsilon\,(0,1)$ there is a solution $\lambda > 1$ such that $E\left[\left(\delta N^2\right)^{\lambda}\right] = 1$ (see, e.g., de Haan et al., 1989). Since for any x, $\tau+\delta x^2$ is increasing in τ, and so is $\left(\tau+\delta x^2\right)^{\lambda}$, it follows that

$$E\left[\left(\tau+\delta N^2\right)^{\lambda}\right] > E\left[\left(\delta N^2\right)^{\lambda}\right] = 1.$$

Note that $E[Q^{\kappa}]$ is differentiable with respect to κ, so that $E[Q^{\kappa}]$ is continuous in κ. By continuity, there exists a κ, $\lambda > \kappa > 1$, such that

$$1 = E[Q^{\kappa}] = E\left[\left(\tau+\delta N^2\right)^{\kappa}\right] < E\left[\left(\tau+\delta N^2\right)^{\lambda}\right].$$

Furthermore, we need to constrain b and β for both sides in Equation A4 to agree. If $\beta > 1$, the second term on the left hand side, $bx^{-\beta}$, is of smaller order than the second-order term on the right had side, $\kappa\omega x^{-1}E[Q^{\kappa}]$. If $\beta < 1$, then $E[Q^{\kappa+\beta}] = 1$, which contradicts with the requirement that $E[Q^{\kappa}] = 1$. Hence we conclude that if the second-order tail index exists β must be equal to 1. This implies that the second-order scale parameter b has the unique form as in Equation 16, since we need to equate bx^{-1}on the left hand side to $\kappa\omega x^{-1}E[Q^{\kappa}] + bx^{-1}E[Q^{\kappa+1}]$ on the right hand side.

Next, we show that the second-order scale parameter b must be negative. Since

$$E\left[Q^{\kappa+1}\right] = E\left[(Q^{\kappa})^{\frac{\kappa+1}{\kappa}}\right] > (E[Q^{\kappa}])^{\frac{\kappa+1}{\kappa}} = 1,$$

by Jensen's inequality. Hence

$$b = \frac{\kappa\omega}{1-E[Q^{\kappa+1}]} < 0.$$

The above arguments were developed as part of lecture notes by De Vries, and partly in private correspondence with Charles Goldie and Laurens de Haan. Later, Baek, Pipiras, Wendt, and Abry (2009) published arguments along similar lines for the forms of β and b in the case of the ARCH(1) process.