

# UC Berkeley

## Recent Work

### Title

On Adaptive Tail Index Estimation for Financial Return Models

### Permalink

<https://escholarship.org/uc/item/2651k8f5>

### Authors

Wagner, Niklas  
Marsh, Terry

### Publication Date

2000-11-01



**Research Program in Finance  
Working Paper RPF-295**

**On Adaptive Tail Index Estimation  
for Financial Return Models**

Niklas Wagner

Terry Marsh

Haas School of Business

November 2000

**ABSTRACT**

Estimation of the tail index of stationary, fat-tailed return distributions is non-trivial since the well-known Hill estimator is optimal only under iid draws from an exact Pareto model. We provide a small sample simulation study of recently suggested adaptive estimators under ARCH-type dependence. The Hill estimator's performance is found to be dominated by a ratio estimator. Dependence increases estimation error which can remain substantial even in larger data sets. As small sample bias is related to the magnitude of the tail index, recent standard applications may have overestimated (underestimated) the risk of assets with low (high) degrees of fat-tailedness.

Keywords: fat-tails, tail index of stationary marginal distributions, Hill estimator, minimal AMSE

JEL#: C13, C14

---

Thanks are to the Department of Mathematical Statistics, Munich University of Technology, for generously providing access to computer facilities. Niklas Wagner acknowledges a Maple program for the calculation of the theoretical tail indices by Klaus Kiefersbeck and a research fellowship by the Deutsche Forschungsgemeinschaft (DFG).

Correspondence: Terry A. Marsh, Walter A. Haas School of Business, U.C. Berkeley, Berkeley CA 94720-1900, Phone: (510) 642-1651, Fax: (510) 643-8460. E-mail: [nwagner@alum.calberkeley.org](mailto:nwagner@alum.calberkeley.org)  
This paper is available on-line at: <http://haas.berkeley.edu/finance/WP/rpflist.html>

## Abstract

Estimation of the tail index of stationary, fat-tailed return distributions is non-trivial since the well-known Hill estimator is optimal only under iid draws from an exact Pareto model. We provide a small sample simulation study of recently suggested adaptive estimators under ARCH-type dependence. The Hill estimator's performance is found to be dominated by a ratio estimator. Dependence increases estimation error which can remain substantial even in larger data sets. As small sample bias is related to the magnitude of the tail index, recent standard applications may have overestimated (underestimated) the risk of assets with low (high) degrees of fat-tailedness.

*Key Words:* fat-tails, tail index of stationary marginal distributions, Hill estimator, minimal AMSE

*JEL Classification:* C13, C14

## 1. Introduction

Early empirical studies on the behavior of speculative asset prices document fat-tails in unconditional return distributions; see for example Blattberg and Gonedes (1974). The effects of non-normality of the distribution of returns regularly gets public attention at the time of events such as the market crash of 1987<sup>1</sup>, the so-called mini-crash in 1989, the Asian crisis in 1997 and the financial crisis in Russia in 1998. Also, various recent cases of individual financial distress add yet more evidence that the “once in a hundred year”-events according to the normal model have in fact been occurring much more frequently. This observation has re-emphasized how important careful modelling of extreme financial returns is to financial risk management applications. Extreme value theory provides a statistical framework characterizing the asymptotic extremal characteristics of stationary distributions. The theory allows us to make inferences about return distributions not only within, but also beyond, the observed range of sample returns and to obtain an adequate characterization of the extremal behavior of returns. To this end, estimation of the so-called tail index is essential, for which theory offers a variety of different approaches.<sup>2</sup> The present paper contributes to the modelling of fat-tailed financial distributions by examining the small sample properties of semi-parametric estimators of the tail index.

The tail index estimation approaches which we examine here are based on the seminal work of Hill (1975). The Hill estimator is established as one of the most suitable for financial applications: the semi-parametric estimation approach is based on the assumption that the underlying distribution is in the maximum domain of attraction of the Fréchet extreme value distribution. This generally holds for fat-tailed distributions as analyzed in finance. Unlike for example the estimation approach based on the generalized extreme value distribution, the assumption does not require that exact asymptotic limits are met. However, the Hill estimator is only optimal under independent draws from an exact Pareto distribution. In financial applications the estimator has to be robust with respect to deviations from the Pareto model and the choice of an appropriate subsample

---

<sup>1</sup>As Rubinstein (1994) and subsequent authors suggest, such events have led market participants to adjust their perceptions of stock return distributions which had a substantial impact on derivatives market prices.

<sup>2</sup>Extreme value theory together with financial applications is outlined in Embrechts et al. (1997). Studies of the tail index of return distributions include Koedijk et al. (1990), Dacorogna et al. (1995), Danielsson and de Vries (1997a), Huisman et al. (1997), Jondeau and Rockinger (1999), Lux (2000), and McNeil and Frey (2000).

of observations from the tail of the return distribution is essential to the estimator's bias/variance trade-off. Furthermore, as returns typically display nonlinear dependence, it is desirable that the estimator is also robust with respect to the independence assumption. Along with deviations from the Pareto model and dependence in returns, the results of recent applications of the Hill estimator in finance leave some room for interpretation. The optimal choice of the number of tail observations is subject to a variety of different adaptive approaches and relatively little is known about the small sample properties of the tail index estimators under ARCH-type dependence. Considering these points, our paper provides a simulation study of the small sample performance of recently suggested adaptive estimation procedures.

Previous pioneering simulation studies on tail index estimation in finance, including Dewachter and Gielens (1994) and Kearns and Pagan (1997), did not address adaptive selection criteria. However, such objective criteria for the choice of the number of tail observations are helpful for judging estimation performance, as for example in Dacarogna et al. (1995) and Huisman et al. (1997). Also, although the generalized autoregressive conditional heteroskedasticity (GARCH) model is prominent for time varying conditional variance in finance, results on the behavior of estimates of the tail index under this type of dependence are sparse (see Danielsson and de Vries (1997b) for an exception). The studies of the performance of bootstrap versions of the Hill estimator given in Dacarogna et al. (1995) are based on independent data. Drees and Kaufmann (1998) and Drees et al. (2000) derive a sequential and a Hill plot-based adaptive estimator and simulate sampling properties under independent random variables.

In this paper, the behavior of three different adaptive versions of the Hill estimator and a moment ratio estimator are studied under three different base models of the return generating process. The chosen return models include stationary stable and student-t distributions as well as the GARCH(1,1) model. Our results indicate that: (i) semi-parametric tail index estimation fails for all stable return models considered, not only for the normal distribution where failure is natural since the latter is not in the maximum domain of attraction of the Fréchet distribution. Hence, the tail index estimation techniques should be used under the *a priori assumption* of the existence of the second moment of the underlying distribution but not in order to make inferences about the existence of moments in the first place; (ii) the bias/variance trade-off is generally important in tail index estimation because unless all data are generated by an exact Pareto model, the Hill estimator is far from optimal in a best unbiased estimator sense. In fact,

our simulation results indicate that, even in a mean squared error sense, the Hill estimator is *not* the best available alternative to semi-parametric tail index estimation for financial applications; (iii) mean squared estimation error can remain substantial even in relatively large data sets when the data are *not* independent, so that the use of high frequency data sampling intervals may come at a cost when high frequency data show stronger deviations from the iid case than lower frequency data; (iv) the superiority of single approaches to adaptive tail index estimation depends on the magnitude of the tail index of the underlying return model and the related estimation bias. No single adaptive approach uniformly dominates in the settings studied, which makes it difficult to give straightforward recommendations, but the bootstrap and the Hill plot-based methods seem to be the most promising candidates; (v) GARCH effects increase estimator variance (which one expects due to the effect of volatility clustering), but surprisingly do not substantially increase mean squared estimation error since estimation bias happens to be reduced under the models studied; (vi) the results show that the magnitude and the sign of the estimation bias is related to the size of the tail index – in particular under GARCH, there is a tendency to *overestimate small tail indices* and to *underestimate large tail indices*. In view of this result, we believe that empirical investigations in the literature indicating a relatively small range of estimated tail indices should be reinterpreted: instead of taking the small range as evidence of quite similar tail characteristics, the finding may simply be due to the statistical tendency to underestimate the difference between the true tail indices.

The organization of the remainder of the paper is as follows. Section 2 gives the methodological background. This includes a review of extreme value asymptotics, the Hill estimator, a selection of adaptive methods for Hill estimation and results on the tail index of ARCH-type return models. The adaptive methods used in our study include sequential, bootstrap and Hill plot-based approaches. The chosen return models and the results of the simulation study are presented in Section 3. The paper ends with concluding remarks in Section 4.

## 2. The Methodological Background

### 2.1. Extreme Value Theory

Classical extreme value theory is concerned with the asymptotic distribution of standardized maxima from a series of independent and identically distributed

(iid) random variables  $(R_t)_{1 \leq t \leq T}$  with a common distribution function  $F$ . For given normalizing constants,  $a_T > 0$ ,  $b_T \in \mathbb{R}$ , and  $M_T = \max(R_1, \dots, R_T)$ , the classical result by Fisher/Tippett and Gnedenko states that if  $H$  exists as the non-degenerate distributional limit of the standardized maximum

$$\Pr\{a_T^{-1}(M_T - b_T) \leq r\} = F^T(a_T r + b_T) \longrightarrow H(r) \quad \text{as } T \longrightarrow \infty, \quad (1.1)$$

then  $H(r)$  is equal to one of three different types of extreme value distributions. The latter are nested within the so-called generalized extreme value distribution

$$H_\xi(r) = \begin{cases} \exp(-(1 + \xi r)^{-1/\xi}), & \text{for } \xi \neq 0 \\ \exp(-\exp(-r)), & \text{for } \xi = 0 \end{cases}, \quad (1.2)$$

where  $1 + \xi r > 0$ . The shape parameter,  $\xi \in \mathbb{R}$ , also denoted as tail index, characterizes the extremal behavior of the distribution function.

Condition (1.1) states that  $F$  belongs to the maximum domain of attraction of  $H_\xi$ ,  $F \in MDA(H_\xi)$ . Fat-tailed distribution functions, which are of interest in financial applications, particularly belong to the maximum domain of attraction of the Fréchet type extreme value distribution,  $F \in MDA(\Phi_\xi)$ , where:  $\Phi_\xi(r) = \exp(-r^{-1/\xi})$ ,  $r > 0$ ,  $\xi > 0$ . From a theorem by Gnedenko, it is well-known that the condition  $F \in MDA(\Phi_\xi)$  is satisfied if and only if the tail  $\bar{F}(r) = 1 - F(r)$  of the distribution function  $F$  is regularly varying at infinity with parameter  $-1/\xi < 0$ , i.e.

$$\bar{F}(r) = L(r)r^{-1/\xi}, \quad r > 0, \quad (1.3)$$

where the function  $L(r)$  is slowly varying at infinity:

$$\lim_{r \rightarrow \infty} \frac{L(tr)}{L(r)} = 1, \quad t > 0.$$

## 2.2. Semi-parametric Tail Index Estimation

The above result demonstrates that the upper<sup>3</sup> tail of a fat-tailed distribution function  $F$  behaves asymptotically like the tail of the Pareto distribution  $G$  given by  $\bar{G}(r) = cr^{-1/\xi}$ ,  $c > 0$ ,  $\xi > 0$ . Based on the order statistics,  $R_{T,T} \leq \dots \leq$

---

<sup>3</sup>The focus on the upper tail is an arbitrary choice. The results of Section 2.1 hold equivalently for the lower tail since  $\min(R_1, \dots, R_T) = -\max(-R_1, \dots, -R_T)$ .

$R_{k,T} \leq \dots \leq R_{1,T}$ , the maximum likelihood estimator (MLE) of  $\xi$  is given by the so-called Hill estimator

$$\widehat{\xi}_{k,T} = \frac{1}{k} \sum_{i=1}^k \ln R_{i,T} - \ln R_{k,T}, \quad (1.4)$$

with  $k = T$ .

### 2.2.1. Asymptotic Tail Behavior: Where does the Tail begin?

As the  $(R_t)_{1 \leq t \leq T}$  have a common distribution function  $F$ , not  $G$ , the usual optimality of the MLE (1.4) does not apply. A subsample fraction  $k = k(T) < T$  has to be selected, i.e. only large observations will be used in the calculation of  $\widehat{\xi}$ . The number  $k(T)$  of order statistics used should increase with the overall sample size  $T$ , while on the other hand, it should be small relative to the overall sample size  $T$ . In the literature this is frequently made precise by the asymptotic condition

$$k(T) \longrightarrow \infty, \quad k(T)/T \longrightarrow 0, \quad \text{as } T \longrightarrow \infty. \quad (1.5)$$

Under model (1.3) and condition (1.5), the Hill estimator can be shown to have following properties: (i) the estimator is consistent (see Embrechts et al. (1997), Example 4.1.12); (ii) under additional assumptions on the asymptotic tail behavior of  $F$ , asymptotic normality follows (see de Haan and Peng (1998), Theorem 1)

$$\sqrt{k}(\widehat{\xi}_{k,T} - \xi) \xrightarrow{d} N(B_\xi; \xi^2), \quad (1.6)$$

where  $B_\xi$  denotes some asymptotic bias term; (iii) the Hill estimator obtains a theoretically derived optimal rate of convergence, being superior to other popular estimators proposed in the literature (Drees (1998)).

Several approaches to the automated determination of an optimal sample fraction  $k(T)$  for the Hill estimator have been studied. Theoretical results on an optimal bias/variance trade-off can be derived using the asymptotic mean squared error (AMSE) as the optimality criterion. The results are based on the so-called Hall model, which forms a generalization of the Pareto model. It imposes a second order condition on the asymptotic behavior of the tail of the distribution function  $F$ . By assuming  $L(r) = c_1 (1 + c_2 r^{-\rho/\xi} + o(r^{-\rho/\xi}))$  in (1.3) it follows that

$$\overline{F}(r) = c_1 r^{-1/\xi} (1 + c_2 r^{-\rho/\xi} + o(r^{-\rho/\xi})), \quad \text{as } r \rightarrow \infty, \quad (1.7)$$



where  $c_1 > 0$ ,  $c_2 \in \mathbb{R}$  and  $\rho > 0$ . The above model gives an asymptotic characterization of the tail of the underlying distribution and at the same time robustifies the semi-parametric estimation approach against deviations from the exact Pareto tail. The model quite generally holds for the well-known distribution functions such as the Fréchet, the student-t and the symmetric  $\alpha$  stable.

We consider three Hill-based and one moment ratio approach to adaptive tail index estimation. The first three proposals rely on model (1.7), while the fourth is a formalized heuristical criterion supported by theoretical results. The estimation approaches are briefly presented in the following.

- Drees and Kaufmann (1998) derive a sequential estimator of the optimal  $k$  by extending previous theoretical work on the asymptotic bias and variance of the Hill estimator. A stopping time criterion for a sequence of Hill estimators is used in order to approximate the number of upper order statistics  $k$  under which the bias in the Hill estimator starts to dominate the maximum random fluctuation of the series  $\sqrt{i} \left| \widehat{\xi}_{i,T} - \xi \right|$ ,  $2 \leq i \leq k$ , given some threshold  $u_T > 0$ :

$$\bar{k}(u_T) = \min \left\{ k \in \{2, \dots, T\} : \max_{2 \leq i \leq k} \left( \sqrt{i} \left| \widehat{\xi}_{i,T} - \widehat{\xi}_{k,T} \right| \right) > u_T \right\}.$$

The second order parameter  $\rho$  in (1.7) is either set equal to a constant or given by a consistent estimator such as

$$\widehat{\rho}_T(u_T, \lambda) = \frac{1}{\ln \lambda} \ln \frac{\max_{2 \leq i \leq [\lambda \bar{k}(u_T)]} \left( \sqrt{i} \left| \widehat{\xi}_{i,T} - \widehat{\xi}_{[\lambda \bar{k}(u_T)], T} \right| \right)}{\max_{2 \leq i \leq \bar{k}(u_T)} \left( \sqrt{i} \left| \widehat{\xi}_{i,T} - \widehat{\xi}_{\bar{k}(u_T), T} \right| \right)} - \frac{1}{2}, \quad (1.8)$$

where  $\lambda \in (0, 1)$  and  $[x]$  denotes the largest integer smaller or equal to  $x$ . Then, for a given  $\rho$  and a consistent initial estimator of  $\xi$ , a consistent estimator of the optimal  $k$  is derived. The procedure yields a Hill estimator with minimal asymptotic MSE (Drees and Kaufmann (1998), Theorem 1).

- Dacarogna et al. (1995) as well as Danielsson et al. (1997) and Danielsson and de Vries (1997a, b) use a subsample bootstrap approach by Hall (1990) to estimate the optimal sample fraction. The approach is based on bootstrap subsamples of size  $T_1 < T$ . The estimate of the optimal subsample

fraction can be derived from

$$\widehat{k}_1 = \arg \min_{1 \leq k \leq T_1} \widehat{E} \left( (\widehat{\xi}_{k,T_1}^* - \widetilde{\xi}_T)^2 \mid (R_1, \dots, R_T) \right), \quad (1.9)$$

where  $*$  denotes estimates based on resamples drawn from  $(R_1, \dots, R_T)$ .  $\widetilde{\xi}_T$  denotes a consistent initial Hill estimator. The expectation operator is approximated by a given number of bootstrap runs. Instead of applying the bootstrap to the MSE of the Hill estimator directly as in (1.9), an asymptotically equivalent criterion which does not depend on the choice of an initial estimator is proposed in Danielsson et al. (1997). Their bootstrap estimate is given by<sup>4</sup>

$$\widehat{k}_1 = \arg \min_{1 \leq k < T_1} \widehat{E} \left( (M_{k,T_1}^* - 2(\widehat{\xi}_{k,T_1}^*)^2)^2 \mid (R_1, \dots, R_T) \right), \quad (1.10)$$

where  $M_{k,T}$ ,  $k < T$ , is a second order moment estimator of the form:

$$M_{k,T} = \frac{1}{k} \sum_{i=1}^k (\ln R_{i,T} - \ln R_{k+1,T})^2.$$

The estimate for the optimal overall sample fraction  $k_{opt}$  can then (most simply as in Hall (1990) and Danielsson and de Vries (1997a)) be derived from:

$$\widehat{k}_{opt} = \left[ \widehat{k}_1 (T/T_1)^{2\rho/(2\rho+1)} \right]. \quad (1.11)$$

Again, the second order parameter  $\rho$  is either estimated by equation (1.8) or set equal to a constant.

- Various generalizations of the Hill estimator have been proposed in the literature. The moment ratio estimator

$$\widehat{\xi}_{k,T}^m = \frac{1}{2} (M_{k,T} / \widehat{\xi}_{k,T}), \quad (1.12)$$

---

<sup>4</sup>In case the maximum sample realization is drawn at least twice in a bootstrap resample, the MSE equivalent criterion (1.10) yields a zero value for  $k = 1$  (and possibly for  $k > 1$ ), whereas the MSE criterion (1.9) yields the squared initial estimator. We enter the squared initial estimator value in this case.

may be a useful approach for financial applications. Danielsson et al. (1996) point out that  $\widehat{\xi}_{k,T}^m$  has lower asymptotic (absolute) bias than the Hill estimator  $\widehat{\xi}_{k,T}$ . Evaluated at their respective AMSE-minimizing  $k$ 's, the moment ratio estimator can not only have lower bias but also lower AMSE than the Hill estimator, depending on the tail index and the second order parameter of the underlying distribution function (de Haan and Peng (1998)). We propose the subsample bootstrap approach and estimate the optimal sample fraction  $k_{opt}^m$  for the moment ratio estimator by equation (1.9).

- Guided by the frequently recommended approach of choosing an estimate of the tail index within a “stable” region of the Hill plot  $\{(k, \widehat{\xi}_{k,T}) : 1 \leq k \leq T\}$ , Drees et al. (2000) propose a maximal occupation time Hill estimator. It can be shown that the so-called alternative Hill plot,  $\{(\theta, \widehat{\xi}_{[T^\theta],T}) : 0 \leq \theta \leq 1\}$ , yields superior results if the underlying distribution is not exactly Paretian. The proposed consistent maximal occupation time estimator is given by

$$\widehat{\xi}_T = \arg \max_{\xi > 0} \int_0^{\bar{\theta}} \mathbf{1}_{\{[T^\theta]^{1/2} |\widehat{\xi}_{[T^\theta],T} - \xi| \leq m \tilde{\xi}_T\}} d\theta, \quad (1.13)$$

with an upper boundary  $0 < \bar{\theta} < 1$  and a scaling constant  $m > 0$ . The parameter  $\tilde{\xi}_T$  denotes a consistent initial Hill estimator which is used as an estimate of the standard deviation of the Hill plot.

### 2.2.2. The Tail Index for ARCH-type Return Series

We now consider the case where the  $(R_t)_{1 \leq t \leq T}$  are not necessarily independent but have a stationary marginal distribution function  $F$ . GARCH is a well-known model class for non-linear dependence in financial time series. Assuming a constant return expectation and excess returns with a conditional time-varying variance  $\sigma_t^2$ , the prominent GARCH(1,1) process has the representation:

$$R_t = \sigma_t Z_t, \quad \sigma_t^2 = \beta_0 + \beta_1 R_{t-1}^2 + \beta_2 \sigma_{t-1}^2, \quad \beta_0, \beta_1, \beta_2 \geq 0. \quad (1.14)$$

The random variables  $Z_t$  are standardized iid draws from some symmetric, possibly fat-tailed, distribution function with density  $g(z) : \mathbb{R} \mapsto \mathbb{R}^+$ . As outlined for example in Mikosch and Stărică (1998), a stationary marginal distribution  $F$  for the GARCH(1,1) process (1.14) exists if:

$$\int_{-\infty}^{\infty} \ln |\beta_1 z^2 + \beta_2| g(z) dz < 0, \quad \beta_0 > 0. \quad (1.15)$$

It can then be shown that the tail index of  $F$  is given as a solution to an integral equation under which it follows that:  $\overline{F}(r) \sim cr^{-1/\xi}$  as  $r \rightarrow \infty$ . For the ARCH(1) process, analytical solutions are obtainable. For the GARCH(1,1) process, the tail index can be determined numerically as a solution to

$$I(\xi) \equiv \int_{-\infty}^{\infty} |\beta_1 z^2 + \beta_2|^{\frac{1}{2\xi}} g(z) dz - 1 = 0, \quad \xi > 0. \quad (1.16)$$

Under condition (1.5), consistency of the Hill estimator can be maintained theoretically under ARCH-type models; see Resnick and Stărică (1998). Also, the results by Hsing (1991) indicate that the Hill estimator is asymptotically quite robust with respect to deviations from independence. However, there remain open questions: the sampling behavior of the Hill estimator in small samples under ARCH-type dependence is unknown. Hence, the asymptotic normality result (1.6) only gives a crude approximation to its bias and standard error. In particular, ARCH effects cause clustering in the extremes of the process (1.14). This potentially causes the estimation error for a given sample size to be larger than it would be under independence. Also, little is known about how ARCH-type dependence influences the bias component of the estimator and the bias/variance trade-off under the adaptive subsample selection criteria.

### 3. Simulation Study

In order to characterize the small sample properties of the various adaptive Hill estimators presented above, we study the estimators' error distributions and particularly their root mean squared error in a series of Monte Carlo simulations.

Without loss of generality, the sample size  $T$  will be defined as the number of positive sample observations. We exploit the property that all our return models are based on distribution functions which are symmetric around zero by simulating  $T$  observations and taking absolute values. The corresponding overall sample size is then on average  $N = 2T$ . This refers to the typical application, where, observing a sample of  $N$  returns from a distribution which is not necessarily symmetric, one is interested in separate inferences about each of both tails of the underlying return distribution.

#### 3.1. Estimator Settings and Definitions

The adaptive estimators that were outlined in Section 2.2 require the choice of several variables, where we rely on the suggestions made in the literature as far

as possible. Following Drees and Kaufmann (1998), we choose a threshold value  $u_T = \frac{5}{2} \widehat{\xi}_{[2\sqrt{T}], T} T^{1/4}$  for the sequential Hill estimator. The second order parameter  $\rho$  in the asymptotic tail expansion (1.7) is either set equal to 1, giving the Hill estimator “H-DKC”, or estimated by (1.8) with  $\lambda = 0.6$ , giving the estimator “H-DKE”. Setting  $\rho$  equal to a constant is advisable for moderate sample sizes as, on the one hand, the estimators of the second order parameter tend to be rather unstable (see Danielsson and de Vries (1997a) and Drees and Kaufmann (1998)) and, on the other hand, do not seem to have a large influence on the results (see Dacarogna et al. (1995)). For the bootstrap approach we choose the subsample sizes  $T_1 = [T/10]$  as in Danielsson and de Vries (1997a) and in Jondeaux and Rockinger (1999). The asymptotically MSE-equivalent criterion proposed by Danielsson et al. (1997) is chosen, which yields an optimal subsample fraction according to (1.10). Handling the second order parameter  $\rho$  as in the Drees and Kaufmann approach, (1.11) gives the estimated optimal sample fraction and the estimators “H-BSC” for constant  $\rho = 1$  and “H-BSE” for  $\widehat{\rho}$  according to (1.8). For the Drees et al. (2000) maximum occupation time estimator, the continuous parameter  $\theta$  is approximated by a grid with stepsize 0.05, where the upper bound  $\bar{\theta} = \ln([T/2]) / \ln(T)$  corresponds to  $k \leq [T/2]$ . Together with the scaling constant  $m = 1$ , this defines the “H-MOT” estimate. The initial estimator,  $\widetilde{\xi}_T = \widehat{\xi}_{[2\sqrt{T}], T}$ , is given as a naive, asymptotically suboptimal, alternative estimator labelled “H-INI”.

We also consider simulation results for the moment ratio estimator (1.12). The initial moment ratio estimator “M-INI” is chosen as  $\widetilde{\xi}_T^m = \widehat{\xi}_{[4\sqrt{T}], T}^m$ . “M-BSHC” and “M-BSHE” represent the moment ratio estimators based on the optimal  $k$  from the bootstrap results for the Hill estimator (“H-BSC” and “H-BSE”). These two estimates allow a comparison of the Hill and the moment ratio estimators given an identical selection of the number of upper order statistics. The MSE bootstrap (1.9) with  $\widetilde{\xi}_T^m$  as consistent initial estimator provides “M-BSMC” for constant  $\rho = 1$  and “M-BSME” for  $\widehat{\rho}$  according to (1.8). The maximum occupation time estimator is calculated as above with  $\widehat{\xi}$  and  $\widetilde{\xi}$  replaced by  $\widehat{\xi}^m$  and  $\widetilde{\xi}^m$ , respectively. Altogether, twelve different estimators are calculated; six Hill estimators: H-INI, H-DKC, H-DKE, H-BSC, H-BSE, H-MOT and six moment ratio estimators: M-INI, M-BSHC, M-BSHE, M-BSMC, M-BSME, M-MOT.

**Table 1**

Return model parameters and the corresponding tail index of the stationary marginal distributions.

Parameter Model:	Label:	$\xi$	$\alpha_{sas}$	$\nu$	$\beta_0$	$\beta_1$	$\beta_2$
I: iid sas	sas/0.67	0.67	1.50	—	—	—	—
	sas/0.53	0.53	1.90	—	—	—	—
	sas/0	0	2	$\infty$	—	—	—
II: iid student-t( $\nu$ )	stud/0.17	0.17	—	6	—	—	—
	stud/0.25	0.25	—	4	—	—	—
	stud/0.33	0.33	—	3	—	—	—
III: GARCH(1,1)-t( $\nu$ )	arch/0.17	0.17	—	9	$10^{-6}$	0.05	0.92
	arch/0.25	0.25	—	5	$10^{-6}$	0.03	0.94
	arch/0.33	0.33	—	4	$10^{-6}$	0.03	0.93

### 3.2. Return Models

Our simulations are based on  $N = 1000, 2000$  and  $3000$  observations from three different return models with zero return expectation.<sup>5</sup> Table 1 contains an overview of the return models. The models include: (I) the iid symmetric  $\alpha$ -stable return model by Mandelbrot (see also Janicki and Weron (1993)) labelled “sas/ $\xi$ ”, as well as (II) the iid symmetric student-t( $\nu$ ) return model by Clark (“stud/ $\xi$ ”). The latter two fat-tailed classes include Bachelier’s thin-tailed normal model as a special case for  $\alpha = 2$  or  $\nu \rightarrow \infty$ , respectively. As a third model class (III), GARCH(1,1) return generating processes (1.14) are simulated under innovations from a student-t( $\nu$ ) distribution (“arch/ $\xi$ ”).

The tail indices  $\xi$  for the marginal GARCH-t( $\nu$ ) distributions are calculated by solving equation (1.15) with  $g(z)$  denoting a student-t density with  $\nu$  degrees of freedom. Note that the model parameters are chosen such that the tail indices are approximately equal to those of the iid student models. When studying the properties of the tail index estimators, this allows us to distinguish between the effects of the magnitude of the tail index on the one hand and ARCH-type dependence on the other hand.

<sup>5</sup>Although the Hill estimator is not location invariant (see Drees (1998)), adding a positive constant to each return realization in earlier unreported simulations did not have a notable effect on the estimation results.

The focus of our simulation study is on the performance of the tail estimators under the GARCH-t model class. Based on the early evidence by Blattberg and Gonedes (1974) and more recent findings, finite variance models seem to dominate  $\alpha$ -stable models for most studied return series. The GARCH models also capture volatility clustering, a phenomenon which is widely documented in the literature. Referring to empirical evidence from GARCH estimation, the parameters  $\beta_1$  and  $\beta_2$  are chosen such that  $\beta_1 + \beta_2 \lesssim 1$ , i.e. the models are what is called “nearly integrated”.

### 3.3. Simulation Results

All our simulation results are based on 500 independent simulation runs, under each of which 100 bootstrap runs are performed for calculating the bootstrap estimators. The statistics reported are mean error (ME), standard deviation (STD) and root mean squared error (RMSE). Results for the Hill estimators are given in Table 2 for sample size  $N = 1000$  and in Table 4 for sample size  $N = 3000$ . Tables 3 and 5 give the results for the moment ratio estimators and corresponding sample sizes.<sup>6</sup> Our conclusions from the simulations are as follows.

#### 3.3.1. Stable versus Non-Stable Models

Estimates of the tail index of the symmetric  $\alpha$ -stable model (sas) turn out to be unreliable, particularly when  $\alpha$  is close to 2 which is the case of interest in financial applications. There is a large negative estimation bias for  $\alpha \lesssim 2$  and a large positive estimation bias for  $\alpha = 2$  as can be seen from the ME statistics in Tables 2 through 5. For example, when the true sas model tail index is  $\xi = 0.53$ , negative estimation bias relates to an expected value of the tail index estimator around 0.3, the bias being almost half the magnitude of the true tail index. These findings do not significantly change for the sample sizes and estimation procedures considered. In contrast, Tables 2 through 5 also demonstrate that the estimation bias is much lower when returns are generated by the student-t or the GARCH models.

Thus, although theoretically justified for symmetric stable laws with  $\alpha < 2$ , applications of Hill-type estimators do not seem promising for stable laws in small samples. An explanation for this finding is that, under the limiting normal model,

---

<sup>6</sup>The simulation results for  $N = 2000$  conform to the findings from the two other sample sizes and hence were not tabulated separately.

the tail  $\overline{F}$  is not regular varying and the Hill estimator is therefore inappropriate. The finding of substantial bias under the *sas* model with  $\alpha < 2$  implies that, without a priori knowledge about the underlying distribution, it will be difficult to distinguish between stable and non-stable underlying distributions solely based on estimates of the tail index. Furthermore, statements about the existence of certain moments of the underlying distribution should only be made conditional on the a priori assumption that the underlying distribution is not stable under addition. In the following we limit our discussion to the non-stable student-t and GARCH models.

### 3.3.2. Magnitude of the Tail Index and Estimation Precision

Turning to the Hill and the moment ratio estimators for student-t and GARCH models, we observe in Tables 2 through 5 that higher degrees of fat-tailedness (i.e. larger values of  $\xi$ ) frequently show quite stable or even decreased RMSE statistics for the estimators. For example, in Table 4, the RMSE for the bootstrap estimator H-BSC is 0.080 for the student-t model with  $\xi = 0.17$  and 0.071 for the student-t model with  $\xi = 0.33$ . The RMSE for the initial Hill estimator (H-INI) is 0.091 for the student-t model with  $\xi = 0.17$  and declines by roughly a third to 0.063 when the tail index is nearly doubled to  $\xi = 0.33$ .

The decrease in RMSE with an increase in  $\xi$  deserves further consideration recalling that  $\text{MSE} = \text{STD}^2 + \text{ME}^2$  and that the asymptotic bounds,  $\xi/\sqrt{k(T)}$ , to the standard deviation of the Hill estimator from equation (1.6) are linearly increasing in  $\xi$ . With, for example  $k(T) = \lceil 2\sqrt{T} \rceil$  and sample size  $N = 3000$ , the asymptotic standard deviations are 0.019 for  $\xi = 0.17$ , 0.028 for  $\xi = 0.25$  and 0.038 for  $\xi = 0.33$ . In Table 4, the corresponding simulated standard deviations for the initial estimator H-INI are 0.028, 0.033 and 0.041 respectively, i.e. especially under the smaller tail indices, the standard deviations have not reached their asymptotic bounds, even with a sample size of 3000. Nevertheless, the standard deviations do increase with the tail index  $\xi$  and estimation variance is high especially for the models with  $\xi = 0.33$ .

The observation that the RMSE remains stable or even decreases for increasing tail index  $\xi$  is partly explainable by a slower convergence to the asymptotic standard deviations for models with small tail index. This reduces the increase in sampling variance for estimation under models with larger tail indices. The main explanation of our findings, however, is the behavior of the sample bias (ME): decreases in absolute bias tend to more than compensate for increasing estima-



**Table 2**

Hill estimators; mean error (ME), standard deviation (STD), and root mean squared error (RMSE),  $N = 1000$

	H-INI	H-DKC	H-DKE	H-BSC	H-BSE	H-MOT
	ME STD RMSE	ME STD RMSE	ME STD RMSE	ME STD RMSE	ME STD RMSE	ME STD RMSE
Model:						
$s\alpha s/0.67$	-0.078 0.093 0.12	-0.096 0.074 0.12	-0.090 0.069 0.11	-0.090 0.079 0.12	-0.098 0.069 0.12	-0.25 0.066 0.26
$s\alpha s/0.53$	-0.25 0.047 0.25	-0.26 0.062 0.26	-0.25 0.061 0.26	-0.26 0.067 0.27	-0.26 0.063 0.27	-0.25 0.068 0.26
$s\alpha s/0$	0.20 0.027 0.21	0.16 0.036 0.16	0.17 0.048 0.18	0.15 0.042 0.16	0.16 0.045 0.16	0.16 0.043 0.16
$stud/0.17$	0.12 0.039 0.12	0.088 0.052 0.10	0.10 0.061 0.12	0.077 0.060 0.098	0.081 0.062 0.10	0.074 0.059 0.095
$stud/0.25$	0.093 0.050 0.11	0.076 0.063 0.099	0.092 0.068 0.11	0.067 0.073 0.099	0.072 0.074 0.10	0.066 0.068 0.094
$stud/0.33$	0.072 0.057 0.092	0.061 0.069 0.092	0.079 0.072 0.11	0.053 0.081 0.097	0.057 0.083 0.10	0.038 0.069 0.079
$arch/0.17$	0.14 0.055 0.15	0.11 0.072 0.13	0.12 0.084 0.15	0.092 0.079 0.12	0.096 0.083 0.13	0.081 0.070 0.11
$arch/0.25$	0.086 0.055 0.10	0.057 0.070 0.090	0.074 0.079 0.11	0.040 0.076 0.086	0.046 0.081 0.093	0.036 0.069 0.078
$arch/0.33$	0.049 0.065 0.081	0.025 0.084 0.088	0.038 0.098 0.11	0.011 0.090 0.090	0.015 0.093 0.094	-0.0097 0.072 0.072

**Table 3**

Moment ratio estimators; mean error (ME), standard deviation (STD), and root mean squared error (RMSE),  $N = 1000$ .

	M-INI	M-BSHC	M-BSHE	M-BSMC	M-BSME	M-MOT
	ME STD RMSE	ME STD RMSE	ME STD RMSE	ME STD RMSE	ME STD RMSE	ME STD RMSE
Model:						
$s\alpha s/0.67$	-0.074 0.099 0.12	-0.064 0.12 0.14	-0.072 0.11 0.13	-0.089 0.084 0.12	-0.082 0.079 0.11	-0.26 0.062 0.26
$s\alpha s/0.53$	-0.23 0.057 0.24	-0.22 0.12 0.25	-0.23 0.11 0.25	-0.25 0.066 0.25	-0.24 0.065 0.25	-0.24 0.084 0.25
$s\alpha s/0$	0.21 0.018 0.21	0.12 0.034 0.13	0.13 0.037 0.13	0.17 0.024 0.17	0.17 0.029 0.17	0.14 0.033 0.14
$stud/0.17$	0.12 0.028 0.12	0.046 0.058 0.074	0.050 0.059 0.077	0.083 0.039 0.092	0.089 0.044 0.099	0.050 0.053 0.073
$stud/0.25$	0.094 0.037 0.10	0.038 0.075 0.084	0.044 0.075 0.087	0.066 0.051 0.083	0.073 0.054 0.091	0.040 0.073 0.083
$stud/0.33$	0.069 0.046 0.083	0.021 0.083 0.086	0.027 0.082 0.086	0.046 0.060 0.076	0.056 0.063 0.084	0.018 0.078 0.080
$arch/0.17$	0.14 0.046 0.15	0.050 0.066 0.083	0.055 0.071 0.090	0.096 0.052 0.11	0.10 0.059 0.12	0.054 0.059 0.080
$arch/0.25$	0.087 0.044 0.097	-0.0019 0.067 0.067	0.0034 0.071 0.071	0.044 0.052 0.068	0.051 0.057 0.076	0.0055 0.066 0.066
$arch/0.33$	0.048 0.053 0.071	-0.039 0.078 0.087	-0.032 0.083 0.089	0.0063 0.067 0.062	0.014 0.067 0.069	-0.038 0.078 0.086

**Table 4**

Hill estimators; Mean error (ME), standard deviation (STD), and root mean squared error (RMSE),  $N = 3000$

	H-INI	H-DKC	H-DKE	H-BSC	H-BSE	H-MOT
	ME STD RMSE	ME STD RMSE	ME STD RMSE	ME STD RMSE	ME STD RMSE	ME STD RMSE
Model:						
$s\alpha s/0.67$	-0.061 0.068 0.092	-0.092 0.045 0.10	-0.097 0.038 0.10	-0.091 0.043 0.10	-0.098 0.040 0.11	-0.26 0.065 0.27
$s\alpha s/0.53$	-0.26 0.040 0.26	-0.25 0.043 0.26	-0.26 0.038 0.26	-0.26 0.038 0.26	-0.26 0.038 0.26	-0.23 0.060 0.23
$s\alpha s/0$	0.17 0.017 0.17	0.14 0.023 0.14	0.13 0.031 0.14	0.12 0.033 0.13	0.12 0.035 0.13	0.13 0.068 0.15
$stud/0.17$	0.087 0.028 0.091	0.073 0.035 0.081	0.080 0.041 0.090	0.065 0.046 0.080	0.067 0.047 0.082	0.053 0.046 0.070
$stud/0.25$	0.064 0.033 0.072	0.057 0.041 0.070	0.067 0.045 0.081	0.050 0.055 0.075	0.052 0.056 0.077	0.039 0.058 0.070
$stud/0.33$	0.048 0.041 0.063	0.048 0.044 0.065	0.060 0.047 0.076	0.046 0.055 0.071	0.048 0.056 0.074	0.030 0.064 0.070
$arch/0.17$	0.10 0.043 0.11	0.085 0.055 0.10	0.090 0.063 0.11	0.070 0.066 0.096	0.069 0.069 0.098	0.049 0.058 0.075
$arch/0.25$	0.069 0.052 0.086	0.054 0.062 0.082	0.060 0.069 0.092	0.039 0.075 0.084	0.041 0.078 0.088	0.016 0.066 0.067
$arch/0.33$	0.030 0.062 0.069	0.020 0.081 0.084	0.027 0.087 0.092	-0.0009 0.097 0.096	0.0001 0.098 0.098	-0.030 0.071 0.077

**Table 5**

Moment ratio estimators; Mean error (ME), standard deviation (STD), and root mean squared error (RMSE),  $N = 3000$ .

	M-INI	M-BSHC	M-BSHE	M-BSMC	M-BSME	M-MOT
	ME STD RMSE	ME STD RMSE	ME STD RMSE	ME STD RMSE	ME STD RMSE	ME STD RMSE
Model:						
$s\alpha s/0.67$	-0.059 0.075 0.095	-0.063 0.070 0.095	-0.073 0.059 0.094	-0.085 0.049 0.098	-0.078 0.048 0.092	-0.27 0.065 0.27
$s\alpha s/0.53$	-0.23 0.056 0.23	-0.20 0.081 0.21	-0.21 0.074 0.22	-0.24 0.045 0.24	-0.24 0.045 0.24	-0.21 0.077 0.22
$s\alpha s/0$	0.17 0.011 0.17	0.10 0.028 0.11	0.10 0.029 0.11	0.14 0.015 0.14	0.14 0.019 0.14	0.14 0.092 0.17
$stud/0.17$	0.088 0.021 0.090	0.044 0.047 0.064	0.045 0.048 0.066	0.067 0.029 0.073	0.070 0.031 0.076	0.035 0.047 0.059
$stud/0.25$	0.063 0.028 0.069	0.031 0.059 0.067	0.033 0.059 0.067	0.049 0.038 0.062	0.053 0.040 0.066	0.014 0.067 0.069
$stud/0.33$	0.047 0.036 0.059	0.027 0.060 0.066	0.030 0.059 0.066	0.038 0.044 0.058	0.044 0.044 0.063	0.0081 0.073 0.074
$arch/0.17$	0.10 0.037 0.11	0.041 0.059 0.072	0.042 0.062 0.075	0.075 0.042 0.086	0.076 0.046 0.089	0.029 0.049 0.057
$arch/0.25$	0.065 0.046 0.079	0.0048 0.068 0.068	0.0071 0.071 0.071	0.040 0.052 0.065	0.042 0.054 0.069	-0.011 0.063 0.064
$arch/0.33$	0.024 0.058 0.063	-0.039 0.087 0.095	-0.036 0.089 0.096	-0.0021 0.059 0.059	0.0021 0.062 0.062	-0.057 0.075 0.095

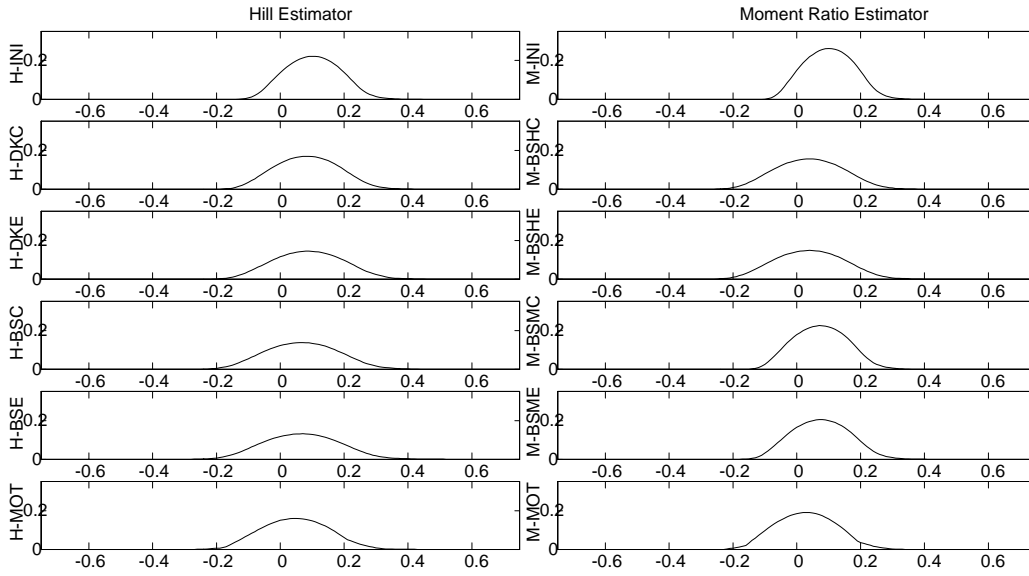
tion variance. For example, in the case mentioned above, when RMSE falls from 0.091 to 0.063, we see that the bias falls from 0.087 to 0.048, more than offsetting the increase in standard deviation from 0.028 to 0.041. Interestingly, for the maximum occupation time estimator (MOT), this compensation effect weakens and absolute bias even increases for the GARCH models with  $\xi = 0.33$  in Tables 3 through 5. In this case, a larger tail index yields reduced estimation precision measured by the STD *and* the RMSE statistic.

Overall, our results indicate that the bias/variance trade-off is complex and hence one should be careful with simple rules relating overall estimation precision to the magnitude of the tail index. Figures 1a and 1b give a graphical illustration of the findings for two of the GARCH models. Figure 1a shows the error densities for the Hill and the moment ratio estimators under the GARCH model with  $\xi = 0.17$ , while Figure 1b shows the respective error densities under the GARCH model with  $\xi = 0.33$ .

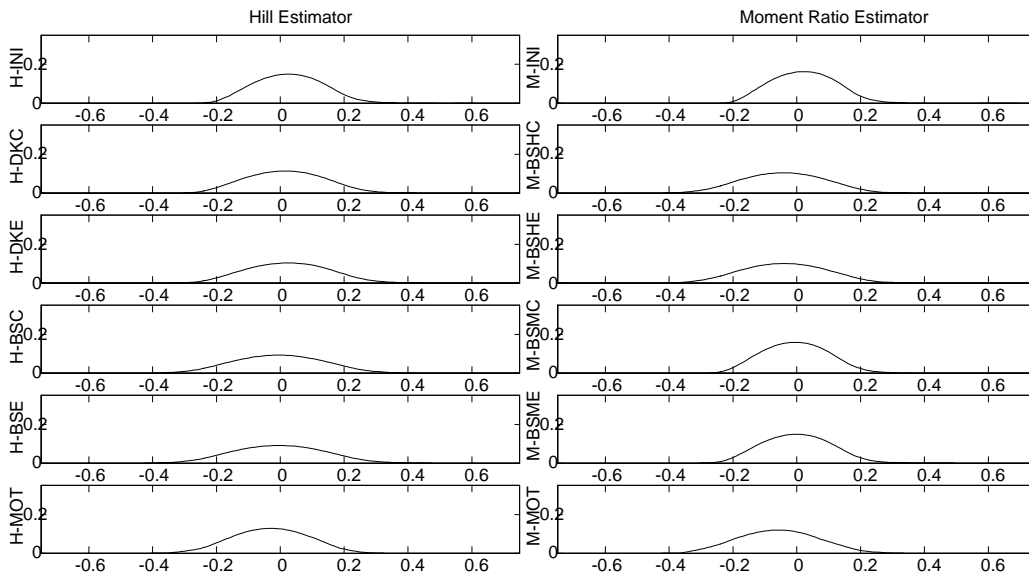
### 3.3.3. Hill versus Moment Ratio Estimators

Comparing the performance of the Hill versus the moment ratio estimators, the simulation results show that the moment ratio estimators' mean squared error statistics are in general substantially smaller than those of the Hill estimators. The results for the bootstrap estimators demonstrate that the moment ratio estimators give superior RMSE performance for an identical subsample size  $k$ . This can be seen by comparing the Hill bootstrap estimators H-BSC and H-BSE in Tables 2 and 4 with the moment ratio estimators M-BSHC and M-BSHE in Tables 3 and 5. Under the estimates of the optimal moment ratio subsample fraction  $k_{opt}^m$  and a sample size of  $N = 3000$ , the performance of the moment ratio estimator further improves for the models with  $\xi = 0.25$  and  $\xi = 0.33$ , but not for  $\xi = 0.17$ ; see the standard errors for the M-BSM estimator as compared to the M-BSH estimator in Table 5. But even for  $\xi = 0.17$  the moment ratio estimators' RMSE statistics still dominate those of the Hill estimators.

The results for the maximum occupation time estimator (MOT) are not as straightforward as those for the bootstrap estimators. It turns out that for models with  $\xi = 0.33$  the moment ratio estimator M-MOT has slightly larger RMSE than the Hill estimator H-MOT while the RMSE is lower for models with  $\xi = 0.17$  and  $\xi = 0.25$ . This suggests that the moment ratio estimators are particularly appropriate for larger tail indices and subsample selection based on the bootstrap estimator.



**Figure 1a:** Epanechnikov kernel densities of the Hill and the moment ratio estimation errors under the GARCH-t model with  $\xi = 0.17$ ,  $N = 3000$ .



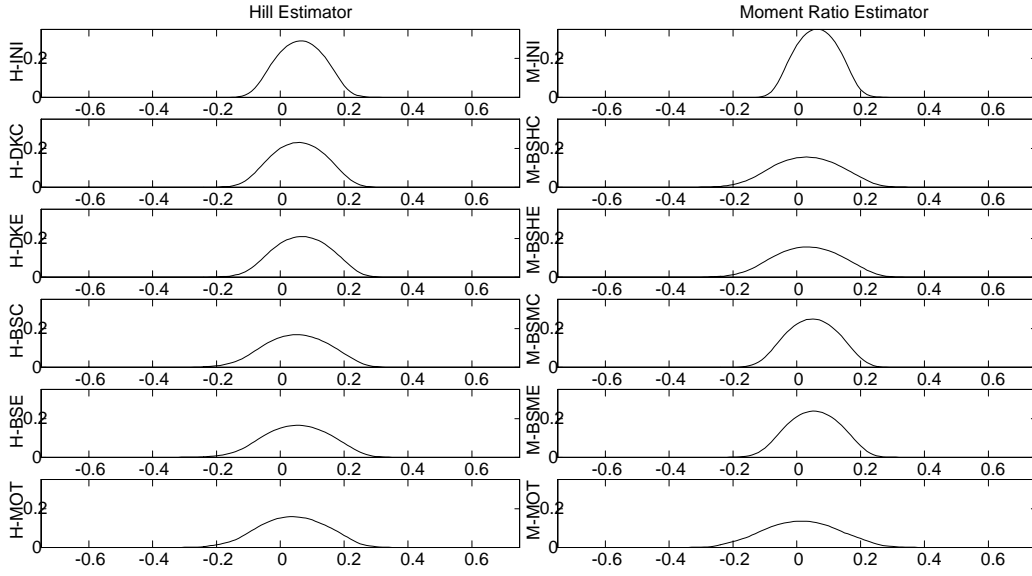
**Figure 1b:** Epanechnikov kernel densities of the Hill and the moment ratio estimation errors under the GARCH-t model with  $\xi = 0.33$ ,  $N = 3000$ .

Overall, we conclude that the best RMSE statistics under the student-t and GARCH models for all sample sizes are always obtained with a moment ratio estimator. Also, for both the Hill and the moment ratio estimators, setting the second order parameter  $\rho$  from equation (1.7) equal to one yields better results than estimating the coefficient (even though it is known that  $\rho = 2/\nu$  for the student-t( $\nu$ ) distribution): for all simulations, the RMSE under the fixed parameter is smaller or equal to the RMSE under the estimated parameter. A graphical illustration of the error densities of the Hill and the moment ratio estimators using the two GARCH models with  $\xi = 0.17$  and  $\xi = 0.33$  is given in Figures 1a and 1b.

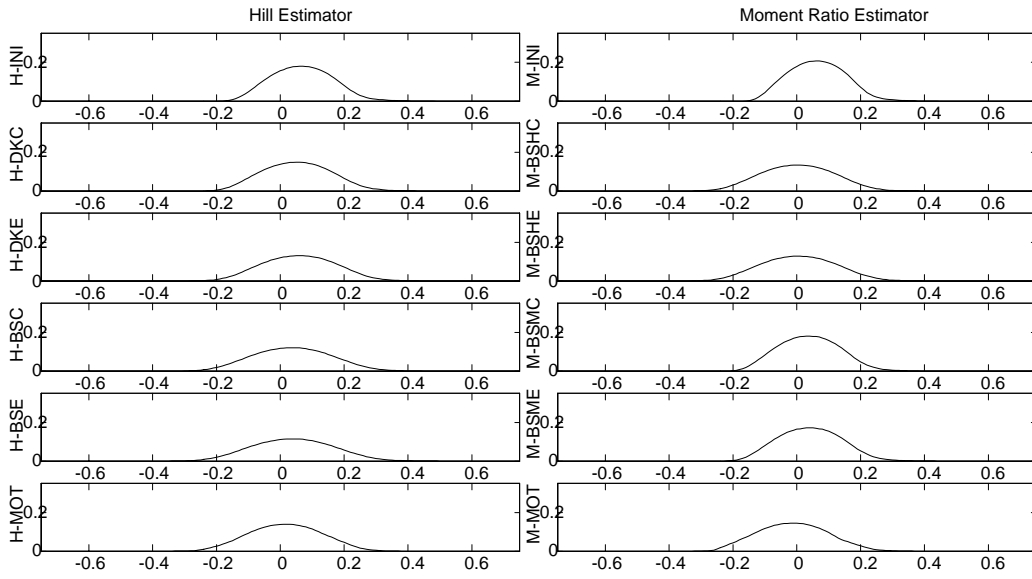
### 3.3.4. Performance of the Adaptive Subsample Selection Approaches

A comparison of the results of the adaptive estimation approaches with those of the initial estimator shows that the optimal choice of  $k$  in the former involves a trade-off between reduced estimation bias and increased estimation variance. This is illustrated for example in Figure 2a, where a shift of the Hill estimators' error density to the left, i.e. a reduction in estimation bias, is accompanied by an increase in estimation variance, i.e. a widening of the density.

In comparing the adaptive estimation approaches, the RMSE-optimal approach depends on the magnitude of the tail index  $\xi$ . For example, the MOT estimator dominates the BS estimator for  $\xi = 0.17$  while the BS estimator dominates the MOT for  $\xi = 0.33$ . This is mostly due to a different bias structure for the two estimation approaches. While the MOT estimator has lowest absolute bias for the GARCH model with  $\xi = 0.25$ , the bootstrap estimator reaches this point around  $\xi = 0.33$ . Assuming one has a priori knowledge about the general magnitude of the tail index, it seems advisable to use the M-MOT estimator for smaller tail indices and the M-BSMC estimator for larger tail indices. Overall, the moment ratio estimators together with the bootstrap (M-BSMC) and the maximum occupation time estimation approach (M-MOT) tend to perform best (see Tables 3 and 5).



**Figure 2a:** Epanechnikov kernel densities of the Hill and the moment ratio estimation errors for  $\xi = 0.25$  under the iid student-t model,  $N = 3000$ .



**Figure 2b:** Epanechnikov kernel densities of the Hill and the moment ratio estimation errors for  $\xi = 0.25$  under the GARCH-t model,  $N = 3000$ .



Interestingly, it turns out that all the approaches to optimal subsample selection show a common tendency to be positively biased for smaller values of the tail index (see Tables 2 through 5). For the GARCH models which are of particular interest regarding financial applications, there is also some tendency towards negative bias for larger values of the tail index especially for the MOT estimator in Tables 3 and 5.<sup>7</sup> In sum, bias as a function of  $\xi$  appears to be a critical issue in tail index estimation in that it causes the estimated degree of fat-tailedness to be systematically biased. Our evidence is that small tail indices (below 0.25) tend to be overestimated and large GARCH model tail indices (0.33 and above) tend to be underestimated.

### 3.3.5. GARCH versus iid Student Models

In comparing the results for the iid student-t and GARCH models, a first glance at the tables may lead to the conclusion that there is a relatively high overall mean squared error-robustness of the tail index estimators with respect to heteroskedasticity. Interestingly, volatility clustering under the GARCH models hardly increases the mean squared error of the optimal adaptive estimation approaches. For the sample size  $N = 1000$  the RMSE criterion yields even smaller values for the GARCH-t model than under the iid student-t model; for  $N = 3000$  the RMSE results for the two model classes are roughly identical. Again, as in the case of increases in the magnitude of the tail index  $\xi$  in Section 3.3.2., quite unchanged or even improved RMSE results come from the compensating effect of a reduction in estimation bias.

Although ARCH-type dependence does not necessarily increase mean squared estimation error in the simulation results, we know that ARCH-type volatility clustering leads to clustering in the extremes which should obviously increase estimation error for the tail index. This is confirmed by our results when we look at the STD statistics of the estimators which mostly show a notable increase in estimation variance under GARCH (Tables 2 through 5).

The estimators' error densities for both the iid student-t and the GARCH-t model with  $\xi = 0.25$  are illustrated in Figures 2a and 2b, respectively. It is apparent that the error distributions in Figure 2b have a larger dispersion than those in Figure 2a. Note that although estimation variance apparently increases

---

<sup>7</sup>Unreported results on return models with  $\xi > 0.33$  showed that negative, increasing bias for increasing  $\xi$  is not specific to the MOT estimator as documented in Table 5, but is also a characteristic of the bootstrap estimator.

under time series dependence in simulated returns, the overall RMSE results for the adaptive estimation approaches in Figure 2a and 2b are mostly quite similar due to reduced estimation bias under dependence. Figure 2 also suggests that the MOT estimator variance is relatively large under iid student-t data, but is hardly affected by ARCH-type dependence. This robustness characteristic under dependence appears to be due to the construction of the MOT estimator which, in contrast to the other estimators, is based on a Hill plot criterion.

### 3.3.6. Increasing the Sample Size

Comparing the results for the sample size  $N = 1000$  (Tables 2 and 3) with those for the sample size  $N = 3000$  (Tables 4 and 5) indicates that the relative performance ranking of the estimation approaches under the student-t and GARCH models is basically unaffected by an increase in the sample size; M-MOT and M-BSMC achieve the best RMSE results irrespective of sample size. Apart from that, the results suggest that an increased sample size yields larger performance improvements for the bootstrap (M-BSMC) as compared to the maximum occupation time (M-MOT) estimator (Tables 3 and 5).

Considering estimation precision as measured by STD, an increase in the sample size causes improvements in the estimators' standard deviations, especially under the iid student-t model with a small tail index (where the distance to the asymptotic bounds for the standard deviations is larger, see Section 3.3.2). Interestingly, there is a notable difference in the reduction of estimator variance between the GARCH and the iid student-t models. Under the GARCH models with  $\xi = 0.25$  and  $\xi = 0.33$  the standard deviation of the Hill as well as the moment ratio estimators can hardly be improved by increasing the sample size from  $N = 1000$  to  $N = 3000$ . This shows that the convergence to the asymptotic lower bounds for the estimator's standard deviation can be very slow under ARCH-type dependence.

## 4. Conclusion

By the very nature of extreme value theory, estimation bias and variance play an important role in studies of the tail index of return distributions. Upper order observations which typically deviate from an exact Pareto-tail cause the well-known bias/variance trade-off. Also, robustness with respect to assumptions about the underlying return model is important in financial applications.

Our small sample simulation results show that estimation bias will depend on the approach to adaptive subsample selection and the magnitude of the tail index itself. Small tail indices tend to be overestimated and large tail indices tend to be underestimated; thus, when tail indices are estimated for a cross-section of distributions, there is the tendency to underestimate the true range of tail index values. Such estimation bias is relevant to applications; e.g. too large holdings of high risk assets (based on underestimating their extremal risk) and too little holdings in low risk assets (where the risk estimate is upward biased) yield an inefficient allocation of capital. It also appears that large sample sizes are a necessary although not sufficient condition for a successful application of tail estimation procedures in finance. Our results indicate that convergence to the asymptotic lower bounds for the estimator's standard deviation can be very slow under ARCH-type dependence. Hence, when dependence is stronger for shorter return intervals, as often seems to be the case, the additional amount of data from higher frequency observations will not necessarily significantly reduce estimation variance.

In general, proper assumptions about the underlying return model and its time series properties are vital to the interpretation of the estimation results gained from extreme value analysis. What the tail index estimation results tell us about the extremal behavior of the returns cannot be separated from a priori assumptions about the return distribution. Our results suggest the following prudent steps in tail index estimation:

- Pay attention to the questions of sum-stability of the underlying distribution and potential time-series dependence in the dataset.
- Assuming a fat-tailed distribution with a finite second moment, apply the moment ratio estimator. Make Hill-type plots for a graphical illustration of the range of potential tail index estimates.
- Use at least two different adaptive estimation approaches. Our results suggest that the maximum occupation time estimator and the bootstrap ap-

proach perform well. Both approaches may be further refined as suggested in the literature.

- Assuming that GARCH-t is a suitable model for the return data, the estimation results may be interpreted using our results as a benchmark. In particular, the findings may be helpful in quantifying potential tail index estimation bias and error variance.

## References

- [1] **Blattberg, R. C., Gonedes, N. J. (1974)**: A Comparison of Stable and Student Distributions as Statistical Models for Stock Prices, *Journal of Business* 47: 244-280
- [2] **Dacorogna, M. M., Müller, U. A., Pictet, O. V., de Vries, C. G. (1995)**: The Distribution of Extremal Foreign Exchange Rate Returns in Extremely Large Data Sets, O&A Preprint, Zürich
- [3] **Danielsson, J., de Haan, L., Peng, L., de Vries, C. G. (1997)**: Using a Bootstrap Method to Choose the Sample Fraction in Tail Index Estimation, Working Paper, Erasmus University Rotterdam
- [4] **Danielsson, J., Jansen, D. W., de Vries, C. G. (1996)**: The Method of Moments Estimator for the Tail Shape Parameter, *Communications in Statistics–Theory and Methods* 25: 711-720
- [5] **Danielsson, J., de Vries, C. G. (1997a)**: Tail Index and Quantile Estimation with Very High Frequency Data, *Journal of Empirical Finance* 4: 241-257
- [6] **Danielsson, J., de Vries, C. G. (1997b)**: Beyond the Sample: Extreme Quantile and Probability Estimation, Working Paper, Erasmus University Rotterdam
- [7] **Dewachter, H., Gielens, G. (1994)**: A Note on the Sum-Stability of Speculative Returns, *Economic Notes* 23: 116-124
- [8] **Drees, H. (1998)**: Optimal Rates of Convergence for Estimates of the Extreme Value Index, *Annals of Statistics* 26: 434-448
- [9] **Drees, H., de Haan, L., Resnick, S. (2000)**: How to Make a Hill Plot, *Annals of Statistics* 28: 254-274
- [10] **Drees, H., Kaufmann, E. (1998)**: Selecting the Optimal Sample Fraction in Univariate Extreme Value Estimation, *Stochastic Processes and their Application* 75: 149-172
- [11] **Embrechts, P., Klüppelberg, C., Mikosch, T. (1997)**: Modelling Extremal Events for Insurance and Finance, Springer, New York

- [12] **de Haan, L., Peng, L. (1998)**: Comparison of Tail Index Estimators, *Statistica Neerlandica* 52: 60-70
- [13] **Hall, P. (1990)**: Using the Bootstrap to Estimate Mean Squared Error and Select Smoothing Parameters in Nonparametric Problems, *Journal of Multivariate Analysis* 32: 177-203
- [14] **Hill, B. M. (1975)**: A Simple General Approach to Inference about the Tail of a Distribution, *Annals of Statistics* 3: 1163-1174
- [15] **Hsing, T. (1991)**: On Tail Index Estimation using Dependent Data, *Annals of Statistics* 19: 1547-1569
- [16] **Huisman, R., Koedijk, K. G., Kool, C., Palm, F. C. (1997)**: Fat Tails in Small Samples, Working Paper, Maastricht University
- [17] **Janicki, A., Weron, A. (1993)**: Simulation and Chaotic Behavior of  $\alpha$ -Stable Stochastic Processes, Dekker, New York
- [18] **Jondeau, E., Rockinger, M. (1999)**: The Tail Behavior of Stock Returns: Emerging versus Mature Markets, Working Paper, HEC-School of Management
- [19] **Kearns, P. Pagan, A. (1997)**: Estimating the Density Tail Index for Financial Time Series, *Review of Economics and Statistics* 79: 171-175
- [20] **Koedijk, K. G., Schafgans, M. M., de Vries, C. G. (1990)**: The Tail Index of Exchange Rate Returns, *Journal of International Economics* 29: 93-108
- [21] **Lux, T. (2000)**: The Limiting Extremal Behavior of Speculative Asset Returns: An Analysis of Intra-Daily Data from the Frankfurt Stock Exchange, Working Paper, University of Kiel
- [22] **McNeil, A. J., Frey, R. (2000)**: Estimation of Tail-Related Measures for Heteroscedastic Financial Time Series: An Extreme Value Approach, Preprint, ETH Zürich
- [23] **Mikosch, T., Stărică, C. (1998)**: Limit Theory for the Sample Autocorrelations and Extremes of a GARCH(1,1) Process, Preprint, University of Groningen

- [24] **Resnick, S., Stărică, C. (1998):** Tail Index Estimation for Dependent Data, *Annals of Applied Probability* 8: 1156-1183
- [25] **Rubinstein, M. (1994):** Implied Binomial Trees, *Journal of Finance* 49: 771-818